UNIVERSITÀ CATTOLICA DEL SACRO CUORE

MILANO

Dottorato di Ricerca in Istituzioni e Politiche
Ciclo XXV

S.S.D.: SECS-P/02

Networks in Science:
Three Essays on Scientific Collaboration

Tesi di dottorato di: Lara Togni
Matricola: 3810439

Anno Accademico 2011/2012

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

MILANO

Dottorato di Ricerca in Istituzioni e Politiche
Ciclo XXV

S.S.D.: SECS-P/02

Networks in Science:
Three Essays on Scientific Collaboration

Coordinatore: Ch.mo Prof. Paolo COLOMBO

Tutor: Ch.mo Prof. Mario A. MAGGIONI

Tesi di dottorato di: Lara Togni
Matricola: 3810439

Anno Accademico 2011/2012

Table of Contents

Acknowledgments

The author truthfully thanks Mario Agostino Maggioni, who supervised this thesis meticulously and with scientific rigour; but, more importantly, he has been guiding me since the beginning of my studies, passing on me his passion for scientific research, and the love for his job.

The author also thanks Teodora Erika Uberti for her extremely useful insights and for having taught me the power of Network Analysis; Simona Beretta, Mario Nosvelli as well as the Faculty of the *Scuola di Dottorato in "Istituzioni e Politiche"* of *Università Cattolica del Sacro Cuore di Milano*, who, at different stages of my Ph.D., have all provided useful advice and support.

Dedico questa tesi di dottorato alle persone che più mi sono state vicine in questo percorso che, non nascondo, essere stato a tratti tortuoso. Questa tesi è il frutto non solo delle mie gioie e dei miei sacrifici, ma anche delle gioie e dei sacrifici delle persone che mi sono care. La mia famiglia, che mi ha sempre supportato nel mio lavoro e nelle mie decisioni "importanti" (che sono state molte negli ultimi anni); grazie a mamma Tiziana, papà Alberto e Stefano: facile non è stato, per molti aspetti, ma ora "il cielo è sempre più blu". E grazie a Cinzia e Italo che "loro ci sono sempre". Alonso, che ho conosciuto all'inizio del mio percorso e che ha vissuto con me questa esperienza e mi ha dato la forza di dare sempre un po' di più *sin llorar*. Infine, Mario, che veramente crede che "Education is not the filling of a bucket, but the starting of a fire".

Chapter 1

Introduction

1.1 On networks and science

The reader might feel puzzled by the way I am introducing this thesis. However I believe that briefly describing the personal story which lies behind this study may be one of the best ways of explaining the topic and methodology used.

During the course of my Ph.D., many people have asked me about the topic of my thesis. At first, I did not know where to start, and I asked myself how I could best explain the relevance of scientific production and collaboration; why it was actually worth investigating the mechanisms behind these two activities, and why I needed to extend the standard economic toolbox of analysis.

Scientific collaboration is an atypical topic, I must admit, and Network Analysis is a technique which is still rarely taught in M.Sc. in Economics around the world. I eventually realised the following was the best way of expressing this.

"We are all involved in the production of scientific knowledge and every day we experience the relevance of knowledge flows and knowledge exchange both in a horizontal (*i.e.*: peer-to-peer) and in a vertical (*i.e.*: lecturing, supervision) way. My research wants to explicitly address the structure and dynamics of the process of scientific production of knowledge by reducing its complexity to its basics components (what is economic thinking actually about?): networks, scientific communities, collaborative behaviours, individual incentives and collective outcomes".

Therefore this thesis is composed of three different essays, each of which approaches the phenomena of scientific production and collaboration from different perspectives, allowing me to highlight the role played by the above mentioned four basic components (*i.e.*: networks, scientific communities, collaborative behaviours, individual incentives and collective outcomes) in shaping the structure and dynamics of the "world of science".

Let me now describing the four themes, by defining and contextualising them in the history of the "economics of science", in order to support the understanding of the rationale behind this thesis.

In recent years, economists have devoted much attention to networks; studying networks, the ways in which they form and evolve dynamically is now widely recognised as the natural way of approaching the investigation of social, economic and organisational aspects of society. A *network* can be defined as the set(s) of actors and of relationships (or relational ties) which are defined on them (Wasserman and Faust, 1994). From this perspective, any kind of community can be defined by the characteristics of the relationships among the actors, rather than by the attributional characteristics of the actor themselves. When studying real life phenomena, including the scientific one, adopting a network's perspective allows us to go more in depth, hence providing explanations that go beyond "the easily observable".

The economic literature on networks has provided good examples of applications to various fields, such as markets (Berkowitz, 1988; Burt, 1988; White, 1981, 1988 and Leifer and White, 1987), occupational mobility (Breiger, 1981 and 1990), trade (Lazerson, 1993 and Nishiguchi, 1994); welfare (Bertrand et al., 2000), job hunting (Holzer, 1987 and Montgomery, 1991); patents (Maggioni et al., 2007) and, last but not least, the world of science (Maggioni and Uberti, 2009, Rivellini and Terzera, 2010, and Cainelli et al., 2010) – which is exactly what we are going to talk about.

In fact, over the last sixty years economists, and more broadly social researchers, have been attracted by the mechanisms which lie behind the world of science, in terms of the peculiar ways in which individuals part of this community interact with each other with a final view to creating scientific knowledge; hence producing scientifically relevant pieces of work. Here it comes the interest in studying *scientific production*, which may be defined as the process by which diverse inputs, such as scientists' training and knowledge, and capital invested in research and development are transformed into equally diverse outputs in the form of scientific publications; for example, patents, but also formal and informal knowledge transfers (Rousseau and Rousseau, 1997; Nagpaul and Roy, 2003; Warning, 2004; Johnes, 2006)[1].

[1] This interest within Academia was also accompanied by the general recognition of science as one of the most important drivers of economic growth and socio-cultural empowerment of every country (Wagner, 2008). This is one of the reasons why governments involved in the attempt to foster scientific research through the implementation of new incentive schemes (Shrum et al, 2007; Boyle, 2008; Frey 2003, 2009), which have contributed over time to

Bibliometrics and Scientometrics are the two disciplines which developed around this research interest, in the attempt to develop metrics to measure the effectiveness and efficiency of research activity by taking into account the mentioned inputs, but, above all, the outputs of the process of the creation of scientific knowledge (Moed, 2005; Weingart, 2005). Hence, indicators which consider the frequency a scientific product has been cited over time, the impact of a scientific journal within a particular discipline; or again the increased or decreased level of appeal of a particular scientific topic over time are all examples of issues bibliometrics is concerned with. However, this approach clearly lacks of an analysis which goes beyond the mere product of a scientific process, since it is widely recognised and emphasised that, in order to understand the outputs of the process, the inputs and the actors involved need to be accounted for. In fact, the process leading to a scientific output cannot be equated to the one leading to the production of a classical and conventionally tradable good.

Following the pioneer work by Crane (1972) and Price (1986), social researchers have found in the study of scientific collaborations a complementary way of explaining scientific production. *Scientific collaboration* is the collaboration among two or more scientists who work together in order to achieve the common goal to produce new and innovative research, hence it is the behaviour that lead to scientific production.

Therefore, if both individual and groups' behaviours are accounted for, we must consider *individual incentives* which move such behaviours. In fact, networks (including scientific networks) emerge from agents' behaviours, who are in turn moved by different kinds of incentives, that is to say, in this specific case, the maximisation of the impact of their publications within a particular scientific community.

This is the rationale behind the three essays this thesis is composed of. It represents a first attempt to answer *four research questions* which arise from the framework described above, by accounting for *three* different *methodological trade-offs* and studying *two scientific communities*.

1.2 Four research questions

This thesis aims at answering four different research questions which relate to the intriguing relationship between patterns of collaboration and scientific productivity.

---

modify exogenously some of the incentives that move scientific production as the ultimate goal of the "Republic of Science" (David, 2008:2).

In the attempt to answer the following questions, different methodological tools of analysis are adopted and integrated together in order to highlight their informative power in light of the peculiarity of the data available.

1. *What is scientific collaboration and is it possible to agree on a unique definition?*

Above we have defined scientific collaboration as a generic phenomenon, which consists of scientists working together in order to produce a piece of research.

However, scientific collaboration is a multi-faceted phenomenon that can take different forms. When we want to study this phenomenon empirically, it is necessary to clarify which type(s) of collaboration we are interested in, and we need to self-impose various "borders" that can help defining the collaboration itself.

The literature on scientific networks has provided different examples; we might attempt to categorise them into two different orders of collaborations:

i. *Formal* collaborations: this is the easiest type of collaboration to measure, since one can adopt as a proxy of collaboration the fact that two or more researchers have collaborated with each other, simply because they have published an article together in a scientific journal or have registered a patent together. Therefore, considering formal collaborations often means accounting for *co-authorships* as the way in which the collaboration between scientists is expressed;

ii. *Informal* collaborations: these are also known as *Invisible Colleges* (Crane, 1972) collaborations, and capture relationships between scientists resulting from behaviours which do not lead to a scientific product per se, but that in turn have influenced the production process. This type of scientific collaboration is harder to study as compared to the formal collaboration described in *i.*, since there is not a tangible output which can be easily measured. However, it is possible to track a scientists' network of informal collaborations by considering the *acknowledgments* list available in most published articles, or the fact that two or more scientists are members of the *same editorial board* of a certain scientific journal; or again that they are affiliated to the *same faculty/university*.

Certainly, the borders between informal and formal collaborations are often weak, since an informal relationship might turn out to be formalised later on, or, on the contrary, a formal collaboration can turn into an informal one. Numerous factor might affect a collaboration, and this is the reason why we like to think about scientific collaboration as a multi-faced phenomenon. For example, the spatial dimension is an important component, in terms of *geographical* or *institutional contiguity.* As well as the scientific dimension itself: the *scientific contiguity* with other scientists is one of the main factors which affects a scientist's willingness to collaborate with another scientist possessing the same expertise, or, on the contrary, the complementary expertise. Moreover, these conditions might change over a scientist's scientific career; hence, we consider scientific collaboration as a constantly evolving phenomenon, and a phenomenon which has "memory". Scientists build and update their scientific knowledge thanks to the knowledge of other collaborators they get in touch with, along different ways, which can range from having studied in the same university at the beginning of their career, or being experts in the same field; to just reading other scientists' publications.

As it will be better explained later on, this is the reason why this thesis attempts to study scientific collaboration by taking into account some of the different forms in which it is manifested (*e.g.*: co-authorship, acknowledgments relationships).

2. ***How can we measure the structure of scientific collaboration?***

Given the multi-faceted nature of scientific collaborations, it is hard to identify a tool of analysis or a unique metric which could help us analysing the complex structure of scientific collaborations and identifying overarching explanations of the phenomenon. It should be clear by now that this thesis will be looking at formal and informal scientific collaborations as a complex networks of relationships among scientists belonging to the same scientific community.

In this regard, we have looked for a tool of analysis which could allow us to describe and measure scientific collaborations so to be able to analyse their structure scientifically. The solution to this methodological problem was found in *Network Analysis* (henceforth, *NA*); in its own nature, this methodology looks at the phenomenon by capturing a statistical description of the features of the network, but also, and more importantly, it permits to carry out inference analysis. In other words, it has the advantage of being able to identify the architecture of the networks (*i.e.*: the *topology* of the networks), and to provide some static indicators (*e.g.*: degree distribution, clustering coefficient, average path length), along with different laws of motion (*e.g.*: random, scale-free, small world networks). And at the same time, its added value lies in the possibility of the statistical properties of the networks to fit the asymptotic requirements of theoretical models and hence to allow for inference analysis.

Although complemented by alternative methodologies such as Econometrics and laboratory experiments, the results obtained by applying NA will be the starting point of the three essays in this thesis.

3. Which is the relationship between scientific collaboration and scientific production?

This thesis aims at studying the relationship between scientific collaboration and scientific production from an *empirical* point of view. In doing so, it is inevitable to ask ourselves whether the patterns of collaboration directly affect the number and the quality of a scientist's scientific production in terms of publications; or, on the contrary, the number and the quality of a scientist's scientific publications influence his/her network of scientific collaborations in terms of the scientist's "attractiveness" and ability to be surrounded by many other scientists who are interested in collaborating with him/her.

This represents a crucial problem that most of the times forces us to choose between a *productivity-oriented* analysis, and a *collaboration-oriented* analysis. The former is usually focused on the determinants of a scientist's productivity, and attempts to model the phenomenon by integrating the traditional set of attributional characteristics with relational attitudes, amongst which the most commonly used is (formal and/or informal) scientific collaboration. On the contrary, the latter focuses on (formal and/or informal) scientific collaborations, and aims at investigating its determinants, amongst which there is also a scientist's productivity. In other words, one approach looks at scientific collaboration as one of the crucial determinants of productivity, whereas the other approach looks at scientific productivity as a crucial factor in making choices of collaboration (or no collaboration).

Therefore, the risk in an empirical analysis of the two phenomena is to underestimate a problem of reverse causality. A problem which will be addressed in this thesis.

4. How is it possible to explain the emergence of complex network topologies from basic individual incentives?

In order for a complete analysis of scientific networks of collaboration to be carried out, this thesis will inevitably attempt to consider the role that the scientist's individual attributes and personal attitudes play on influencing individual's choices of collaborations.

This is a crucial point, since we believe that production motives do not provide a fulfilling explanation of the phenomenon. Moreover, factors which relate to the features of the collaboration networks themselves, and to the influence of social preferences do need to be taken into account.

More in detail, we are interested in two different orders of evidence, which are extrapolated by the implementation of a pilot lab experiment. On one hand, we aim at understanding whether real networks of scientific collaboration match the theoretical predictions of scientists to converge into the creation of efficient and stable network architectures, since scientists internalise the negative externalities which are generated by networks themselves. On the other hand, we aim at analysing the influence that social preferences (cooperativeness and imitation of others) and fairness attitudes (equity and symmetry) might have on the decision-making process embedded in network creation.

Finally, we will also account for the potential behavioural difference between networks generated endogenously or exogenously.

1.3 Three trade-offs

This thesis takes into account three different (mostly methodological) trade-offs, which typically emerge from the study of networks, and are related to each other. Here they are described in an order that follows a "general to particular" scale.

In addressing the following trade-offs, different solutions are proposed as potential ways to either disentangle or reconcile them.

1. ***Micro* vs *Macro perspective***

The first trade-off is concerned with one of the most debated issues in the theory of economic networks. In fact, every network can be studied from two antithetical perspectives, which are hard to be reconciled with each other.

As Schweitzer et al. (2009) point out, the micro perspective is typical of network studies which develop in the field of economics and of sociology; whereas the macro perspective is adopted mainly in the literature on complex systems in physics and computer science. However, we believe that this trade-off emerges also within economics and sociology, in the extent to which the starting point of the analysis relies either on agents' incentives in the development of formal and informal links or on the statistical regularities of the network as a whole.

The *micro perspective* emphasises the role of individual's incentives in shaping the structure of the network, for then analysing the local structure of the network itself. On the contrary, adopting a *macro perspective* means to consider the global structure of the network, the rules of network evolution, for then analysing individual's performance within the network.

Each of the perspective has its own advantages and disadvantages, and one might choose one in spite of the other according to his/her research goals. However, we believe that the real issue that arises when studying network is the failure to develop a perspective that merges the micro with the macro. For example, in most of the cases, the micro perspective is not complemented by a macro analysis which allows to identify the complex system behind the network; in this way, loosing important aspects of the phenomenon under investigation which specifically regard the mutual influence between micro and macro characteristics (*e.g.*: exogenous and endogenous elements) of the phenomenon itself.

In this thesis, we take into account this crucial trade-off, and we also attempt to address the question of whether individual's characteristics directly influence the role played by the individual within the network; or, on the contrary, the network's architecture itself influences the role played by the individual within the network.

2. *Theoretical* vs *Empirical approach*

The trade-off between the theoretical and the empirical approach is directly linked to the first of our trade-offs. In fact, adopting a micro perspective typically reduces to a *theoretical analysis* of networks, while adopting a macro perspective to an *empirical analysis* of networks.

In this case, we are more concerned about the analytical tools which are used to carry out the analysis. A theoretical approach makes use of game theory, by considering a network as the outcome of a network formation game among either competing or cooperating agents, based on the concept of multiple strategic interactions. On the contrary, an empirical approach adopts mathematic, statistic and econometric tools of analysis in order to understand the structural regularities of a network in its formation and dynamical evolution.

As in the first of our trade-offs, here a general synthesis is missing: the core issue dividing the theoretical and empirical approach is individual rationality, and especially the relationship between individual incentives and overall societal welfare (Schweitzer et al., 2009). Seldom is this issue addressed in the "canonical" NA literature, while it constitutes the object of study of a branch of game theory literature dealing with "network formation: stability and efficiency"[2], whose aim is to answer questions such as : "How are such network relationships important in determining the outcome of economic interaction? Which networks are likely to form when individuals have the discretion to choose their connections? How efficient are the networks that form and how does that depend on the way that the value of a network is allocated among individuals?" (Jackson, 2003:12).

This thesis follows some of the attempts that have been made available in the literature[3] from an empirical perspective, mainly by means of behavioural and experimental economics, which provide answers to the questions behind the rationale of network formation (hence based on NA concepts) by creating scenarios which proxy the reality of networks in a controlled environment. Moreover, experiments could be considered as the natural research path to follow in order to integrate and push forward both "theory and practise" of networks; in fact, they can represent a valid tool for testing theoretical models, and at the same time providing a support to the analysis

[2] To quote the title of a well known paper, Jackson (2003).
[3] See, for example, Squazzoni et al. (2012).

of empirical findings. As Weibull (2001) points out, the integration of game theoretical models, field data, and lab experiments is the key to picture a more exhaustive framework behind several phenomena: "moving from armchair theorizing to controlled laboratory experiments may be as important a step in the development of economics as it once was for the natural sciences to move from Aristotelian scholastic speculation to modern empirical science" – and of course, networks are included among these phenomena.

3. Collaboration vs *Production analysis*

The last trade-off is directly connected to our third research question, which asks which kind of relationship can be drawn between scientific collaboration and scientific production. And whether this relationship might suffer of a reverse causality problem.

In fact, the trade-off between a *collaboration analysis* and a *production analysis* arises mainly because, in the operational reality of the research around networks, researchers are forced to take a view, although recognising the risks of both analysis. Assuming that scientists' collaborative behaviour affects the outputs of scientific production; or that scientific production affects the scientists' patterns of collaboration, inevitably drives the researcher into issues of endogeneity or omitted variable bias, as long as the problem of reverse causality is not correctly addressed, or at least the limitations of the research are not clearly defined.

Along this thesis, we are aware of the issues which are generated, and the limitations of each of the essays are clearly described.

1.4 Two scientific communities

In the analysis of networks, scientific networks in our case, the researcher needs to make important decisions in the early steps of his/her study. One of them is the correct definition of the sample.

In this thesis, we look at two different scientific communities in different essays, which are the result of a sampling process which moves from two different criteria. In fact, along this thesis we will discuss two opposite ways of looking at scientific communities, which mainly differ in the definition of the boundaries of the network and on the relevance of either relationships or individuals in the analysis of the network itself. Furthermore, it has to be specified that the results obtained from studies on networks, or from studies which use networks' characteristics to inform and integrate the analysis, are highly influenced by two main factors. On one hand, by the fact that the databases (*e.g.:* ISI Web of Science and MIUR-Cineca which are used in this thesis) which are accessed in order to construct the datasets are typically designed to carry out bibliometric research, rather than research with a view to analysing relationships themselves. On the other hand, by the boundaries that the researcher imposes to the network he/she wants to investigate; that is to say, by the criteria used to restrict or extend the relationships of each single actor part of the network. For example, these boundaries might be geographical, temporal or related to a particular scientific discipline.

1. Top Geographers

The first scientific community we look at is composed of *Top Geographers,* that is to say the networks of geographers who have published in the Top Journals (*i.e.*: the Journals with the highest Impact Factor) relevant to the discipline in a particular time span (from 2000 to 2007).

The sampling process originated from a common characteristics to our sample units: having published at least a scientific article in one of the Top Journals in the broad field of Geography. By doing this, our sample criteria is a criteria which focuses on the *relationship* (of co-authorship) between scientists, rather than on the scientist himself. Hence, the importance is assigned to the network and the relationships that is composed of, rather than to some attributes which are common to the agents part of the network (*e.g.*: geographical position, institutional framework).

This sampling method has the advantage of capturing all the possible relationships within the scientific community of the Top Geographers. However, it looses potential relationships which the scientists might have in place with other scientists outside the community of Top Geographers. For example, if a Top Geographer is also a Top Anthropologist who has published in the Top Anthropology Journals, we cannot capture neither the attribute nor the additional relationships that the Top Geographer/Anthropologist might have in place in the scientific community of Anthropology.

2. Italian Economists

The second scientific community is the community of Italian Economists, that is, the whole group of economists who are affiliated to an Italian university in a particular time span (from 1990 to 2010).

In this case, the sampling process originated from the subjects, and we did not define any particular boundary but being affiliated to an Italian university.

By doing this, our sample criteria is a criteria which focuses on the *actors* (and his/her attributes), rather than on the relationships between the scientists. Hence, the attention here is posed on the attributes of the scientists, and the network of relationships turns out to be limited to the extent to which relationships can be made possible only among Italian Economists.

Hence, we cannot analyse the whole set of relationships that our sample units have in place, because, for example, if an Italian economist has collaborated with another scientist affiliated to a foreign university, this relationship is not captured in our sample, although the attribute of the Italian economist to have collaborated outside the Italian community is.

1.5 Thesis outline

The thesis is divided into three different essays which illustrate the results of three studies with the specific aim to answer the research questions discussed.

Chapter 2:Scientific Networks and Co-authorship in Economic Geography

The aim of this essay is to investigate the role that different scientist's attributes and relational attitudes play in the decision-making process of building scientific collaborations. More precisely, we take into account different orders of differences, such as geographical, relational, experience, and production distance between pairs of scientists who are part of the scientific community of Top Geographers.

In particular, the attention is shifted from an "ego-perspective" to a "pair-perspective". Hence, a scenario in which we analyse all possible pairs of co-authors is built, and by matching different orders of characteristics we are able to suggest some mechanisms that could influence the success of a collaboration in terms of its "intensity", that is the number of published papers in the Top Journals of the scientific field considered. Moreover, we take into account difference/similarity between any two researchers potentially forming a pair, by constructing different orders of distance/proximity between authors, according to gender and experience attributes; geographical and relational factors.

We believe this approach needs to be considered as alternative, since the single author's productivity over time is not considered as a response variable, but rather as a possible explanation for collaborative behaviour.

The results of a Zero-Inflated Poisson model suggests that all the relational explanatory variables we include in this study do play a (positive) role in shaping scientific collaboration decisions and that within-between scientific sub-communities and informal collaborations – together with the "traditional" form of collaboration, namely co-authorship, are crucial determinants of scientific behaviour. As well as the spatial distance still affects negatively the likelihood or the success of a long-distance collaboration, despite the availability of new technologies.

The essay is structured as follows. Section 2.2 aims to describe the network of Top Geographers as the starting point of our analysis, while in section 2.3 the data are presented, together with the rationale behind the selection of the sample. Then, a detailed description of the variables we have built in order to run the econometric analysis is provided in section 2.4, followed by the econometric analysis (section 2.5). Finally, in the last section (section 2.6), we draw some conclusions and highlight important issues to be considered in our research agenda.

Chapter 3: Incentives and Behaviours in the Formation of Scientific Networks: An empirical estimate of Jackson and Wolinksy (1996)

This essay can be linked back explicitly to the trade-offs between micro *vs.* macro perspectives, and between theoretical *vs.* empirical approach, which were described above.

In fact, the aim of this essay is to build a bridge between the micro and macro approaches by developing an empirically testable version of two models (co-authorship *versus* connections) developed by Jackson and Wolinsky (1996) in order to test the relative importance of direct *versus* indirect relations on the scientific productivity of an individual scientist on the entire population of Italian academic economists.

The econometric analysis shows that the number and quality of co-authors are positively related to an economist's scientific production, while a negative and significant coefficient is registered for the *Connections* variable.

Further attributional (gender, tenure, localisation) and relational (centrality and clustering coefficient in the co-authorship network) variables affect the scientific productivity of academic economists in Italy.

The essay is structured as follows. Section 3.2 describes the theoretical model of Jackson and Wolinsky (1996) which our study originated from, whereas in section 3.3 we suggest an empirical version of it. Then, section 3.4 describes the variables which are taken into account in order to test the empirical model, and section 3.5 illustrates the sampling process and the data gathered, together with some descriptive statistics. Finally, section 3.6 provides the empirical results obtained by a Tobit estimation of the scientific production of Italian economists, while section 3.7 concludes with some suggestions for further research.

Chapter 4:Co-authorship Networks: An experiment on Network Formation, Efficiency and Stability

The aim of this essay is to provide a first insight into the process of network formation in the more specific context of collaborative networks. In other words, we aim at studying the way in which networks emerge in a cooperative setting, and their features in terms of efficiency and stability, by letting people play connections game in a controlled laboratory experiment.

Following the co-author model by Jackson and Wolinsky (1996) and its experimental implementation suggested by Vanin (2002), two co-author games have been designed and implemented in a controlled laboratory environment in order to shed light on the emergence of the trade-off between network stability and efficiency. In particular, we aim at studying the way in which networks emerge in a cooperative setting, and results show that even when we let people form their own collaborative network *under the most favourable conditions* (*e.g.*: communication amongst participants), it is not clear whether people systematically reach the strongly efficient network configuration or not. Moreover, evidence of the trade-off between pairwise stable configurations and efficiency is provided: most of the time participants converged and agreed on building a strongly efficient architecture, which, on the other hand, is not pairwise stable. With regard to people deviating from rational decision-making, results show that people do not care only about utility maximisation, but they are also concerned with fairness motives and with payoff symmetry amongst individuals.

Finally, through a simple statistical analysis and the elaboration of both an individual and a group index of cooperativeness, we showed the importance of investigating the role of people's cultural features and degrees of cooperativeness. We provided a first insight into the crucial role these factors could play on affecting individuals' connection decisions. However, it was not possible to take into account individual characteristics (*e.g.*: gender and years spent in the UK) because our sample did not show variance with respect to them; in other words, we could not make any consistent observations regarding statistically significant differences among groups of individuals, due to the fact that our sample was too small to allow us to infer such information.

The essay is organised as follows. Section 4.2 provides the reader with the basic concepts and definitions which are adopted in our study, followed by a formal presentation of the co-author model of Jackson and Wolinsky (section 4.3). Next, section 4.4 presents the hypotheses, the design and the procedures used in the experiment, together with the obtained results and possible explanations with reference to participants' personal and cultural features, and their level of cooperativeness. Section 4.5 concludes by drawing some final observations and suggesting improvements for further research.

Chapter 2

Scientific networks and Co-authorship in Economic Geography

2.1 Introduction

We must recall from the general introduction of this thesis that the interest of economists towards the "Republic of Science" (David, 2008) has grown out of the methodological framework of Network Analysis (*henceforth*, NA). Following the pioneer works by Crane (1972) and de Solla Price (1986), back to the '70s and '80s social scientists began to investigate the mechanisms which lie behind scientific production by using appropriate tools of analysis, which can help to explain its own patters of knowledge creation and to go beyond mere bibliometric indicators. Thanks also to the technical contributions provided by physicists (Watts and Strogatz, 1998; Newman, 2001; Albert and Barabási, 2002; Barabási et al, 2002) and computer scientists (Broder et al., 2000; Daley and Gani, 2000), a more systematic analysis of scientific networks has been carried on, allowing to unveil some of the drivers of each particular scientific community.

In this framework, we have already explained the emergence of a trade-off between a *production approach*, focused on the determinants of a researcher's productivity, which attempts to model the phenomenon by integrating the traditional set of attributional characteristics with relational attitudes, amongst which the most commonly used is co-authorship[4]; and a *collaboration approach* whose focus is on scientific collaborations, and aims to investigate its determinants, amongst which there is *also* productivity. This second approach to the study of networks has also introduced into this field many innovative tools of analysis, such as NA (Wasserman and Faust, 1994; Scott, 2000), Textual Analysis (Maggioni *et al*, 2009; Roth and Cointet, 2010), and

[4] Other measures which relate to researchers' relational attitudes which have been already used in the literature include the attitudes of researchers to collaborate with the same co-authors over time (Cainelli et al. 2010 and 2012), or a researcher's propensity to get involved in informal collaborations (Togni, 2011).

Experimental Economics (Jackson and Wolinsky, 1996; Vanin, 2004; Maggioni and Togni, 2012; Togni, 2011) alongside the most traditional ones.

In this essay, we will be adopting a collaboration approach, since the author believes that formal and informal scientific relationships do matter in regards not only to scientific productivity, but also to the strategic motives which might boost scientists to choose a particular collaborator in spite of another. In fact, one of the most common reasons which have encouraged economists to investigate the intriguing relationship between scientific productivity and collaborative behaviour (both formal and informal) relies on the belief that collaboration might enhance both quality and quantity of publishable scientific work. Nevertheless, this assumption has proved to be also one of the most controversial, given the antithetical evidence provided by empirical works on the subject[5] (Hollis, 2001). Moreover, the majority of contributions focus on the single individual's attributes, characteristics, and networks features, and little or no attention is devoted to what we could call a "reverse causality" effect: which is, in fact, the role that individual characteristics play on the decision of collaborating with other scientists? In which way scientific productivity might affect collaboration choices?

2.2 The scientific network of Top Geographers

Owing to the fact that each scientific discipline has its own features and "informal rules" which contradistinguish itself from any other, our first goal was to select one that satisfied our sample requirements to be big enough and interdisciplinary, but at the same time could allow us to draw a subsample in order to carry out a second-order analysis[6]. Hence, we opted for the scientific community of Geographers, which is knowingly characterised by multidisciplinarity; additionally, to our knowledge, there is no empirical contribution available in the literature that directly studies Top Geographers' patterns of collaboration, and it seems to be perfect to investigate the role played by different dimensions of spatial and non-spatial "distances".

As it will be detailed in section 2.3, we refine and extend an original dataset built by Togni (2009, 2011), whose information are extracted from *Web of Science (WoS)*[7], one of the largest bibliographic datasets provided by *Thomson Reuters*[8]. The information we are interested in

[5] For example, Hollis (2001) has shown that higher co-authorship is correlated with higher quality, greater length, and higher frequency of publication. Nonetheless, if we take into account the relationship between co-authorship and the "output" attributable to the single researcher (economist in Hollis' study) is negative, if we discount it for the number of co-authors.
[6] For additional details, please refer to section 2.3 of this chapter.
[7] http://www.thomsonreuters.com [16th October 2012].
[8] Please note that *Thomson Reuters WoS* was formerly called ISI *WoS*.

concerns the scientific discipline of Geography (which is also characterised by a strong interdisciplinarity with other social sciences, including Economics) and its pattern of publications during eight years (2000-2007).

In order to construct and analyse the network of Top Geographers, we decided to gather all the bibliometric information regarding the first five "Top Journals", which have been downloaded. A scientific journal is classified as "Top" if its Impact Factor index (*henceforth*, IF) is ranked in the first 5 per each year. IF is calculated as follows:

$$IF = \frac{two - year\_tot\_citations\_t_3}{tot\_art\_t_1 + t_2\_A\_journal} \quad \textbf{(2.1)}$$

Where $two - year\_tot\_citations\_t_3$ is the amount of citations that the papers published in a generic "A journal" in a two-year period receive in the year following the publication; and $tot\_art\_t_1 + t_2\_A\_journal$ is the total number of articles published in such "A journal" in those two years.

Hence, IF rankings per each single year (2000-2007) were accessed in order to build the dataset, and the Top 5-Journal ranking was downloaded per each year. It has to be remarked here that the total number of Journals we are considering in order to draw our sample amounts to eight, rather than five. This is due to the fact that the annual IF ranking is not completely stable over the years we consider, and some Journals that were ranked in the first five positions per IF in one year might not be ranked the same in following years. Hence, some Journals enter the Top-5 ranking, and some others are scored down instead. More in detail, the *Annals of the Association of American Geographers (AAAS)* is occupying the fourth place in the Top-5 ranking in 2000, for then being scored down to the sixth place in 2002, and being replaced by *Economic Geography (EG)*. In turn, *EG* and *Global Environmental Change (GEC)*, which in 2002 occupies the fourth and fifth places respectively, will then be replaced by *Environment and Planning – D (EPD)* and by *Political Geography (PG)* the following year[9]. **Table 2.1** shows the annual IF ranking and the eight top journals which were selected.

[9] Please note that, although the mentioned Journals exit the Top-5 positions of the rankings, they anyway occupy a position which is within the first 10 in the ranking (over the time period considered).

Table 2.1 – Top Journals and Impact Factor

		2000		2001		2002		2003		2004		2005		2006		2007	
ID	**JOURNAL**	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank
TIBG	TRANSACTIONS OF THE INSTITUTE OF BRITISH GEOGRAPHERS	4.067	**1**	3.500	**1**	2.218	**3**	2.388	**3**	2.438	**2**	2.574	**3**	3.093	**1**	2.698	**1**
GEC	GLOBAL ENVIRONMENTAL CHANGE - HUMAN AND POLICY DIMENSIONS	3.915	**2**	2.600	**3**	1.952	**4**	/	/	/	/	/	/	/	/	/	/
PHG	PROGRESS IN HUMAN GEOGRAPHY	3.762	**3**	3.440	**2**	2.616	**2**	2.943	**2**	3.653	**1**	2.762	**1**	2.288	**2**	2.386	**2**
AAAG	ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS	2.962	**4**	2.141	**5**	/	/	2.115	**5**	1.972	**5**	2.586	**2**	1.855	**3**	2.109	**4**
JEG	JOURNAL OF ECONOMIC GEOGRAPHY	2.679	**5**	2.519	**4**	3.222	**1**	3.139	**1**	/	/	/	/	/	/	/	/
EG	ECONOMIC GEOGRAPHY	/	/	/	/	1.757	**5**	2.325	**4**	/	/	2.455	**4**	/	/	1.909	**5**
PG	POLITICAL GEOGRAPHY	/	/	/	/	/	/	/	/	2.250	**4**	/	/	1.519	**5**	/	/
EPD	ENVIRONMENT AND PLANNING D-SOCIETY & SPACE	/	/	/	/	/	/	/	/	2.269	**3**	2.377	**5**	1.583	**4**	2.152	**3**

The sampling process originated a list of eight "Top Journals" which published a total of 2,474 articles (in the period 2000-2007) by 2,436 Top geographers (authors). To be reminded here that the dataset has been painstakingly cleaned in order to avoid misspelling problems and other trivial mistakes[10].

The attributional characteristics we retrieved from this dataset will be further detailed in section 2.3. The aim of this section is in fact to provide a first glance at the structure of the Top geographers' network, and to understand which network's architecture can best describe it (*e.g.*: scale-free network, random network, small world network).

Table 2.2 lists the typical network indices computed in a NA analytical framework, as compared to the average of the same indices for 10 random networks which were generated by taking into account the dimension of the real network of Top geographers[11].

Table 2.2 – Network topology (n2436)

	Random Net	**Top Geographers' Net**
Nodes	2,436	2,436
Average Degree	2.83	2.83
CC	0.001	0.956
APL	7.472	4.929
CC'	956	
APL'	0.66	
Q *	1,449.23	

**Uzzi et al. (2007) index. Please see footnote 12 for more details.*

In particular, we might suspect that the Top Geographers's network shows the typical features of a *Small World* (Watts, 1999 and Watts and Strogatz, 2001), that is, a type of network whose characteristics are common to the majority of real and social network. Small World networks are identified in NA literature as networks in which nodes display a low degree centrality (see Appendix G) but are nonetheless strongly connected with each other thanks to the role played by "strategic" nodes, commonly referred as "hubs" of the network. Hence, those networks report an high level of *density*, which can be defined as the ratio between the actual number of links within

[10] Deriving, for example, from the inclusion or not of middle initials, "Mortimore, M." appeared in ISI sometimes as "Mortimore M. L." or as "Mortimore M. J. L.". We opted for the latter option.

[11] Typically, the number of nodes in the network and the average degree of the network are the two measures/dimensions which are considered when generating a random network to be benchmarked against the real one.

a network (L), and the potential number of links that could be built within the same (Wasserman and Faust, 1992). In other terms, the density of a network (D_g) is given by

$$D_g = \frac{L}{\frac{n(n-1)}{2}} = \frac{2L}{n(n-1)} \qquad \textbf{(2.2)}$$

where n represents the number of links populating the network. This index takes values between 0 (in the case of a completely disconnected network – the network does not actually exist, since it would be composed of isolated nodes only), and 1 (in the case of a maximally connected network – all nodes are connected to each other).

As it is possible to assume from table 2.2, in order to understand whether the network of Top geographers is actually a small world or we need to consider other network's topologies, we also need to take into account other two indices of NA, typically the Clustering Coefficient (CC), which corresponds to the density described above, and the Average Path Length (APL), that is the shortest average distance between any two nodes in a network, and then compare these indices derived from the observed real network with one which is randomly generated according to the number of nodes, average degree, and density which characterise the real network. More formally,

$$CC' = \frac{CC}{CC_C} \qquad \textbf{(2.3)}$$

$$APL' = \frac{APL}{APL_C}$$

where CC_C and APL_C are, respectively, CC and APL of the randomly generated network. If CC' is greater than 1, and APL' is approximately equal to 1, then we are analysing a Small World network.

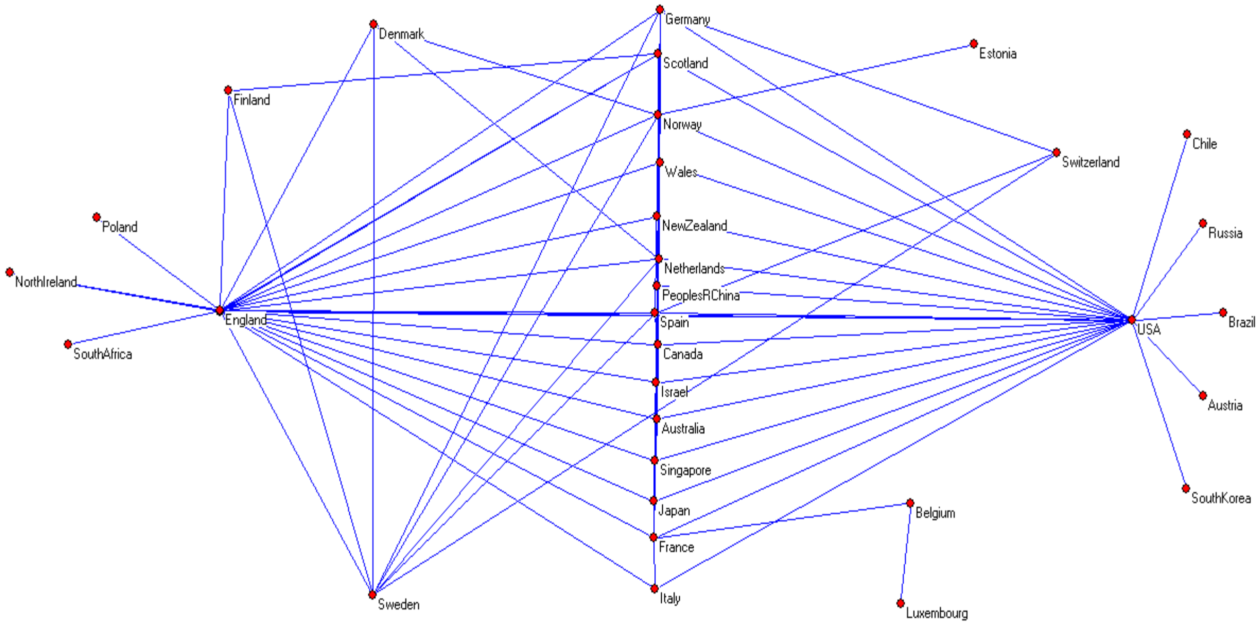
By using the indices suggested both by Watts (1999) and Watts and Strogatz (2001), and Uzzi *et al* (2007)[12], we prove that our network displays high level of clustering, but at the same time a relative short distance between any scientist in the network: the Top geographers' network is a small world.

Moreover, it is also interesting to have a look at the spatial network which is generated by considering as nodes the countries where the universities in which Top geographers are affiliated

[12] Uzzi et al. (2007) built an alternative method to test a network against Small World features, according to which a Small World Q is characterised by a ratio between CC' and APL' which needs to be largely greater than 1.

to are based. For the sake of simplicity, we extracted the main component of the spatial network, that is the sub-network in which every node is maximally connected: no isolated nodes are part of a main component. For example, suppose that there are only two Italian geographers who have published an article in our Top Journals in the time period considered. Let us assume that they wrote one paper alone and one jointly, but none of them has written a paper in co-authorship with a geographer affiliated to a University in the UK. In this hypothetical scenario, Italy would be an isolated node in the network, hence it would not be part of the main component of the spatial network. **Figure 2.1** depicts the main component of the spatial network of geographers.

Figure 2.1 – Spatial network of Top Geographers – Main component (n33)



On the extreme left and extreme right of the network's representation, the two most central countries are represented, namely England and the USA[13]. On their left and right respectively, countries that would be isolated if it was not for their connection to the most central countries are pictured. In the vertical line in between England and the USA, countries which have scientific relationship with both most connected countries are displayed, but also with the other countries. Moreover, countries in the space between England and the "vertical line" are those that are directly connected to England, and indirectly to the USA; whereas in the space between the USA and the "vertical line" are placed those countries which have indirect relationship with both the

[13] Please note that we opted for distinguishing between England, Wales, Scotland, and Northern Ireland because the results were not affected by aggregating them into the UK; moreover, we believe this choice reflects the actual network of Institutions across the UK better than considering the constituent countries together.

USA and England. It is worth deriving a ranking of these countries (**table 2.3**), in order to support the graphical representation which anyway tells us at a glance about the hierarchical composition of the spatial network. In order to do so, the traditional indices of centrality used in NA were computed[14].

Table 2.3 – Ranking of countries with respect to network centralisation indices

		DEGREE CENTRALITY		CLOSENESS CENTRALITY		BETWEENNESS CENTRALITY	
NODE	**COUNTRY**	DEGREE	RANK	CLOSENESS	RANK	BETWEENNESS	RANK
2	USA	22	1	74.419	1	44.873	1
1	England	22	1	74.419	1	37.805	2
5	Germany	9	3	56.140	3	4.174	6
4	Canada	9	3	56.140	3	1.654	9
18	Sweden	8	5	49.231	15	2.662	7
20	France	7	6	55.172	5	12.500	3
11	Spain	6	7	51.613	6	2.032	8
3	Scotland	6	7	50.794	8	1.109	10
8	Australia	6	7	50.794	8	0.151	14
15	Norway	5	10	51.613	6	6.979	4
9	Netherlands	5	10	50.794	8	0.796	11
29	Singapore	5	10	50.794	8	0.796	11
31	Japan	5	10	50.000	12	0.722	13
27	Belgium	2	20	36.782	31	6.250	5

As predictable, in the field of Geography, England and USA emerge as the two most powerful countries in the network. And not only because of the mere number of articles which the Top geographers affiliated to their universities produce, but also for their ability to get (and keep) the whole network connected, by playing the crucial role of "hubs" amongst all the other countries. In other words, England and the USA prove to be "attracting science" from other countries, hence talented geographers, whose aim to be interconnected to those colleagues working there;

[14] For more details about these indices, please refer to Wasserman and Faust (1994), Scott (2000) and Appendix G of this thesis.

but they also believe that collaborating with the USA and England is strategic in order to be connected with other countries.

2.3 Data and sample selection

Section 2.2 described some general features of the scientific network we are considering in our analysis, together with a presentation of the source of our data. We now proceed with the full description of the information we gathered and the rationale behind a further restriction of our sample.

Table 2.4 – Original dataset description

Data	**Number**
Years	8
Top Journals	8
Published articles	2,379
Number of authors	2,436
Number of acknowledgments	7,730
Number of acknowledged authors	636
Number of acknowledgments per author (average)	3.17
Number of acknowledgments per article (average)	3.25

As already mentioned in the previous section, our initial units of analysis consist of all geographers who published at least one article in one or more of our five “Top Journals”[15] in the field of Geography in the 2000-2007 time-span. Hence, it should be remarked here that our sample selection follows a sort of “macro approach”, according to which we start from a feature common to all of our units of analysis (*i.e.*: the geographers who supposedly have published high-quality scientific works, since we consider only the Top Journals), in order to derive the structure of the whole network; instead of considering each single Top geographers and its own ego network at the first place.

Following Togni (2009, 2011), we took into account information concerning both authors and acknowledgments available in most of the articles part of our sample: 636 of the acknowledged persons were also geographers who have published in the Top Journals. **Table 2.4** summarises this information.

[15] Please recall from section 2.2 that the selected Top Journals are actually eight. This is due to the fact that we considered the top-5 ranking for each of the eight years considered: some of the Journals appear as top-5 every year, some others do not being replaced by better performing Journals which were not amongst the top-5 on previous years.

Nonetheless, it should be made clear at this point that the scientific discipline of Geography is strongly characterised by interdisciplinarity. Therefore, accounting also for the fact that we needed to restrict our sample in order to perform the econometric analysis (section 2.5), we decided to limit our attention to those Top Journals which are mostly devoted to scientific findings in the field of Economic Geography. Additionally, this decision allows us to benchmark the scientific and collaborative behaviours of those authors who publish only in the above mentioned Journals against those who do not; and against those who tend to publish in both (Economics related and Geography-only related Journals).

Table 2.5 – Economic Geographers' dataset description

Data	**Number**
Years	8
Top Journals	8
Published articles	320
Number of published articles written by 1 author only	163
Number of published articles written in pair	117
Number of published articles written by more than 2 authors	40
Number of authors	409
Number of possible pairs	83,436

Hence, we reduced our sample to those Top geographers (and their acknowledgments) who have published at least an article in the period 2000-2007 on the following two Top Economic Geography Journals[16]: *Journal of Economic Geography (JEG)*, and *Economic Geography (EG)*. Our dataset is now composed of 409 authors who wrote 320 Journal articles between 2000 and 2007 in the two Journals we have just mentioned. Moreover, for the sake of our analysis, we did consider also the articles that the Top Geographers composing our new sample published in the other six Top Journals[17]. Note that our unit of analysis is not the single author but each possible combination of *ij* pairs; excluding *ii* pairs, a total of 83,436 pairs[18] were included in our dataset. **Table 2.5** reports a summary of the description of the dataset.

For each author, besides bibliometric information, different orders of data were gathered: *attributional*, such as gender and first year entry in WoS database; *spatial*, such as University of affiliation and city; *relational*, such as number of acknowledgments and number of co-authors; and related to their patterns of *scientific production*.

[16] Please refer to table 2.6 for further details.
[17] Please refer to section 2.4 of this chapter for further details.
[18] Since $\frac{n\times(n-1)}{2}=\frac{409\times(409-1)}{2}=83{,}436$

2.4 Variables description

Having described the data we are using in our research, it is now necessary to further detail the variables we constructed in order to address our research questions.

Given the fact that our main interest lies on the role that different attributional, relational, and spatial dimensions might play on influencing a scientist's decision to collaborate with a certain colleague rather than any other, we need to construct a dependent variable which in its own nature allows to account for all available collaboration choices that each scientist could potentially have. Hence, our unit of observation is represented by a pair of authors, rather than a single individual. In other words, our observations are all the possible combinations of pairs, therefore, all potential pairs of co-authors in our dataset.

To do so, we constructed a dependent variable ($Pair\ Art_{ij}$) which is the total number of articles each single pair has *jointly* written in at least one of the two Top Journals between 2000 and 2007.

With regard to the explanatory variables, we like to think about them as different kinds of "distance/proximity" (hence difference/similarity) between the two authors forming the pair, which could possibly influence their likelihood to collaborate in the production of a scientific piece of work. In particular, we divide our independent variables into 5 groups according to the type of "distance" they proxy: *gender*, *experience*, and *geographical* distance; *relational* and *production* distance.

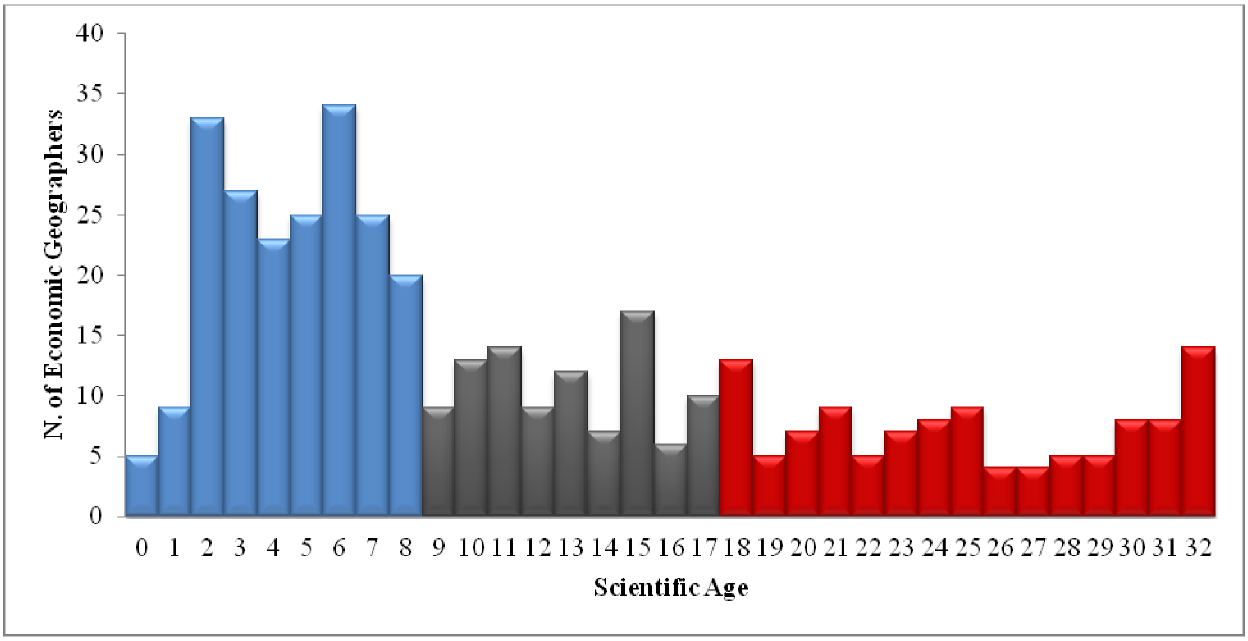
a. Gender distance

We recorded the gender of each of the two authors (i and j) forming the pair, in order to assign to the pair itself the attributional status of either same gender or mix gender. In particular, the variable M_i-M_j is a dummy variable which takes value 1 if both authors forming the pair are males, and 0 otherwise. On the contrary, the variable F_i-F_j is a dummy which takes value 1 if both authors are females, and 0 otherwise. Finally when the dummy M_i-F_j takes value 1, that means that the pair is formed by one female and one male.

b. Experience distance

In order to derive a proxy of an author's experience in the field of Geography, we first recorded his/her first year entry in WoS. In other words, we looked for the year in which each author had first published an article in any of the Journals covered by WoS since 1975.

Figure 2.2 – Distribution of Top Economic Geographers according to their Scientific Age *



Note that the Young ($0 \leq$ Sci $Age_i \leq 8$) group is coloured in blue, whereas the Old ($18 \leq$ Sci $Age_i \leq 32$)group is coloured in red.

** Information extracted from Web of Science database as of 31st December 2007.*

Then, we computed the author's scientific age as of 31st December 2007 by calculating the difference between the last year we are taking into account (2007) and the author's year of entry in WoS database. We can expect that the older the author, the greater the experience he had gained. Finally, since we are interested in a measure of distance/proximity of experience between the authors forming the pair, we derived our variable *Sci* Age_{ij} by simply computing the difference between the scientific age of author *i* and the scientific age of author *j*.

In addition to a continuous explanatory variable which captures the scientific experience of the Top economic geographers (*Sci* Age_{ij}), we built a complementary categorical variable, *Sci* $Generation_{ij}$, with the view to capturing the potential effect of scientific relationships which are often described in the literature as "mentor/PhD student" collaborations. In order to do so, we looked at the distribution of the scientific age along our sample of Top economic geographers (**Figure 2.2**), and we split this into two age-intervals: a group of Top economic geographers who are scientifically less than 8 year old; and a group of Top economic geographers who are scientifically more than 18 year old. Therefore, we created two categories, the *Young* ($0 \leq$ *Sci* $Age_i \leq 8$) and the *Old* ($18 \leq$ *Sci* $Age_i \leq 32$). In this regard, by looking at the distribution in Figure 2.2, we can easily see that the number of Young economic geographers is bigger than the number of Old economic geographers. This might be due to the fact that the topics covered by the discipline of economic geographers might have been more attractive for a new generation of economists as compared with the older generations. Finally, *Sci Generation*$_{ij}$ is a dummy variable which takes value 1 if the two geographers forming a pair belong to different scientific generations (*e.g.*: they are either *Young/Old* or *Old/Young*), and 0 otherwise.

c. Geographical distance

Information regarding the authors' affiliations were downloaded when our dataset was built. In doing so, we were able to gather data regarding the geographical location of each of the observations. In particular, we designed a matrix in which the geographical distance between any two cities in the dataset was calculated according to the shortest distance in kilometres. In this way, it was possible to assign a spatial distance to each pair in the dataset, and create the explanatory variable *Geo Dist*$_{ij}$.

d. Relational distance

The capacity of each author to build scientific relationship has been considered as multi-faceted. In fact, there are different type of relational attitudes that an author could display, and are not only related to the phenomenon of co-authorship. More in detail, we created four different explanatory variables which address this aspect, which may be divided into three groups.

First of all, we constructed two different variables which account for the author's tendency of writing scientific articles in co-authorship with other colleagues. The variable *Co-auth*$_{ij}$ is based on the classical index of co-authorship: at first, we divided the number of co-authors of author *i* by the number of articles published by author *i*. In this way we could get the author's average number of co-authors per published article. Then, for each pair *ij* we computed the difference between *i*'s co-authorship index and *j's* co-authorship index. While this variable allows us to account for an author's tendency to collaborate, it does not tell us anything about the stability of his/her collaborations over time. Hence, we constructed the variable *Stab Co-auth*$_{ij}$ which is the sum of *i* and *j* co-authors' stability index. The co-authors' stability index (Cainelli *et al*, 2010 and 2012) captures the author's preference over stable collaborations, and its values ranges between 0 (absolute instability) and 1 (complete stability). In order to construct this variable, differently from *Co-auth*$_{ij}$, we did not consider just the mere number of co-authors, but we also needed to distinguish their identities. These two variables are to be considered as the attempt to capture the capacity of building scientific and collaborative relationship within the scientific community.

Additionally, since we were also interested in patterns of informal scientific collaboration, we used the number of acknowledgments per author available in our dataset, in order to construct a variable (*Invisible Coll*$_{ij}$) which could act as a proxy of the distance or proximity in *i* and *j* tendency to create a network of informal collaborations (that of course do not result in a published paper, and therefore is more difficult to track). In order to do so, we simply calculated the difference between *i*'s total number of acknowledgments and *j*'s total number of acknowledgments.

Moreover, we were curious to understand whether a sort of "contamination" amongst different branches of Geography with exclusion of Economic Geography could play a role. Hence, the variable *Out Community*$_{ij}$ served as a measure of the similarity/dissimilarity between *i* and *j* in terms of openness and inclusion in the broader scientific community of Geography. This variable was constructed as the sum of the (not necessarily co-authored[19]) articles that *i* and *j* wrote in the same Top Journal(s) part of the original dataset which is neither JES or EG. Note that a necessary condition in order for this measure to take value greater than 0 is that both authors wrote at least an article in the *same* Top Journal.

Finally, we decided to take into account two measures of network centrality per each Top economic geographer part of our sample, by computing the classical indices of *betweenness* and *closeness centralities* within the *macro* network of Top geographers. The rationale behind our choice of considering indices derived from the network of the entire community of Top geographers lies on the fact that we cannot consider the network of economic geographers as an actual network; indeed, we can think about it as a network which was created and imposed *exogenously* by the author, but that actually originated *endogenously* from the scientific relationships built within the network of the Top geographers' community. To give an example of the issues which could arise, authors who seem to be disconnected in our superimposed network might actually have a (direct or indirect) scientific relationship between each other within the *macro* network.

In particular, we are interested in accounting for the difference between the roles that any two geographers forming a pair play in the whole network as "hubs"; that is to say, how important they are in terms of acting as bridges between the other geographers in the network, hence assuring connectivity of the whole network. In order to do so, we derived the index of betweenness for both geographers forming a pair. And their role of assuring ease of "achievability" of any node within the network, which is captured by the closeness centrality computed for all Top economic geographers. Finally, we calculated the difference in

[19] This would be beyond our scope.

betweenness (*Betweenness$_{ij}$*) and closeness (*Closeness$_{ij}$*) indices of the authors forming a pair; these two variables need to be considered as the similarity/dissimilarity between potential co-authors in terms of the strategic roles they play within the network.

e. Production distance

As a measure of distance/proximity in terms of each author's production capability, we simply computed the difference between the total number of article published by *i* and the total number of article published by *j* in the time span and Top Journals considered. We named this variable *Prod$_{ij}$*.

Table 2.6 reports the descriptive statistics.

Table 2.6 – Descriptive statistics

Variable name	Observations	Mean	Std. Dev.	Min. Value	Max. Value
Pair Art$_{ij}$	83,436	.0034877	.0619285	0	4
M_i-M_j	83,436	.6287933	.4831305	0	1
F_i-F_j	83,436	.0390719	.1937671	0	1
M_i-F_j	83,436	.3213682	.4670046	0	1
Sci Age$_{ij}$	83,436	10.1719	8.024816	0	32
Sci Generation$_{ij}$	83,436	.6909967	.4620853	0	1
Geo Dist$_{ij}$	83,436	5631.374	4422.308	0	19,854
Co-auth$_{ij}$	83,436	1.081592	1.240492	0	8
Stab Co-auth$_{ij}$	83,436	.4890612	.4137633	0	1
Invisible Coll$_{ij}$	83,436	7.508155	18.35405	0	413
Out Community$_{ij}$	83,436	.1433074	.9643211	0	33
Betweenness$_{ij}$	83,436	.0051429	.0216592	0	.182
Closeness$_{ij}$	83,436	.0138546	.019021	0	.044
Prod$_{ij}$	83,436	.5492407	1.092859	0	6

2.5 The econometric analysis

In this section, the model we have developed and the results of a Zero-Inflated Poisson Regression are presented and commented. First, we built a model which allowed us to bring together the different orders of explanatory variables illustrated in the previous section, in order to assess their potential influence on the possibility that any two authors in the dataset collaborate or do not collaborate in the production and publication of a scientific paper. Hence, our generic model can be written as:

$$PairArt_{ij} = f\begin{pmatrix} M_i - M_j, \beta_2 F_i - F_j, SciAge_{ij}, SciGeneration_{ij}, GeoDist_{ij}, \\ Co-auth_{ij}, StabCo-auth_{ij}, OutCommunity_{ij}, InvisibleColl_{ij}, \\ Betweenness_{ij}, Closeness_{ij}, \Pr od_{ij} \end{pmatrix} \quad \textbf{(2.4)}$$

Nonetheless, we must recall that our units of observations are all the possible pairs of authors. This means that we can expect our dependent variable to be equal to 0 many times. Therefore, we need to think about a method of estimation which allows us to take into account the high number of zero counts. In fact, the number of zero-counts is greater than one would expect for a Poisson distribution: 99.6% of all possible pairs of authors in our dataset did not publish any article together in the Top Journals we are considering (*e.g.*: they are not a "real" pair). Hence, we opted for using an extension of the Poisson model: typically the so-called *Zero-Inflated Poisson Regression (henceforth,* ZIP), in which the response variable needs to be a count variable and its variance needs to be relatively close to its mean[20]; and each unit has the same length of observation in time (Greene, 1994). Basically, this type of regression consists of two separate models that are then combined together: a *Logit* model is generated for "certain zero" observations in order to understand whether each observation is part of this group or not, and then a *Poisson* model is estimated for the prediction of the counts of those observations that are certainly non-zero.

Therefore, we cannot use an OLS estimation procedure, and we need to rely on a Poisson regression, which is estimated by means of maximum likelihood estimation techniques. The independent variable $PairArt_{ij}$ can take values

$$PairArt_{ij} \approx \begin{cases} 0 & \text{with probability } p \\ Poisson\left(\mu_{ij}\right) & \text{with probability } 1\text{-}p \end{cases}$$

Hence, in our model the number of article any possible pair *ij* has written together has a Poisson distribution with a conditional mean (μ_{ij}) that is a function of the independent variables.

$$\mu_{ij} = \exp\begin{pmatrix} \alpha + \beta_1 M_i - M_j + \beta_2 F_i - F_j + \beta_3 SciAge_{ij} + \beta_4 SciGeneration_{ij} + \beta_5 GeoDist_{ij} + \\ + \beta_6 Co-auth_{ij} + \beta_7 StabCo-auth_{ij} + \beta_8 OutCommunity_{ij} + \beta_9 InvisibleColl_{ij} + \\ + \beta_{10} Betweenness_{ij} + \beta_{11} Closeness_{ij} + \beta_{12} \Pr od_{ij} \end{pmatrix} \qquad \textbf{(2.5)}$$

Finally, in order to test whether our model choice was correct against the use of a Standard Poisson regression, we run a Vuong test (based on the Vuong's statistic (Vuong, 1989)), which allows us to reject the hypothesis that a Standard Poisson is preferable to the ZIP.

[20] As we could see in table 2.6, which reports also the descriptive statistics of our response variable, this is exactly the case. This is also the reason why we could not use a *Zero-Inflated Negative Binomial Regression* (Lambert, 1992): this method requires, in fact, excessive number of zeros *but also* overdispersion (typically, the conditional variance needs to be significantly larger than the conditional mean).

Table 2.7 reports the estimates of six different models we run (A to E). In order to get to model E, which we consider the final model, we first tested A and B as base models, by adding some of the explanatory variables according to the criteria described in section 2.4 (*i.e.*: variables related to gender, relational, and geographical distances). More in details, model A encloses the measures of gender and spatial distance/proximity; model B adds the NA indices proxying the similarity/dissimilarity in terms of strategic role in the macro network of Top geographers. Furthermore, models C and D enrich our estimates with a the measure of production distance/proximity, together with all the measures related to relational similarity/dissimilarity between the pairs generated in our scenario (both in the Poisson and in the Logit component of the equation). Finally, model E tests the role played by scientific experience and scientific generation.

Since we believe that model E is the most exhaustive amongst the others, we now proceed on commenting the results by referring to it exclusively.

Table 2.7 – Empirical Results (Zero-Inflated Poisson Regression)

	POISSON (Dep. Var. Pair Art$_{ij}$)									
	[A]		*[B]*		*[C]*		*[D]*		*[E]*	
	coeff.	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*
Constant	-2.67***	.334	-2.66***	.328	-3.11***	.231	-2.13***	.245	-1.63***	.275
M_i-M_j	.260*	.137	.249*	.136	.232*	.136	.270*	.138	.272*	.139
F_i-F_j	.836**	.252	.913***	.251	.877***	.245	.739***	.253	.748***	.254
Geo Dist$_{ij}$	-.00009	.00006	-.0001*	.00006	-.0001***	.00003	-.0001***	.00003	-.0001***	.00003
Betweenness$_{ij}$	…	…	-2.68	3.184	-5.49	3.76	-6.41	4.03	-6.20	4.01
Closeness$_{ij}$	…	…	-95.2***	14.23	-96.6***	14.86	-106.0***	16.13	-106.2***	16.37
Prod$_{ij}$	…	…	…	…	.068	.046	-.180***	.047	-.169***	.048
Invisible Coll$_{ij}$	…	…	…	…	-.011	.013	-.029**	.015	-.023*	.014
Stab Co-auth$_{ij}$	…	…	…	…	…	…	.161	.188	.172	.188
Sci Age$_{ij}$	…	…	…	…	…	…	…	…	-.007	.023
Sci Generation$_{ij}$	…	…	…	…	…	…	…	…	-.890*	.363
Constant	2.50***	.340	2.003***	.350	.903***	.277	3.28***	.261	3.92***	.333
Geo Dist$_{ij}$	.00005	.00007	.00003	.00006	-.00003	.00004	.00004	.00004	.00003	.00003
Out Community$_{ij}$	…	…	…	…	-1.54***	.274	-.647***	.151	-.829***	.219
Invisible Coll$_{ij}$	…	…	…	…	.171***	.027	.106***	.021	.130***	.026
Co-auth$_{ij}$	…	…	…	…	…	…	-.847***	.069	-.895***	.073
Sci Age$_{ij}$	…	…	…	…	…	…	…	…	.011	.030
Sci Generation$_{ij}$	…	…	…	…	…	…	…	…	-1.37***	.485
Vuong-statistic	1.46*	…	1.33*	…	2.61***	…	6.04***	…	6.04***	…
Log Likelihood	-1881.7	…	-1784.431	…	-1717.933	…	-1612.832	…	-1606.12	…
N	83436	…	83436	…	83436	…	83436	…	83436	…
Non-zero obs	281	…	281	…	281	…	281	…	281	…

*Legend: *** coeff. significant at 1%; ** coeff. significant at 5%; * coeff. significant at 10%.*

First of all, if we look at the measures related to gender, we see that a pair in which both authors are either females or males positively affect the number of predicted published papers, as compared against the reference group (a mixed-gender pair). Additionally, evidence on female-female pair is stronger and more significant than evidence on male-male couple. Hence, we could assume that intra-gender scientific collaboration pays off.

With regard to the role played by spatial distance, as expected, the coefficient of *Geo Dist*$_{ij}$ is highly significant (1%) and negative: the higher is the distance that spatially separate any two authors forming a pair, the fewer the predicted papers published together. Therefore, we could comment that, although new technologies and transport enhancement facilitate high-distance collaboration, face-to-face communication (Cowan and Jonard, 2004) is still an important factor in determining whether or not two authors decide to collaborate. However, the spatial distance does not significantly affect the likelihood of any pair in our sample to be part of the "zero-group", as we can assume from the estimate of *Geo Dist*$_{ij}$ included in the Logit part of model C. Moreover, we must take into account the fact that, when building our dataset and its subsample, we recorded as an author's affiliation the most recent one, irrespectively of whether the same author was affiliated to different Universities in the past. This is relevant in explaining our results, because they could suggest that what matters is not the geographical distance per se, but the *point in time* in which two authors got to meet each other. In fact, it might be that two authors were affiliated to the same University in the past (and therefore they were geographically close to each other), but as of 2007 they were not. Nonetheless, they established a formal or informal scientific relationship that brings them not to collaborate with each other at present, due to the spatial distance. In other words, if they were closer, they would have published together.

As far as the measures of relational distance are concerned, we notice that stability in the choice of co-authors over time does not affect the number of published papers, as shown by the coefficient of *Stab Co-auth*$_{ij}$. On the contrary, the related coefficient of co-authorship (*Co-auth*$_{ij}$), which can tell us about the authors' attitudes towards building relationships within their scientific community, appears to be significant and negative (Logit component). This result suggests us that the more the two authors forming the pair differ in the average number of co-authors per paper they publish, the less likely is the pair itself not to publish any paper jointly, hence to be part of the zero-count group[21].

[21] It is important to recall that coefficients in the Logit part of a Zero-inflated Poisson regression have to be interpreted in reverse as compared to the Poisson component. Hence, a coefficient presenting a negative sign in the

Instead, if we take into account the role that the author's propensity to build relationships outside the sub-field of Economic Geography, the coefficient of *Out Community*$_{ij}$ is highly significant and negative (Logit component): the higher the number of articles the two authors have published in the *same* Top Journal which is not directly related to the field of Economic Geography but to Geography broadly, the more unlikely is that they do not publish together any article in JEG and EC (hence, in their specific scientific community). This result could suggest us that scientific relationships which are extended beyond an author's restricted field of expertise and community might help reinforcing an author's role in his/her own scientific community.

Additionally, the coefficient of *Betweenness*$_{ij}$ is not significant; this result suggests that the dissimilarity in the strategic role played by any two geographers forming a pair in their macro scientific community does not significantly affect a pair's production. On the contrary, the distance in terms of ability to keep the network connected and to facilitate interactions between geographers (*Closeness*$_{ij}$) within it does influence a pair's production: the higher the difference between the two authors the fewer the articles they publish together, hence the intensity of their collaboration[22].

Furthermore, one of the questions we posed ourselves was whether informal collaborations which do not end up being "formalised" in the co-authorship of one or more papers do play a role. The estimate of *Invisible Coll*$_{ij}$ is positive and highly significant (1%) in the Logit component: the higher is the difference between the two authors forming a pair in the engagement within the informal community of pairs (in the form of informal collaborations which are recorded for our purpose in the "Acknowledgments" section of each paper we downloaded to construct our dataset), the higher is their probability of being part of the zero-count group – of not publishing an article together. This result is important in the sense that it would suggest that informal scientific collaborations do matter. On the contrary, it is negative and significant at the 10% confidence interval in the Poisson component. Hence, it is suggested that the smaller is the difference in the size of the network of informal collaborations between any two economic geographers forming a pair, the higher would be the chance of them

---

Logit part of the regression shows that a 1% change in the predictor decreases the chance of belonging to the "zero-observation group", holding all other factors fixed.

[22] In this regard, it has to be reminded here that we are considering the *difference* in the betweenness indices, and the *difference* in the closeness indices between any two pairs in the generated scenario, which is different from taking into account the NA index itself (for further details, please see Section 2.4, point *d.* of this chapter concerning the variables description).

collaborating together, but the lower would be the intensity of their collaboration[23]. In other words, having a strong network of informal collaboration increases the chance that the same collaboration gets formalised, but does not increase the "intensity" of a formal collaboration when already in place.

With reference to the difference in total production between the authors forming a pair, it is interesting to notice that $Prod_{ij}$ is significant at the usual confidence levels (1%), and shows a negative coefficient. This could be interpreted as the fact that the higher is the difference between the number of articles published by the two authors, the fewer the predicted papers published together. This result might suggest that if we consider scientific production as a measure of an author's quality and we could divide our sample in high and low "quality" authors, we could infer that it would be more likely (or better, it would pay off) for authors of the same "quality" (high-high or low-low) to be working together than for authors of "mixed quality" (high-low) [24]. Clearly, this interpretation generates issues related to whether we could consider the total number of an author's publications as a measure of the quality of his/her work, but this would open a completely different topic of discussion.

The results we obtained by adding $Sci\ Generation_{ij}$ could help us shedding light on this phenomenon. The variable is significant in both components of the regression (although the result is stronger in the Logit component than in the Poisson component), and displays a negative coefficient. On the contrary, $Sci\ Age_{ij}$ is not significant in any case. This result is particularly interesting; for example, it might suggest that it could be better for a scientifically old author – in terms of "intensity" of the collaboration (measured as number of published JA with the same co-author), to be collaborating with a scientifically young author. Nonetheless, this relationship does not pay off if the co-authors are too distant in terms of age (*i.e.* "too young and too old"). A possible explanation might lie on the role played by individual incentives: the young and the old might be attracted to work together (because of matching incentives[25]), but, when they actually do it, they realise that the collaboration behaviour is too different, up to diminishing their

[23] This relationship will be investigated further in Chapter 3 of this thesis, where we attempt to build an empirically testable model which can capture more explicitly the potential difference in terms of formal and informal collaborations.

[24] On this point, we should not incur in the error of disregarding individual's incentives and willingness to collaborate. In fact, in order for a collaboration to be established, there needs to be mutual agreement between the two authors. This reasoning links to an important distinction we will make clear in chapter 4, where we distinguish between *two-sided links* and *two-fold links* (Bala and Goyal, 2000). For further details, please refer to Chapter 4, section 4.2 of this thesis.

[25] As an example, the reader could think about the scientific relationship between Ph.D students and their mentors/supervisors.

productivity at the end (relatively to the productivity level they could have achieved with a co-author belonging to the same scientific generation).

Moreover, if we assume that there is a correlation between an author being “high quality” and his/her scientific age, then we could infer that the older the author the higher the number of papers he/she had published. Nevertheless, this interpretation is not endorsable, in the way that we did include in our model *Sci* Age_{ij} with the specific aim of controlling for this possible misinterpretation. Hence, we might infer that being “scientifically old” could also imply that, for example, an experienced author belongs to a certain generation of researchers in his/her field, and could potentially be more exposed to collaboration with other authors belonging to the same/the closest generation. This result could also lead us to the conclusion that there exists a sort of “invisible college” authors belong to, which builds on the University in which authors had studied together, or on the very first Department in which authors had worked together, and so on. The same applies to “scientifically young” authors.

2.6 Conclusion and research agenda

This essay aimed at studying the potential determinants of scientists’ collaboration choices. Rather than assuming as unit of analysis the single researcher, we wanted to look at pairs of researchers, in our case Top economic geographers. In particular, we built a scenario in which we analysed all possible pairs of co-authors, and by matching different orders of characteristics we were able to suggest some mechanisms that could influence the success of a collaboration in terms of its “intensity”, that is the number of published papers in the Top Journals of the scientific field considered. Moreover, we took into account difference/similarity between any two researchers potentially forming a pair, by constructing different orders of distance/proximity between authors, according to gender, experience; geographical and relational factors. We believe this approach needs to be considered as alternative, since the single author’s productivity over time is not considered as a response variable, but rather as a possible explanation for collaborative behaviour.

The results of a Zero-Inflated Poisson model suggested that all the relational explanatory variables we included do play a (positive) role in shaping scientific collaboration decisions and that within-between scientific sub-communities and informal collaborations – together with the “traditional” form of collaboration, namely co-authorship, are crucial determinants of scientific behaviour, as well as spatial distance still affects negatively the likelihood or the success of a long-distance collaboration, despite the availability of new technologies.

Nevertheless, our research is limited to an extremely small scientific community, and further research should think about enlarging the sample, by including additional fields of research, and by extending the time-span. Additionally, no attributional features of the single geographer were accounted for. This was due to the fact that pairs of individuals represented our units of analysis. On this purpose, it would be useful to extend this research to a sort of mirror image analysis that matches the results we obtained in this essay with evidence from an analysis based on the same sample, but that looks at groups of geographers distinguished and categorised by common attributional features. Hence, looking at the tendency that certain groups of economic geographers (*e.g.:* male *vs.* female; scientifically old *vs.* scientifically young) rather than others have towards collaboration. Another aspect related to the above concerns the possibility of better accounting for other than the one-to-one collaborations which we considered in this analysis, and that naturally emerge from the real network of Top geographers, for example, triplets.

Moreover, it would be useful to replicate our analysis by considering two different points in time, so that it could be possible to account for the year in which two co-authors published a paper together, if they do; hence, producing more accurate estimates, and controlling for the specific collaboration pattern over time for every geographer.

Furthermore, it might be interesting to re-build our dependent variable by weighting it according to different criteria. For example, it could be weighted by the total number of issues or the total number of articles each Journal had published in each year; in this way we would be able to normalise the number of articles published by each pair according to the actual slots offered by different Journals in different years. Or again, we could weigh the depend variable using a measure of the quality of the authors forming a pair; this could be applied by taking into account the authors' impact factor index, h-index, or g-index.

With regard to the estimation method, it would be interesting to extend or replicate our analysis using a development of the Tobit regression, such as the Heckman's sample selection model (Heckman, 1979). This model, being slightly more sophisticated than the one used in this study, could help in shedding light on our results, although it better performs with continuous rather than count data.

Finally, we believe that more light should be shed on the influence of an author's productivity on choices of collaboration, by using complementary tools of analysis, such as those which are now typically adopted in the field of Experimental Economics.

Chapter 3

Incentives and Behaviours in the Formation of Scientific Networks: An Empirical Estimate of Jackson and Wolinsky (1996)

3.1 Economics and Network Analysis: a promising but difficult relationship

When Network Analysis (NA) emerged, originated from the initial intuitions of J.L. Moreno back in the ‘30s and ‘40s (Moreno, 1946), and developed in the ‘50s and ‘60s in two distinct but intertwined fields of sociology and anthropology, economists were very suspicious about an approach “which did not explicitly include prices and individual incentives in the analysis”. As it often happens in the history of economics, the interest raised when networks started to be studied adopting the so-called “complex systems approach”[26].

In particular, these disciplines went back to the original mathematical literature dealing with random graph theory (Erdos and Reny, 1959) and introduced in the NA literature the concept of topology. In other words, they defined the architecture of networks, including some static indicators (*i.e.*: degree distribution, clustering coefficient, average path length), and different laws of motions (*i.e.:* random vs. preferential attachment) that allowed a comparison between actual networks and benchmark ones (*i.e.:* random networks, scale free networks, small world networks, regular networks, etc.). Hence, even if they did not “include prices and individual incentives”, they succeeded in building a framework able to describe the structure and evolution of large complex networks, whose statistical properties could easily fit the asymptotic requirements of theoretical models and allow for inference analysis.

In the meantime, by the end of the ‘80s, game theory had become one of the leading approaches in the economic literature to model agents’ behaviour at the micro level and, later, the concept of

[26] On this point, please see Chapter 2, section 2.1 of this thesis; and for a thorough discussion of this issue, see Maggioni and Uberti (2011).

multiple strategic interactions (both in cooperative and non-cooperative games) was included in the realm of economics.

Thus, as we explained more in detail in the introduction of this thesis, Economics was confronted with a double line of research: a micro (mostly theoretical) perspective studying the way in which the strategic behaviour of agents is influenced by, and in turn influences, the relatively simple structure of a "local" network, and a macro (mostly empirical) perspective focussing on the statistical regularities of the network as a whole. In this scenario different economists choose their own research path, while a general synthesis was, and it is still, missing.

However, an attempt to reconcile these two different perspective has been provided by a branch of game theory, which is concerned with network efficiency and stability, with the final view to introduce individual incentives and rationality into the study of networks. Nonetheless, while this (mostly theoretical) literature, which has developed from the middle of the '90s, has been recently organized and extensively surveyed in three recent books (Vega-Redondo, 2007; Jackson, 2009; Goyal, 2009), no systematic review has been written in order to organise the much more fragmented literature on the empirical side[27]. Therefore, the aim of this essay is to build a bridge between these two separated realms by developing an empirically testable version of one (or more) of these micro-founded models.

3.2 Theoretical micro-founded models

Looking for a model and a dataset to be used in conjunction to achieve this aim, we focussed on one of the seminal papers in the micro theoretical literature on networks (Jackson and Wolinsky, 1996), and we decided to further exploit a database that was already built on the co-authorship behaviour of Italian academic economists (Cainelli, Maggioni, Uberti and De Felice, 2012) by enriching it with new fresh data[28]. We opted for the models by Jackson and Wolinsky (*henceforth*, J&W), because they are, to our knowledge, among the few, if not the only, theoretical microeconomic models which explicitly try to address scientific collaboration phenomena. The main innovation brought about the authors consists in the application of their models to specific "allocation mechanism of non-market goods" (J&W:46) in a *cooperative* setting.

In their paper, J&W introduce two distinct models that take into account different value functions which can be defined on the network. More specifically, the authors develop a

[27] A first limited attempt (focussing on the "economic geography of knowledge flows") is provided by Maggioni and Uberti (2011).

[28] See section 3.5 of this chapter.

connection model – where benefits and costs of forming and maintaining links (hence, relationships) with other agents are considered[29] – and a *co-author model*, where links are interpreted as collaborations amongst agents, and the direct benefits and the indirect costs of the collaboration are included in the utility function[30]. Both models have the common feature that agents can only build *two-sided links*[31]. While the focus of their paper is on the trade-off between network pairwise stability and network efficiency, an issue that will drive most of the subsequent contributions (Slikker and van den Nouweland, 2000; Johnson and Gilles, 2000; Bala and Goyal, 2000; Watts, 2001; Jackson and Watts, 2001), what mostly interested us was the trade-off between the *connections model*, which is focussed on the individual's (knowledge) value determined by his/her relations in the network (with benefits deriving both from "direct" and "indirect" partners, although with weights being decreasing with the length of the relational distance), and the *co-author model* in which indirect connections enter the utility function of the agent in a negative way, since they detract time from one's co-author(s). In fact, one of the main limitations shown by J&W models is the fact that they consider the implications of the two models separately; they do not take into account the empirical evidence that both "co-authors" and "connections" matter in the dynamics of cooperative networks (such as scientific ones). As Goyal (2007:212) suggests, a richer model should be developed, so that, for example, the two models could be integrated and different value functions could be assigned to each node, hence, allowing agents' heterogeneity.

The original theoretical models presented in J&W are as follows.

The first is the *connections model* which "models social communication among individuals" (J&W: 49). This model specifically distinguishes between "direct" and "indirect" connections. A node (an economist in our case) benefits both from the knowledge he/she can get from nodes is directly connected to, and from the knowledge accessed from his/her adjacent nodes' connections. The only difference between these two kinds of communication consists in the costs, which are dependent on the distance of the connections. The less a connection is distant, the more the communication is valuable between those two nodes. Therefore, value and costs of communication depend on the distance between any pair of nodes in a network. These features

[29] In the *connections model* links represent relationships.
[30] In the *co-author model* links represent co-authored papers.
[31] *Two-side link formation* implies that the creation of new links between any two node in a network requires mutual agreement between the two parties, while severance of existing links is unilateral. Please note that this concept is different from the concept of *two-flow links* (Bala and Goyal, 2000), which refers to mutual knowledge exchange between two or more nodes in a network when a link is established, independently of who proposes the link to whom.

of the *connections model* imply that "individuals must weigh the benefits of a link against its cost" (J&W, *ibid*).

More formally, if $w_{ii} \geq 0$ stands for the value that agent *i* assigns to himself (in terms of self-perception of his value), $w_{ij} \geq 0$ represents the "intrinsic value" assigned to agent *j* from agent *i*; and c_{ij} represents the cost that agent *i* bears in order to maintain a link (*ij*) with *j*, then we can represent the utility that any agent *i* gets in the network *g* as:

$$u_i(g) = w_{ii} + \sum_{j \neq i} \delta^{t_{ij}} w_{ij} - \sum_{j:ij \in g} c_{ij} \qquad \textbf{(3.1)}$$

"where t_{ij} is the number of links in the shortest path between i and j [32], $0 < \delta < 1$ captures the idea that the value that i derives from being connected to j is proportional to the proximity of j to i. Less distant connections are more valuable than more distant ones, but direct connections are costly" (J&W, *ibid*). Therefore, the *connections model* specifically accounts for positive externalities which are generated from collaborative processes; and since also indirect connections are beneficial, the more the number of indirect links, the higher the value that an individual is able to extract from his collaboration network.

The second is the *co-author model*. The phenomenon of multi-authored scientific papers has been recently investigated from an empirical perspective by economists (see, for example, amongst the first important contributions, Newman, 2001; Goyal et al., 2006; Maggioni and Uberti, 2009), which shows that co-authorship is a constantly increasing trend amongst scientists.

Let us consider a group of authors (researchers) whose main aim is to produce scientific knowledge in the form of published papers. According to the J&W *co-author model*, authors cannot write papers alone, but they can only collaborate with other researchers in doing so. Clearly, this assumption is restrictive, in particular if we consider that in real networks researchers do publish also alone. Nevertheless, it serves the purposes of the model.

Each author is a node in the network of potential co-authors and has a certain amount of time available to spend collaborating with others. As Goyal (2007:208) points out, "starting a new project allows access to the skills of a new partner and this is attractive, but a new project also takes time away from existing projects, which reduces their worth". Therefore, the crucial

---

If agents *i* and *j* are not connected (neither directly nor indirectly), hence there is no path between them, then $t_{ij} = \infty$.

feature of the model is that it intrinsically generates *negative externalities* from the collaborations (that is, links). In fact, adding more links increases the number of projects/papers and possibly the number of co-authors a researcher has, but at the same time, the quantity and the quality of the time they can invest in each project is reduced (providing that authors have time constraint – *i.e.*: a fixed amount of available hours –).

J&W suggests two different versions of the *co-author model*; firstly, it is useful to briefly detail J&W *general model*.

Suppose g is a network of researchers. We define i, j as two different authors in the network, while n_i, n_j are the projects (or papers) author i and j are respectively working on. The utility (or productivity in this specific case) of author i is given by

$$u_i(g) = \sum_{j:ij\in g} w_i(n_i, j, n_j) - c(n_i) \tag{3.2}$$

where $c(n_i)$ represents the cost author i has to bear in order to maintain n_i connections (therefore, projects). Moreover, $w_i(n_i, j, n_j)$ is the utility author i gets when he/she is directly connected with j, and i and j are working on n_i, n_j projects respectively. Hence, one's utility is given by the sum of the utilities he/she gets from his/her connections minus the costs of maintaining them.

However, a more *specific version* of the model is provided by J&W, due to the fact that it better details the phenomenon under investigation.

In a network in which n_i is positive, agent i's utility function is given by

$$u_i(g) = \sum_{j:ij\in g}\left[\frac{1}{n_i} + \frac{1}{n_j} + \frac{1}{n_i n_j}\right] = 1 + \left(1 + \frac{1}{n_i}\right)\sum_{j:ij\in g}\frac{1}{n_j} \tag{3.3}$$

where agent i's utility function is equal to 0 if he/she is not involved in any collaboration (see above). Also in this version of the model, n_i represents i's number of links and n_j represents j's number of links. Given an initial time endowment which is equal to 1, utility depends on the time all researchers involved spend in total on a specific project $\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$ plus what J&W define as *synergy effect*, $\left(\frac{1}{n_i n_j}\right)$, which is a sort of benefit from the joint "production process"

(J&W:56). As the number of project increases, the synergy effect decreases; hence, they are inversely proportional to each other.

Notice that, if we compare this version of the model with the previous one, we can realise that direct costs of connections do not enter the utility function in an explicit way. This is explained by the presence of *negative externalities*: adding new links represents an indirect cost for the authors, which diminishes the synergy effect, therefore the utility they get.

3.3 Towards an empirical version to be tested

As explained above, these two models seem perfect to describe the kind of environment in which academic researchers live and work. Nevertheless, as Carayol et al. (2008:340) remarks, "The (…) structures in these two models are very simple (complete network, empty network, complete star, disconnected pairs) and have little in common with real social or economic networks". Hence, we must integrate the theoretical models by accounting for two limitations that emerged from J&W contribution.

First, an interesting point consists of merging the two models in an encompassing model, since in real life the general creativity (and productivity) of a scientist is definitely related to his/her networks of direct and indirect relations, but the actual performance depends on both his/her co-authorship behaviours and the co-authorship behaviours of his/her co-authors which are competing for a limited amount of "working" time (Cowan and Jonard, 2004); therefore, it does not depend just on the mere number of direct and indirect connections as in J&W original models, but also on the scientific behaviour of such connections.

Second, when trying to implement a testable model based on a theoretical one, the main problem consists of identifying the theoretical hypotheses which may be tested as such and the hypotheses which should be modified in order to allow (for example) for individual heterogeneity, as recently underlined by the results obtained in controlled experiments on network formation, such as Goeree et al. (2009). In order to do so, we must integrate J&W models with other variables which reflect more closely empirical scientific collaborations. In particular, while J&W are concerned with differences between individuals in terms of direct connections (in the *co-author model*) and indirect connections (in the *connections model*) only, we decided to integrate our testable version of the model by introducing two additional aspects which allow us to introduce heterogeneity amongst economists, therefore to reflect more closely the processes of scientific collaboration. First, following somehow the suggestions of the connection model, we account for the scientific production of each Italian Economist at time *t-1*; second, we introduce some attributional variables, whose aim is to grasp individual's characteristics in terms of talent and disciplinary sector, together with "classical" features, such as gender and geographical origin. In doing so, we go beyond the simplicity of the original models and the underlined individual homogeneity implied by the theoretical assumptions (*e.g.*: available time equally split amongst different projects).

Therefore, while Cainelli et al. (2010, 2012) showed that (at least for Italian academic economists in the period 1969-2006) productivity is dependent on ‘attributional’ variables (such as age, gender, academic position, tenure, scientific sub-discipline, geographical location, etc.), ‘relational’ variables (such as propensity to cooperate and the stability of cooperation patterns) and ‘positional’ variables (such as betweenness and closeness centrality indices and clustering coefficients) but with no sound microfounded theoretical model; here we build a testable model based on both the *connections* and the *co-authorship models* of J&W, to test the effects of direct (co-authors) versus indirect (co-authors of co-authors) scientific relations on an individual scientist’s productivity.

The first stage involved the construction of a “generic” model describing productivity of an individual scientist i at time t by “merging” the co-authorship ($co-authorship_i^{t-1}$) and the connections ($connections_i^{t-1}$) models in a single equation, and integrating it with two orders of control variables, regarding attributional (AV_i^t, AV_i^{t-1}) and relational (RV_i^t, RV_i^{t-1}) individuals’ characteristics at time *t* and *t-1*:

$$Prod_{it}(g) = f\left\{co-authorship_i^{t-1}, connections_i^{t-1}, AV_i^t, AV_i^{t-1}, RV_i^t, RV_i^{t-1}\right\} \tag{3.4}$$

Equation (3.4) can be re-written as follows, by explicitly describing the way in which we computed the J&W variables, as it will be better detailed in section 3.4,

$$Prod_{it}(g) = f\left\{\left[1+\left(1+\frac{1}{Prod_{\mathrm{i}}^{\mathrm{t\text{-}1}}}\right)\sum_{j:ij\in g}\frac{1}{Prod_{\mathrm{j}}^{\mathrm{t\text{-}1}}}\right], \frac{\sum_{z:iz\in g} Prod_{\mathrm{z}}^{\mathrm{t\text{-}1}}}{Dist_{\mathrm{iz}}}, AV_i^t, AV_i^{t-1}, RV_i^t, RV_i^{t-1}\right\} \tag{3.5}$$

where *i*, *j*, *z* are different types of individual scientists in the network *g*.

In the equations above, subscript *i* refers to the individual scientist under analysis; *j* indicates his/her co-authors (*i.e.*: any individual who is at geodesic distance equal to 1 from *i* in the co-authorship networks); *z* refers to the co-authors of i’s co-authors (*i.e.:* any individual who is at geodesic distance greater or equal to 2 from *i* in the co-authorship network) .

We thus define $Prod_{it}$ as the number of WoS articles published by individual *i* at time *t* (2007-2010); $Prod_j^{t-1}$ as the number of WoS articles published by *i*’s co-author(s) at time *t-1* (1990-

2006); finally, $Dist_{iz}$ refers to the geodesic distance[33] (greater or equal to 2) linking *i* to his/her indirect connections (*z*). Moreover, RV_i^{t-1} captures an individual economist attitude to collaborate with foreign economists in the past (1990-2006). On the contrary, RV_i^t refers to a measure of individuals' centrality in the present network of collaborations (2000-2007). With regard to AV_i^{t-1}, we consider two different types of variables: one relates to the JAs stored in WoS at time *t-1*; the other relates to a series of variables which account for economists' gender and origin; disciplinary sector, scientific age, and position within the University hierarchy. Finally, AV_i^t represents the amount of other types of publication (books, article in books, working papers, etc.) which are recorded in the Econlit database during the period 2007-2010[34].

The second stage consisted of writing the effective testable version of equation (3.5), starting from the most simple linear version to be tested through an OLS procedure.

$$\begin{aligned}\mathrm{Prod}_{it} &= \alpha + \beta_1\left[1+\left(1+\frac{1}{\mathrm{Prod}_\mathrm{i}^\mathrm{t\text{-}1}}\right)\sum_{j:ij\in g}\frac{1}{\mathrm{Prod}_\mathrm{j}^\mathrm{t\text{-}1}}\right]+\beta_2\frac{\sum_{z:iz\in g}\mathrm{Prod}_\mathrm{z}^\mathrm{t\text{-}1}}{\mathrm{Dist}_\mathrm{iz}}+\\ &+\gamma_1 AV_i^t+\gamma_2 AV_i^{t-1}+\gamma_3 RV_i^t+\gamma_4 RV_i^{t-1}\end{aligned} \qquad \textbf{(3.6)}$$

However – since our dependent variable is the number of scientific papers written by a given economist weighted for a measure of the economist's "quality", we cannot used an OLS estimation procedure, due to the fact that our data presents overdispersion and excessive number of zero counts (since more than 40% of economists who published a WoS paper in the period 1990-2006 have not published a WoS paper in the more recent period). Nonetheless, since the dependent variable has a continuous distribution, we cannot rely on a zero-inflated negative binomial regression (ZINB), because this would be the case of count data.

Hence, we opted for a censored regression model; more precisely, although we reckon that there is no "best" or "perfect" estimation method due to the nature of our data, we are convinced that a *Tobit model* (Tobin, 1958) could provide us with a good estimation of the effects which our analysis focuses on, and that could help us to overcome the distinguished feature of our

[33] In general, in a graph, the (geodesic) distance between two nodes *i* and *j* is "*the length of the shortest path between them*" (Wasserman and Faust, 1994:161).

[34] We added this variable in order to test whether other types of scientific publications, apart from journal articles recorded in WoS, are complements or substitutes in the production of an individual scientist.

dependent variable of being continuous[35], and to show clustering of observations at the 0-constraint.

The Tobit model is a special case of censored regression, which allows us to estimate linear relationships between variables that have been either left- or right-censored in the dependent variable. In our case, we opt for the so-called "censoring from below", since our threshold is represented by observations taking value 0 in the dependent variable, which are censored. More precisely, this model – which is estimated by means of maximum likelihood estimation techniques, relies on the assumption that the independent variable y_{it} is non-negative and is treated as a latent variable (y_{it}^*) for certain values, which therefore cannot be observed: in our case y_{it} is observed only if it is above the defined 0-threshold.

The structural equation of a general Tobit regression is given by

$$y_{it}^* = X_i\beta + \varepsilon_i$$

where and y_{it}^* is the latent (dependent) variable, X_i is a vector of observed independent variables, and ε_i is the error term[36].

Hence, using equation (3.6), the structural equation of our model is given by

$$\mathrm{Prod}_{it}^* = \alpha + \beta_1\left[1+\left(1+\frac{1}{\mathrm{Prod}_i^{t-1}}\right)\sum_{j:ij\in g}\frac{1}{\mathrm{Prod}_j^{t-1}}\right] + \beta_2\frac{\sum_{z:iz\in g}\mathrm{Prod}_z^{t-1}}{\mathrm{Dist}_{iz}} + \\ + \gamma_1 AV_i^t + \gamma_2 AV_i^{t-1} + \gamma_3 RV_i^t + \gamma_4 RV_i^{t-1} + \varepsilon_i \qquad \textbf{(3.7)}$$

where Prod_{it}^* is the latent dependent variable which is observed for values equal to the threshold $\tau = 0$, and censored otherwise.

$$\mathrm{Prod}_{it} = \begin{cases} \mathrm{Prod}_{it}^* & \text{if } \mathrm{Prod}_{it}^* > 0 \\ 0 & \text{if } \mathrm{Prod}_{it}^* = 0 \end{cases} \qquad \textbf{(3.8)}$$

Then, the observed dependent variable Prod_{it} is defined by the above equation.

[35] As it will be discussed in section 3.4, our dependent variable is not constructed as a measure of production itself, but rather as a "quality-adjusted" measure of production (*i.e.*: production is weighted by a proxy of the value of the scientific article and by the single author's contribution).

[36] Note that $\varepsilon_i \approx N(0,\delta^2)$.

In other words, the log-likelihood of the model (which we do not report here[37]), is composed of two parts: one that corresponds to the classical regression for the uncensored observations, and the other that accounts for the relevant probabilities that an observation has been censored.

3.4 Variables Description

In order to test the accuracy of the two J&W models, an econometric analysis was implemented. Before presenting and discussing the empirical results, it is necessary first to describe the variables which were included in the model; second, to explain the way in which we built a bridge between the theoretical model and the empirical data, by presenting the rationale which moved us in constructing the two main variables "co-authors" and "connections", referring to the J&W models.

Our dependent variable (*WoS Prod*) is a "quality adjusted" measure of the scientific production of Italian economists; therefore, we considered the articles published at time *t* in one of the Journals listed in WoS. In particular, we gathered information concerning the Impact Factor (IF)[38] of each WoS Journal each of our Italian economist published in that specific year, together with the overall WoS production of each economist at time *t* and the article authorship. Italian Economists in our dataset published in a total of 72 WoS Journals between 2007-2010.

If we assume that the quality of an economist depends on the number of article he/she delivers, on his/her percentage of contribution to that article, and on the impact of the article in the research community, we need to build a dependent variable which takes into account all these factors. In fact, *WoS Prod*[39] was derived in a two-step process:

- the IF of the Journal in which each article was published was divided by the number of authors who wrote it (as proxy of the economist's contribution, following Moed, 2005);
- the indices derived in the previous point (one for each article published by the same economist) were summed up[40].

[37] For a detailed presentation of the Tobit model, please refer to Tobin (1958) and Wooldridge (2002, 2009) amongst the others.

[38] The Impact Factor of a Journal is defined as $IF = \frac{two-year\_tot\_citations\_t_3}{tot\_art\_t_1+t_2\_A\_journal}$, where *two-year_tot_citations_*t_3 is the amount of citations that the papers published in a generic "A journal" in a two-year period receive in the year following the publication; and *tot_art_*t_1+t_2*_A_journal* is the total number of articles published in such "A journal" in those two years. Although a lively debate on the adequateness of IF as a measure of the quality of publications is still ongoing, we opted for using this index since it is still one of the most commonly used.

[39] Note that 3 years (2007-2010) were assumed to be a period long enough to smooth for idiosyncratic (both positive and negative) shocks (such as pregnancy, illness, sabbaticals) which may impact upon an individual scientific production.

[40] Although the idea of summing up the IFs might seem unconventional to the reader, anecdotal evidence gathered by talking to some Italian senior academic economists has confirmed that this method has already been used as a

Using an example[41], suppose the economist W in year *x*, *y*, *z* published 3 articles in 3 different WoS Journals: Journal A in year *x* has an IF of 1.334; Journal B in year *y* of 1.239; and Journal C in year *z* of 1.558. Additionally, we know that the article published in Journal A was written by 2 authors (including W), the one in Journal B by 3 authors (including W) and the one in Journal C by 4 authors (including W). **Table 3.1** summarises this information.

Table 3.1 – Example: computing the value for the Italian economist W for the construction of the dependent variable *WoS prod*

WoS Journal	*IF in year of publication*	*Authorship*	*IF/ Authorship*
Journal A	*1.334*	*2*	*0.667*
Journal B	*1.239*	*3*	*0.413*
Journal C	*1.558*	*4*	*0.389*
WoS Prod			***1.469***

Therefore, the economist W has a *WoS Prod* value that equals to 1.469. As **table 3.2** reports, overall, the dependent variable has a minimum value of 0 (in the case the Italian economist did not publish any article listed in WoS at time *t*) and a maximum value of 117.9.

Table 3.2 – Descriptive statistics of the dependent variable, *WoS Prod*

Dependent variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
WoS Prod	*967*	*1.178*	*4.205*	*0*	*117.923*

As far as the independent variables are concerned, we can categorise them into three different groups: *J&W*, *attributional*, and *relational* variables.

a. J&W variables

The first group of explanatory variables aims to test the *co-authors* and *connections models* developed by J&W in the above mentioned article. Following step by step the *co-author model*, the variable *Co-authorship* was constructed by assuming that each economist has a limited amount of time (set at 1) he/she can spend writing articles among his/her other academic duties. Therefore, we divided the available time by the number of articles written at *t-1* by each Italian Economist. This ratio was then computed for all of the economist's co-authors (distinguishing their identities[42]), therefore taking into account also the WoS articles that the co-authors wrote

criteria adopted by Selecting Committees for the selection of candidates in Italy. This is also confirmed by the increased tendency to use a 5-year IF, rather than the usual IF.

[41] Example taken from the original dataset.

[42] We considered co-authors distinguishing their identities because of the possibility that if an economist wrote more than one article with the same co-author, for example, we would have duplicated that co-author's relational behaviour, in this way biasing the results since at our eyes he/she would have appeared as a different co-author.

with other economists. The sum of the above ratios and a synergy effect deriving from the collaboration(s) (calculated dividing the amount of time by the product of the scientific production of each pair of economists) gave us the *Co-authorship* value for each of the Italian economists in our sample. More formally,

$$Co-authorship_i = \left(1+\frac{1}{Prod_i^{t-1}}\right)\sum_{j:ij\in g}\frac{1}{Prod_j^{t-1}} \qquad \textbf{(3.8)}$$

In order to test for the *connections model*, we considered the geodesic distances of all economists in our sample, excluding the direct connections (*i.e.*: economists with geodesic distance equal to 1, that is economists' co-authors) in the network structure at time *t-1*. We named the new variable *Connections*, and we built it by computing the ratio between each economist's number of WoS articles at time *t-1* over the geodesic distance that separates (or links) him/her to any other economists in the network. More formally,

$$Connections_i = \frac{\sum_{z:iz\in g}\mathrm{Prod}_z^{t-1}}{\mathrm{Dist}_{iz}} \qquad \textbf{(3.9)}$$

Table 3.3 summarises the descriptive statistics for this set of explanatory variables.

Note that both *Co-authorship* and *Connections* can take value 0 in the cases in which an economist did not write articles with any co-author. In the case of the former, this is due to the fact that the synergy effect is not computable for these economists; and in the latter because if an economist does not have co-authors is isolated within the network and cannot reach any of the other economists (in fact he/she has null values of network centralities).

Table 3.3 – Descriptive statistics of J&W Explanatory Variables

J&W Explanatory Variable	***Obs***	***Mean***	***Std. Dev.***	***Min***	***Max***
Co-authorship	*967*	*1.250*	*1.257*	*0*	*8.287*
Connections	*967*	*159.630*	*322.294*	*0*	*3,266.57*

b. Attributional variables

In this group of explanatory variables we included a set of individual's characteristics which may influence his/her current scientific production. The variable *Female* is a dummy variable which takes value 1 if economist *i* is female, 0 otherwise; as well as *South*, which takes value 1 if economist *i* is affiliated to an University located in the South of Italy, 0 otherwise. Moreover, we added other three dummy variables: *Tenure*, which takes value 1 if economist *i* has a tenured academic position, 0 otherwise; *Faculty of Economics*, which takes value 1 if economist *i* is affiliated to a faculty of Economics, 0 otherwise; and *Scientific Field*, which stands for 6 different dummies which refer to the scientific field economist *i* belongs to, according to MIUR-Cineca database[43]. Finally, the variable *Scientific Age* measures economist *i*'s years of scientific activity (as of 31st December 2006), and it is computed as the difference between 2006 and First Year Entry in Econlit database (since 1969).

Table 3.4 – Descriptive statistics of the Attributional Explanatory Variables

Attributional Explanatory Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Female	*967*	*0.228*	*0.420*	*0*	*1*
South	*967*	*0.126*	*0.332*	*0*	*1*
Scientific Age	*967*	*16.609*	*9.492*	*0*	*38*
Tenure	*967*	*0.659*	*0.474*	*0*	*1*
Economics	*967*	*0.557*	*0.496*	*0*	*1*
Economic Policy	*967*	*0.179*	*0.384*	*0*	*1*
Public Econ	*967*	*0.112*	*0.316*	*0*	*1*
Econ Thought	*967*	*0.023*	*0.152*	*0*	*1*
Econometrics	*967*	*0.053*	*0.225*	*0*	*1*
Applied Econ	*967*	*0.072*	*0.259*	*0*	*1*
Faculty of Econ	*967*	*0.641*	*0.479*	*0*	*1*
WoS Production t-1	*967*	*4.538*	*4.810*	*1*	*45*
Other Publ t	*967*	*0.211*	*0.677*	*0 **	*7*

** The value can be equal to 0 in case the economist has not published any scientific product different from a JA stored in Econlit database.*

We took into account the different publication propensities of different Italian economists, in terms of their past productivity (*i.e.*: the absolute number of WoS article published at time *t-1)*, proxied by the explanatory variable we named *WoS Production t-1*; and of other types of scientific publication apart from Journal articles (and stored in Econlit database at time *t)*. In fact, the production variable *Other Publ t* refers to the sum of economist's publications at time *t* which are not listed as JAs, but as Collective Volume Articles, books, working papers and dissertations. We decided to include this explanatory variable, in order to test whether these types of "alternative" publications are complements or substitutes to the currently dominating publication type, namely the scientific journal article.

In our model, we included these two sets of attributional variables in order to control for individual is salient characteristics and to introduce individuals' heterogeneity, so that it was

[43] Academic economists in Italy may be affiliated to different faculties: Economics, Political Science, Engineering, Law, etc.

possible to isolate factors that could have caused noise in the data, hence producing distorted results. **Table 3.4** reports the descriptive statistics of this group of explanatory variables.

c. Relational variables

The third group of explanatory variables (**Table 3.5**) includes a series of measures whose aim is to grasp the relational structure built around Italian economists. To do so, we included information about the macro structure of the network of economists at time *t* and about their co-authors at time *t-1*.

In particular, for each scientist, we computed the value of his/her betweenness centrality[44] in the co-authorship network at time *t*; and his/her value of clustering coefficient[45]. As compared to the J&W variables described above (which anyway provide us information concerning economists' relational activity), these measures allow us to take into account and exploit the topological structure of the co-authorship network and the specific role played by an individual scientist in it. Thus, betweenness centrality refers to the strategic positioning of a scientist which may act as a bridge between two otherwise separated groups within the same scientific community; while the clustering coefficient measures the cliquishness of an individual scientist's co-authors.

Table 3.5 – Descriptive statistics of the Relational Explanatory Variables

Relational Explanatory Variable	***Obs***	***Mean***	***Std. Dev.***	***Min***	***Max***
Betweenness centrality t	*967*	*0.070*	*0.370*	*0*	*4.438*
Clustering t	*967*	*0.174*	*0.317*	*0*	*1*
Co-auth Stab t-1	*967*	*0.401*	*0.407*	*0*	*1*
Foreign Prop t-1	*967*	*0.148*	*0.297*	*0*	*1*

In addition, we decided to include other two measures which are related to the network, but focus on the characteristics of economists' co-authors. Firstly, we constructed the variable *Foreign Prop t-1*, whose aim is to capture an economist's propensity to co-author with economists outside the Italian scientific community, that is economists who are not included in the MIUR-Cineca database at 31st December 2006 (hence, they are not affiliated to an Italian University)[46]. This variable was computed as the ratio between the number of an economist's foreign co-authors (distinguishing their identities) over the total number of co-authors (distinguishing their identities). Note that *Foreign Prop t-1* assumes values that range between 0

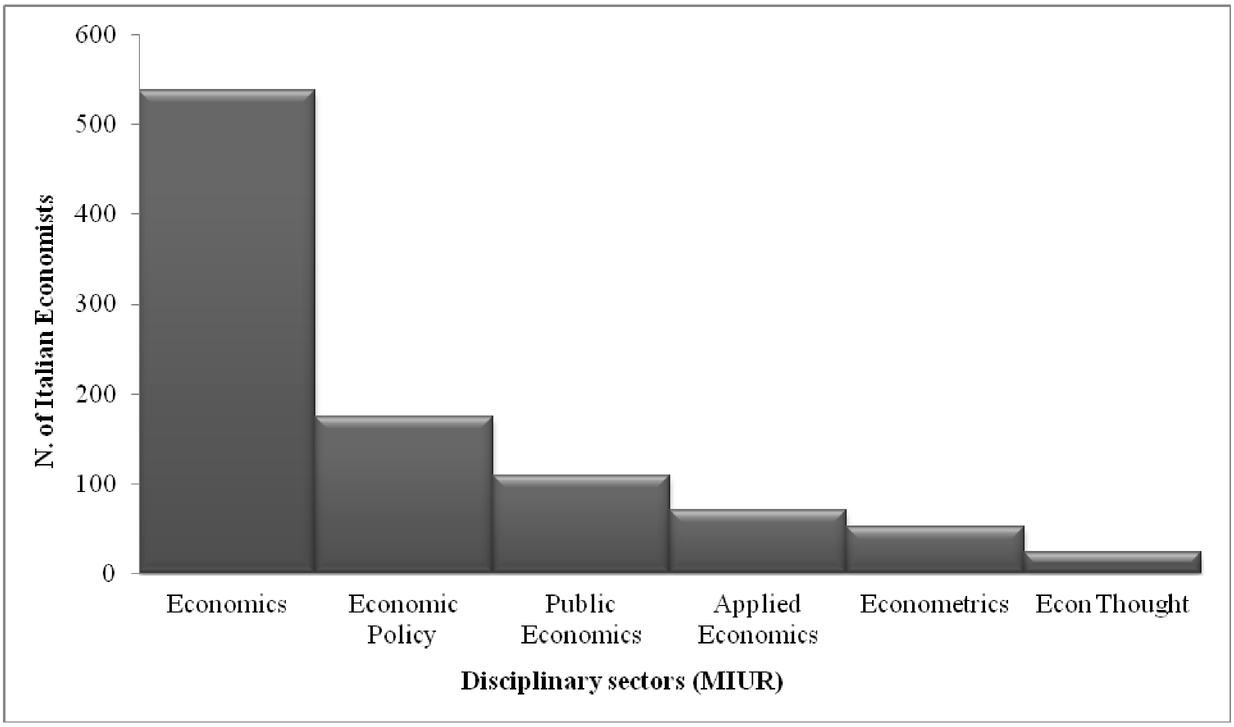
[44] For further information concerning this index of centrality, see Wasserman and Faust (1994) and Scott (2000), and Appendix G of this thesis.
[45] For further information on clustering, please see Chapter 2, section 2.2.
[46] Here we are talking about those economist that we have defined above as "in-network but out-of-sample".

(all co-authors are Italian) and 1 (all co-authors are foreigners). Secondly, we built a variable named *Co-auth Stab t-1*, which accounts for the strength of an economist's scientific collaborations. In other words, following Cainelli et al. (2012), we created an index which measures the extent to which an economist tends to collaborate with the same co-authors over time. It is the ratio between an economist's number of co-authors distinguishing co-authors' identities over the number of an economist's co-authors without taking into consideration their identities. *Co-auth Stab t-1* can take values between 0 (an economist always writes with different co-authors) and 1 (an economist always writes with exactly the same co-authors).

3.5 The data

In order to test equation (3.7), we used an original dataset[47] built by matching three different data sources: the Italian economists population drawn from the official database of the Italian Department for University and Research (Ministero dell'Università e della Ricerca – MIUR in collaboration with Cineca); the Econlit database of the American Economic Association, and Thomson Reuters Web of Science (WoS)[48].

Figure 3.1 – Distribution of Italian economists according to their disciplinary sector (at 31st December 2006 on MIUR-Cineca Database)



[47] Partly used in Cainelli et al. (2010 and 2012) with concern to data extracted from both MIUR-Cineca and Econlit databases regarding scientific production during the period 1990-2006.
[48] Formerly, ISI Web of Science.

Data from the MIUR-Cineca database represents the starting point of the construction of our sample, and refers to a population of 1,620 authors who, at 31st December 2006, held one of the following academic positions: Tenured Full Professor (TFP), Full Professor (FP), Tenured Associate Professor (TAP), Associate Professor (AP), Senior Lecturer (SL) and Lecturer (L). The official MIUR definition of Economics includes six disciplinary groups; in the empirical analysis (**Figure 3.1**) we named these groups as Economics (SECS-P/01), Economic Policy (SECS-P/02), Public Economics (SECS-P/02); Economic Thought (SECS-P/04), Econometrics (SECS-P/05) and Applied Economics (SECS-P/06).

After having identified the population of Italian economists[49], we extracted from Econlit database the information regarding their scientific production, as categorised according to Econlit "product" groups – Journal Articles (JA), Collective Volume Articles (CVA), books (B), working papers (WP) and dissertations (D). Moreover, we extracted information concerning Journal Articles (JA) from WoS.

Then, for the empirical analysis, we divided these products in just three different groups: *Econlit Journal Articles*, *Other Econlit Scientific Products* and *WoS Journal Articles*. These records were downloaded between August 2007 and February 2008, and between October 2011 and April 2012. They were painstakingly corrected for errors in people's names and double entries.

Data have been divided into two different time periods according to the scientific products stored in the above mentioned databases: 1990-2006 *(t-1)* and 2007-2010 *(t)* for a total of 20 years.

In order to extract the sample to use in the empirical analysis, we selected Italian economists who (in the period 1990-2006) have published at least one article in a Journal indexed in WoS. Hence, we obtained a total of 967 Italian economists recording 3,050 Econlit JAs; 1,914 other Econlit publications, and 3,605 WoS JAs in the period 1990-2006. With regard to the period 2007-2010, 338 Italian economists did not publish any article at all; hence, 2,109 Econlit JAs, 1,327 other Econlit publications, and 1,711 WoS JAs were considered (**Table 3.6**).

Table 3.6 – Number of Scientific Products published by the selection of Italian Economists (967) at *t* and *t-1* as stored in Econlit and WoS databases

Publication type	*1990-2006 (t-1)*	*2007-2010 (t)*	*1990-2010*
Econlit JAs	*3,050*	*2,109*	*5,159*
Other Econlit publications	*1,914*	*1,327*	*3,241*
WoS JAs	*3,605*	*1,711*	*5,316*

[49] Please note that 18 out of 967 economists in our sample do not come originally from Italy, but are included in the MIUR database since they are affiliated to one of the Italian Universities.

Nevertheless, in order to derive the attributional and relational characteristics of the population, which are used for the construction of all *(t-1)* variables and consequently of the variables related to the connections and the co-author models, we needed to enlarge our dataset by researching for information regarding scientists (not necessarily economists) who, although part of the considered network, were not included in our sample. This is due to the fact that Italian economists, obviously, do not write Journal articles exclusively in collaboration with other Italians, but they do also with foreign co-authors, whose information were not available to us after the first data download. Therefore, we collected information on scientific production (Econlit and WoS) also about these "out-of-sample but in-network" authors for both time periods.

It has to be remarked here that the construction of this dataset follows an "ego approach", which could also be named as "micro approach"[50]. In fact, starting from a sample of economists who have in common the characteristics of being Italians and being amongst the most "productive" (hence, influent within the Italian economists community) in the period 1990-2006, we decided to track their scientific behaviour in the following period (2007-2010), so that a potential change in the relational attitudes of the economists could have been analysed.

Table 3.7 – Academic Position and Gender (at 31st December 2006 on MIUR-Cineca Database)

Academic Position	*Male (n.)*	*Male (% by rows)*	*Female (n.)*	*Female (% by rows)*	*Total (n.)*	*Total (% by rows)*
TFP	*351*	*36.30*	*45*	*4.65*	*395*	*40.95*
FP	*83*	*8.48*	*23*	*2.38*	*105*	*10.86*
TAP	*107*	*11.07*	*52*	*5.48*	*160*	*16.55*
AP	*90*	*9.41*	*35*	*3.72*	*126*	*13.13*
SL	*43*	*4.45*	*38*	*4.03*	*82*	*8.48*
L	*72*	*7.45*	*25*	*2.59*	*96*	*10.03*
Total *	***746***	***77.15***	***220***	***22.85***	***100***	***100.00***

***** *Totals might be different due to rounding.*

Before proceeding with the description of the variables we built, it is worth spending few words on the composition of our dataset, according to the following criteria: gender and academic position; Athenaeum and faculty[51].

[50] In the existing literature the majority of the sample constructions embraces a "macro approach", as underlined also in the introduction of this thesis. Another example of this antithetic approach is provided by the study presented in Chapter 1 of this thesis.
[51] Information updated at 31st December 2006.

Our sample is composed of 967 Italian Economists, among which just 221 females (22.9%). Moreover (and more interesting), if we consider the gender distribution across all Academic positions, the finding by Cainelli et al. (2012), according to which Italian Academia reflects gender-biased distortion characterising, more generally, the Italian Labour Market, is confirmed.

Figure 3.2 – Italian Economists according to their Faculty of affiliation (at 31st December 2006 on MIUR-Cineca Database)

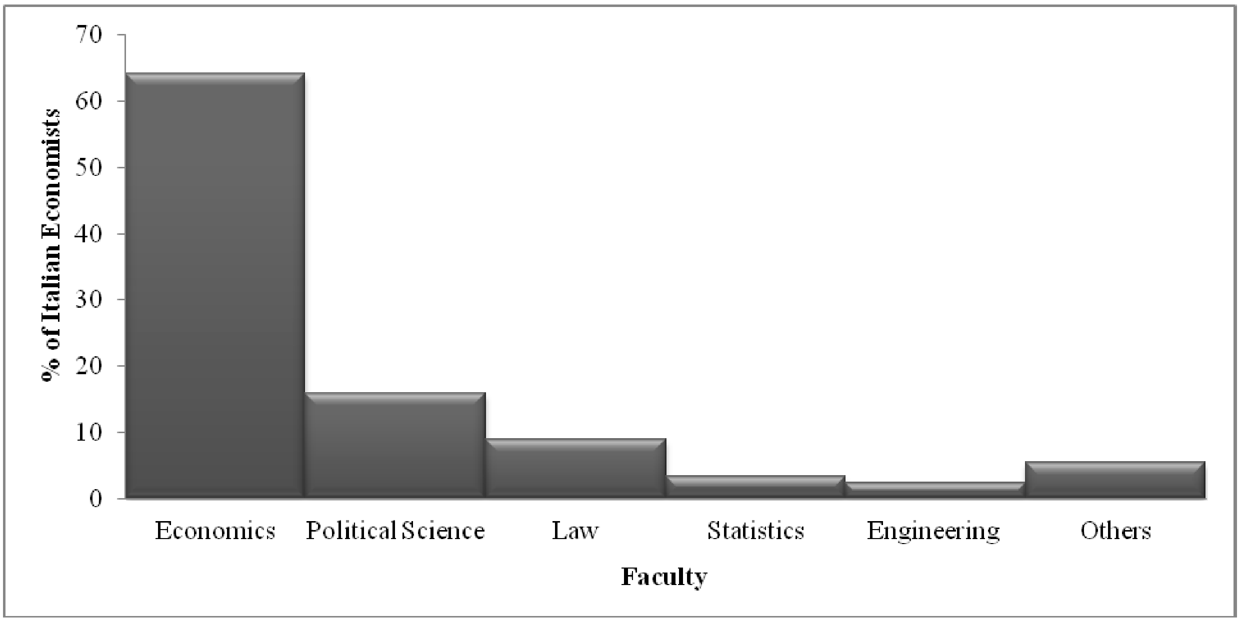


Table 3.7 reports percentages of male and female Italian economists per Academic position. As we can observe, the highest percentage of male economists (36.3%) occupies the top position of Tenured Full Professor, while the majority of female economists is Tenured Associate Professor (5.48%), followed by a 4.03% of women occupying Senior Lecturer roles, whereas only the 4.45% of men is in the same position. Although this result is discouraging, it slightly deviates from the results obtained by Cainelli et al. (*ibid*), in the sense that they found that most of the women occupies the position of Senior Lecturer (a lower position in the Academic ranking as compared to TAP). This difference might be explained by the fact that, despite we built our dataset starting from the database used by Cainelli et al. (*ibid*), we restricted our attention to the Italian economists who have published at least one article in a Journal listed in WoS. In this way, we believed we made a selection of the most talented Italian economists, since it is widely agreed that articles listed in WoS have an higher impact on the scientific community than Econlit ones do.

With concern to the faculty Italian economists belong to, as illustrated in **figure 3.2**, 64% is affiliated to the Faculty of Economics and 15.9% to the faculty of Political Science. The remaining 19.9% is affiliated to diverse Faculties, including Law (9%), Statistics (3%), and Engineering (2%).

Finally, as **table 3.8** shows, the majority of Italian Economists (51.4%) is affiliated to an University located in the North of Italy.

Table 3.8 – Italian Economists and Academic Affiliation according to geographical areas (at 31st December 2006 on MIUR-Cineca Database)

Geographical Area	***Economists (n.)***	***Economists (% by rows)***
North	*498*	*51.40*
Centre	*291*	*30.09*
South and Islands	*178*	*18.51*
Total *	***967***	***100.00***

***** *Totals might be different due to rounding.*

3.6 The empirical results

In this section, the results of a Tobit estimation of the scientific production of Italian economists are presented. As shown in **table 3.9**, we tested 7 different models, starting from a base model (model A) which was further extended to get to a richer model (model I) able to provide a more exhaustive explanation of the phenomena. The estimates of the Tobit regression are presented and followed by information on the truncated observations, together with an ancillary statistic (*sigma*), which is analogous to the square root of the residual variance in an OLS regression.

More in details, table 3.9 presents the estimates for two series of models we decided to run, in order to check the influence of both attributional variables and relational variables. Models A to C refer to the set of attributional variables, while Models D to G are designed for including the set of relational variables.

Table 3.9 – Determinants of Individual economist Scientific Production, 2007-2010 (Tobit regression weighed at WoS Prod ≤ 0)

TOBIT REGRESSION (dep. var. WoS Prod) – weighed at WoS Prod ≤ 0								
	[A]		*[B]*		*[C]*		*[D]*	
	coeff.	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*
Constant	*-235.5****	*52.48*	*-239.8****	*52.6*	*-221.6****	*52.14*	*-198.9****	*51.85*
WoS Production t-1	*.612****	*.071*	*.599****	*.071*	*.557****	*.070*	*.469****	*.073*
Co-authorship	*.421***	*.173*	*.430***	*.175*	*.369***	*.173*	*.245*	*.175*
Connections	*-.003****	*.001*	*-.003****	*.001*	*-.003****	*.001*	*-.002****	*.001*
Betweenness Centr t	…	…	…	…	…	…	…	…
Clustering t	…	…	…	…	…	…	…	…
Foreign Prop t-1	…	…	…	…	…	…	*2.69****	*.657*
Co-auth Stab t-1	…	…	…	…	…	…	…	…
Other Publ t	…	…	…	…	*.762****	*.163*	*.791****	*.162*
Female	*-7.08*	*.472*	*-.038*	*.472*	*-.026*	*.467*	*-.065*	*.463*
Tenure	*-.959***	*.465*	*-.875**	*.465*	*-.724*	*.461*	*-.581*	*.458*
South	*-.965**	*.598*	*-.991**	*.604*	*-1.10**	*.598*	*-1.09**	*.593*
Econ. Policy	…	…	*-.617*	*.532*	*-.666*	*.526*	*-.512*	*.523*
Public Econ	…	…	*-1.98****	*.673*	*-1.86****	*.665*	*-1.74****	*.659*
Econ. Thought	…	…	*-1.35*	*1.34*	*-1.20*	*1.33*	*-.892*	*1.31*
Econometrics	…	…	*.615*	*.836*	*.939*	*.828*	*1.08*	*.822*
Applied Econ	…	…	*-.471*	*.784*	*-.810*	*.783*	*-.721*	*.774*
Scientific Age	*.116****	*.026*	*.119****	*.026*	*.109****	*.026*	*.098****	*.025*
Faculty of Econ	*.409*	*.407*	*.343*	*.407*	*.277*	*.402*	*.225*	*.399*
FIT STATISTICS								
Sigma	*5.39****	*.162*	*5.37****	*.161*	*5.30****	*.159*	*5.25****	*.157*
Left-censored obs	*394*	…	*394*	…	*394*	…	*394*	…
Uncensored obs	*573*	…	*573*	…	*573*	…	*573*	…
Right-censored obs	*0*	…	*0*	…	*0*	…	*0*	…

*Legend: *** coeff. significant at 1%; ** coeff. significant at 5%; * coeff. significant at 10%.*

[Table 3.9 continues on following page]

TOBIT REGRESSION (dep. var. WoS Prod) – weighed at WoS Prod ≤ 0						
	[E]		*[F]*		*[G]*	
	coeff.	*Std. Err.*	*coeff.*	*Std. Err.*	*coeff.*	*Std. Err.*
Constant	*-216.7****	*52.49*	*-206.3****	*52.21*	*-210.0****	*51.95*
WoS Production t-1	*.564****	*.071*	*.557****	*.070*	*.571****	*.070*
Co-authorship	*.283*	*.200*	*.271*	*.174*	*.334***	*.173*
Connections	*-.003****	*.001*	*-.003****	*.0009*	*-.004****	*.001*
Betweenness Centr t	…	…	…	…	*1.45****	*.496*
Clustering t	…	…	*3.61****	*.577*	…	…
Foreign Prop t-1	…	…	…	…	…	…
Co-auth Stab t-1	*-.494*	*.577*	…	…	…	…
Other Publ t	*.762****	*.163*	*.694****	*.162*	*.706****	*.163*
Female	*-.021*	*.468*	*-.144*	*.468*	*.014*	*.465*
Tenure	*-.733*	*.461*	*-.849**	*.461*	*-.782*	*.459*
South	*-1.08**	*.599*	*-1.25**	*.601*	*-1.05**	*.595*
Econ. Policy	*-.658*	*.527*	*-.729*	*.527*	*-.703*	*.524*
Public Econ	*-1.85****	*.666*	*-1.80****	*.668*	*-1.85****	*.661*
Econ. Thought	*-1.13*	*1.33*	*-1.12*	*1.33*	*-1.19*	*1.32*
Econometrics	*.915*	*.829*	*.920*	*.824*	*.655*	*.831*
Applied Econ	*-.826*	*.784*	*-.876*	*.786*	*-.886*	*.780*
Scientific Age	*.107****	*.026*	*.101****	*.026*	*.103*	*.026*
Faculty of Econ	*.268*	*.403*	*.300*	*.403*	*.313*	*.400*
FIT STATISTICS						
Sigma	*5.30****	*.159*	*5.26****	*.157*	*5.27****	*.158*
Left-censored obs	*394*	…	*394*	…	*394*	…
Uncensored obs	*573*	…	*573*	…	*573*	…
Right-censored obs	*0*	…	*0*	…	*0*	…

*Legend: *** coeff. significant at 1%; ** coeff. significant at 5%; * coeff. significant at 10%.*

We consider A as our baseline model, in which we include the production variables related to the *co-author* and *connections models* by J&W together with *WoS Production t-1* as a proxy of economists' talent and attitude; and the set of attributional variables. Therefore, all the models presented have these variables in common. As expected, the scientific production in the period 2007-2010 is a function of the individual idiosyncratic characteristics, as shown by the positive and significant coefficient of *WoS Production t-1*.

Most important for the empirical test of J&W models are the coefficient of *Co-authorship* and *Connections*. The first refers to the availability of time, capacity and mental resources of the individual scientists' co-authors. The coefficient is positive and significant; thus showing that writing papers with high-talented co-authors (and taking into account the fact that they will divide their time and attention between their respective co-authors) will improve the economist's scientific production, according to Jackson and Wolinsky's predictions related to the *co-author model*.

A negative and significant coefficient is registered for the *Connections* variable, which relates (along the lines of the *connections model*) to the scientific productivity of the co-authors of co-authors weighted by their relational distance to the individual economist. This result seems at odds with the *connections model* (which postulates a positive effect of indirect relations in a network) and, indirectly, adds further support to the *co-author model* (the co-authors of my co-authors compete with me over the time of my co-authors) and confirms the results of Cainelli et al. (2010 and 2012). Nonetheless, it is important to take into account the fact that co-authorship is not the only way in which scientists can (and do) collaborate with each other, and the original version of the connections model does not restrict its areas of application to the sole co-authorship relationship. On the contrary, in our empirical analysis we do restrict our attention on co-authored papers only; this might be one of the reasons why our evidence suggests a negative impact of the connections variable on scientific production. Further developments of this work should also consider alternative scientific relationships (*e.g.*: acknowledgments), that could possibly show a positive impact on production, thus confirming Jackson and Wolinsky's theoretical predictions.

With regard to the coefficient of *Scientific Age*, it presents a positive and significant coefficient; as expected, economists who started their careers earlier are more likely to have contributed to the current relevant scientific literature (scientific journal articles in the period 2007-2010) due to their longer experience (relevance of "learning by doing" dynamics), and, possibly, higher reputation and wider scientific network. Moreover, being affiliated to the Faculty of Economics does not seem to be an advantage for Italian economists; the coefficient of *Faculty of Econ* is not statistically significant at the usual confidence intervals. Indeed, the chances of writing WoS articles in period *t* are not increased by being affiliated to a Faculty of Economics.

If we look at the relationship between belonging to a certain disciplinary sector and an Italian economist's productivity (introduced in model B), any of the effects but that of *Public Economics* (negative and significant) appear to be statistically different as compared with the impact deriving from belonging to the disciplinary sector of *Economics*, which we considered as our base group; in other words, being in *Public Economics* actually decreases production at time t as compared to *Economics*. Somehow unexpectedly, as underlined by the estimates reported for model E, being a woman, and holding a tenured position do not have a significant impact on the ability of an economist to publish high-quality scientific articles. On the contrary, being affiliated to an Italian university located in the South of Italy decreases an individual's chance of publishing relevant pieces of research[52].

Furthermore, model C introduces the role played by "alternative" scientific production, by enriching the base model with *Other Publ t*, which specifically addresses the role that scientific production stored in Econlit database in the period 2007-2010 plays on WoS JAs production. The coefficient of *Other Publ t* is positive and statistically significant: writing WP, CVA and B increases the number of relevant JA stored in WoS.

With regard to models D and E, we enrich our base model with additional variables related to more specific aspects of co-authorship. More in details, model D is concerned with the propensity of Italian economists to co-author with foreign scientists. In fact, if the variable *Foreign Prop t-1* might be interpreted as a proxy of the level of scientific internationalisation of the Economics community in Italy, it can also be construed as a specific cost that Italian economists have to bear in order to compete internationally; hence, increasing their chance of being published in relevant Journals. Costs are present in both models developed by J&W. In the *connections model* they are explicitly modelled as costs that the individual agent has to bear in order to establish and maintain direct links within the network; whereas in the *co-author model* they are indirectly represented by the fact that any "new link decreases the strength of the interaction term with existing links" (J&W, 1996:56). Since it is very difficult to find a way to measure the cost that each individual economist is bearing when deciding to activate a co-authorship decision in favour of a foreign scientist as compared to the one related to "intra-national" co-authorship, *Foreign Prop t-1* was used as a proxy of this cost, which has already been exploited in previous works (Cainelli et al., 2010 and 2012). The higher a given economist's propensity to co-author with foreign scientists, the less are the costs of "international" co-authorship for that particular individual. In this sense, we expected the coefficient to be positive and significant as shown by the results of the regression.

Moreover, model E adds an indicator of the stability of the co-authorship behaviour of an economist (*i.e.*: the extent to which the economist tends to write with the same economists over time) in the period 1990-2006 (*Co-auth Stab t-1*). The coefficient is not significant at the usual confidence intervals. Thus, we cannot confirm our intuition about the possibility that changing

[52] Different explanation for this "geographical variance" can be put forward. For example, it might well be that working in a Southern Italian University makes more difficult to be part of the relevant scientific community and, therefore, to publish in highly ranked journals; or that the total amount of (private and public) funds available to individual researchers is higher in the North of Italy than in the South; or even that the average "quality" of Southern academics is lower due to the bias of local selection procedures. However, this discussion goes beyond the scope of the present analysis and will not be pursued further.

co-authors, therefore having the opportunity to work with colleagues who have different expertise and knowledge, pays off in terms of scientific production in the future. Further analysis is needed to understand whether the benefits deriving from establishing new scientific relationships clearly overcome the costs in terms of time and knowledge efforts that one has to put in building it; or, on the contrary, having strong and lasting scientific relationships might drop the costs in term of time as well, but does not pay off in terms of scientific production.

Nevertheless, it has to be noted here that introducing in our model the variables *Co-auth Stab t-1* and *Foreign Prop t-1* reduces the significance of one of our main variables of interest, that is, *Co-authorship,* which is now significant just at 16%; this is the reasons why we decided to drop these effects, and not to include them in our final model I.

Finally, with regard to models F and G, we included NA indices of centralities one by one, in order to understand which of them better explains scientific production. As it can be assumed looking at table 3.9, we opted for *betweenness centrality* (hence, for model G), since, in general, it is the index of centrality that an economist (as any other scientist) cannot be easily aware of; and in this way we can exclude an issue of reverse causality[53]. In fact, the other measures of centrality seem to bias other coefficients that showed to be solid in the base model illustrated above. This is due to the fact that measures such as *Degree centrality* (whose effect are not showed in the models presented) and *Clustering t* are specifically related and indirectly included in the variables *Co-authorship* and *Connections*, generating problems of endogeneity. On the contrary, betweenness centrality provides us with a different information concerning the network; that is, concerning the role played by “stars” or “bridges” in the network. Moreover, J&W (1996) identified the so-called *star network* to be one of the efficient network architectures. Hence, the estimate of *betweenness centrality* (positive and significant) in model I goes in the direction of the *connections model,* since this kind of centrality reflects the individual position in the co-authorship network, taking into account both direct and indirect relationships.

3.7 Conclusion and research agenda

This essay has presented a first attempt to build a bridge between the, mostly theoretical, micro literature devoted to the analysis of the incentives and mechanisms of network formation and the,

[53] This is due to the fact that, if economists in the network are aware of those economists who play a strategic role within it, they will attempt to reach them by seeking to collaborate with them. This is the phenomenon which is also known as *“rich-get-richer”* behaviour or *“preferential attachment”* (Barabási, 2002). Moreover, the betweennes index is very difficult to compute for a researcher that is not a bibliomentrician; therefore it is less likely to generate problems of reverse causality (as compared to the other indices of centrality).

mostly empirical, macro literature concerning the structure and dynamics of networks of scientific collaboration and co-authorship.

The distance between the very sketchy and simplified models of "small" network formation depicted in the theoretical literature, the complexity and the large scale dimension of real life networks are still huge.

We are convinced that there are two possible paths to be followed in order to reduce, if not overcome, such distance. The first (along the lines of Goyal et al., 2006) deals with the construction and empirical testing of "micro-founded" models of network structures at the aggregate level by using data on "real" networks along the lines depicted in this essay. The second path – along the lines of Vanin (2002), Callander and Plott (2005), Di Cagno and Sciubba (2010), Togni (2011) – relies on experimental methods, which may allow the researcher to reproduce in a controlled environment (reducing "data noise") the mechanisms underlined networks formation processes (*i.e.*: the role of heterogeneity in determining individuals' knowledge endowment and connection costs), provided that a sufficient number of experimental observations is considered.

A possible intersection of these two streams of research will include the performance of a sufficiently large numbers of experiments so to be able to use inferential statistics to draw some conclusion on how "real life people" build, maintain and, above all, enjoy networks. And in doing so, to let theoretical models and empirical data give mutual feedback to each other by the "experimental bridge".

Chapter 4

Co-authorship Networks:

An Experiment on Network Formation, Efficiency and Stability

This essay is based on a laboratory experiment which was designed and implemented at The University of Nottingham, where the author has spent a period of study and research (August 2010-October 2011) under the supervision of Dr. Alex Possajennikov.

4.1 Introduction

In this chapter, we are concerned with the lack of experimental investigation of networks, since we believe that laboratory experiments can be a useful tools to investigate and push forward both "the theory and the practise" of networks: on one side, they might represent a valid tool to test theoretical models, and on the other, they could support the analysis of empirical findings by isolating factors which often create noise in the data empirically collected.

Integrating game theoretical models, field data and laboratory experiments might give a more exhaustive framework of the phenomena investigated, including networks. Laboratory experiments allow researchers to specifically control for variables, such as costs and technology (Falk and Kosfeld, 2003:2), in order to make *ceteris paribus* comparison and to set up simple interaction structure in small groups (Callander and Plott, 2005:1471), which would not be possible using field data only. As Weibull (2001) underlines, "moving from armchair theorizing to controlled laboratory experiments may be as important a step in the development of economics as it once was for the natural sciences to move from Aristotelian scholastic speculation to modern empirical science".

The literature provides diverse theoretical approaches which can be followed, starting from the first contribution by Myerson (1977); the author is concerned with a cooperative game in which agents have the possibility of either communication or cooperation. In this way, people have the opportunity to form "coalitions" and the value function is assigned based on those coalitions. This represents the main limit of the work of Myerson: the value function is not defined on the network itself, hence it is not possible to isolate the factors that could affect the process of network formation.

On the contrary, Jackson and Wolinsky (1996) propose two models which try to overcome the above mentioned problem, by taking into account different value functions which are defined on the network itself. However, their attention is put on the trade-off between network pairwise stability and network efficiency; an issue that will drive most of the subsequent contributions. More specifically, the authors develop two different models: a *connection model*, where benefits and costs of forming and maintaining links (hence, relationships) with other agents are considered; and a *co-author model*, which is the starting point of the analysis proposed in this essay, where links are interpreted as collaborations amongst agents, and the indirect costs and the direct benefits of the collaboration are included in the utility function. It is important to underline that the models presented by the authors follow a *cooperative approach* to network formation.

Following the contributions by Myerson and by Jackson and Wolinsky, a series of other models have been developed, suggesting different approaches to network formation.

For instance, Slikker and van den Nouweland (2000) and Johnson and Gilles (2000) develop *static* theoretical models, while Watts (2001) and Jackson and Watts (2001) go further by introducing *evolution dynamics*. Moreover, Bala and Goyal (2000) propose an alternative approach to the one by Jackson and Wolinsky, by designing a model which follows a *non-cooperative* approach. This difference in approaches (cooperative *versus* non-cooperative) represents a crucial assumption when studying network formation processes, which might lead to significant different results, due to the fact that in the first case people do need mutual agreement of the counterpart in order for a link to be formed (*two-sided link* formation); whereas, in the second case, people can build links with other agents unilaterally (*one-sided link* formation). For instance, Jackson and Wolinsky do emphasise the trade-off between network efficiency and stability which emerges from their model, while Bala and Goyal do not witness any conflict of this sort. Moreover, in the Jackson and Wolinsky connection model the star network emerges as strongly efficient but not pairwise stable; on the contrary, Bala and Goyal show that, according to their model, the star network is the sole efficient but also strict Nash architecture (given specific parameters).

Based mainly on the theoretical models of Jackson and Wolinksy and Bala and Goyal, some economic experiments have been designed.

As far as the experiments based on the model by Bala and Goyal are concerned, two remarkable studies have to be mentioned; on one hand, the experiment of Callander and Plott (2005) who consider network formation in which only one-way flow links are allowed. On the other hand, Falk and Kosfeld (2003) set up an experiment in which both one-way and two-way flows are taken into account. The main difference between the two approaches is the value assigned to the knowledge that is implicitly shared when two individuals build a link: in the two-way case, both agents reciprocally share their own knowledge, while in the one-way scenario only the agent who proposes the link has access to the partner's knowledge, while the latter does not. It has to be considered as a crucial aspect; Falk and Kosfeld, in fact, obtained significant differences in the comparison between the results obtained in the two settings: in both conditions a Nash network configuration is achieved, but more stable structures are displayed in the two-way flow condition. Additionally, the authors highlight the role that people's "fairness considerations" might play in connection decisions.

Moreover, to Goeree et al. (2009) is attributed the merit of considering agents' heterogeneity in knowledge values.

With regard to the experiments following Jackson and Wolinsky model, one of the first seminal works was designed by Deck and Johnson (2002). The authors test the model proposed by Johnson and Gilles (2000), which integrates the connection model proposed by Jackson and Wolinsky. Deck and Johnson considers a spatial dimension which accounts for a monotonically increment in the connection costs in relation to the spatial distance between any two agents in the network.

Vanin (2002) designs a pilot experiment which apply both the connection and the co-author models by Jackson and Wolinsky. In particular, to our knowledge[54], it is the only attempt available in the literature which studies experimentally the formation of co-author networks and this represents one of the reasons why (as it is explained in the following sections of this chapter) it has been selected as the empirical starting point of our study.

[54] It is compulsory to quote other two contributions which follow the Jackson and Wolinksy model(s): van Dolder and Buskens (2009) and Mantovani et al. (2011). However, the authors' attention is devoted to aspects which are not directly related to ours. van Dolder and Buskens are concerned with the influence that other people's outcomes might play on the single individual's connection decisions; whereas Mantovani et al. with the concept of "myopic" or "farsighted" pairwise stability.

4.2 Concepts and definitions

Before further proceeding, it is necessary to clarify some concepts which will be used in the following sections of this chapter. This section formally defines and clarifies some principles of network formation, such as network stability and efficiency.

The *stability* of a network (or graph) discloses the idea that nodes have the opportunity both to build and to severe links with agents in the network. Since we apply the definition provided by Jackson and Wolinsky (1996) – *henceforth*, J&W –, the creation of new links requires mutual agreement between the parties, while the severance of existing links is unilateral. Therefore, we are considering *two-side link formation*[55] (J&W, 1996).

If we define V as the set of all value functions in a graph g and i as a node (agent) in the graph, $\mathrm{v(g)} = \sum_{\mathrm{i}} \mathrm{u_i(g)}$ represents the sum of individual utilities. As J&W (*ibid*:47) suggests, we can define an allocation rule, Y, which tells us the way in which V is distributed amongst the nodes in the network. Hence, $Y_i(g, v)$ defines player i's payoff with value function v in graph g.

Moreover, we are concerned with a weaker definition of stability: *pairwise stability*[56]*;* nonetheless, according to J&W, this definition is sufficient to provide strong results, by making a selection between the different graphs (*ibid*:48).

It follows that, given V and Y, a network is pairwise stable if the following conditions hold:

$\forall i, j \in g \qquad Y_i(g,v) \geq Y_i(g - ij, v)$ and

$Y_j(g,v) \geq Y_j(g - ij, v)$ **(cond 4a)**

$\forall i, j \notin g \qquad$ if $Y_i(g,v) < Y_i(g + ij, v)$

then $Y_j(g,v) > Y_j(g + ij, v)$ **(cond 4b)**

where $g + ij$ is the graph created by adding a link nodes i and j, while $g - ij$ is the graph created by severing the link between nodes i and j.

---

[55] Despite the following definitions could create confusion, please note that *two-sided links* are different from *two-flow links* (Bala and Goyal, 2000). The latter refers to a mutual exchange of knowledge between two or more agents when a link is formed, independently of who proposes the link to whom.

[56] For alternative measures of stability, please refer to Aumann and Myerson (1988) and Dutta and Mutuswami (1997) amongst the others.

Condition 4a tells us that, if the two agents are directly connected to each other, the payoff each agent gets under the appropriate value function is greater or at least equal to the payoff they would get if they were not directly connected in the network. On the contrary, **condition 4b** implies that, if the two agents are not directly connected at first and if agent i strictly prefers to be connected with j (because he/she can get a higher payoff under the appropriate value function) and j is indifferent between creating a link with i or not, the link will be built (*ibid*).

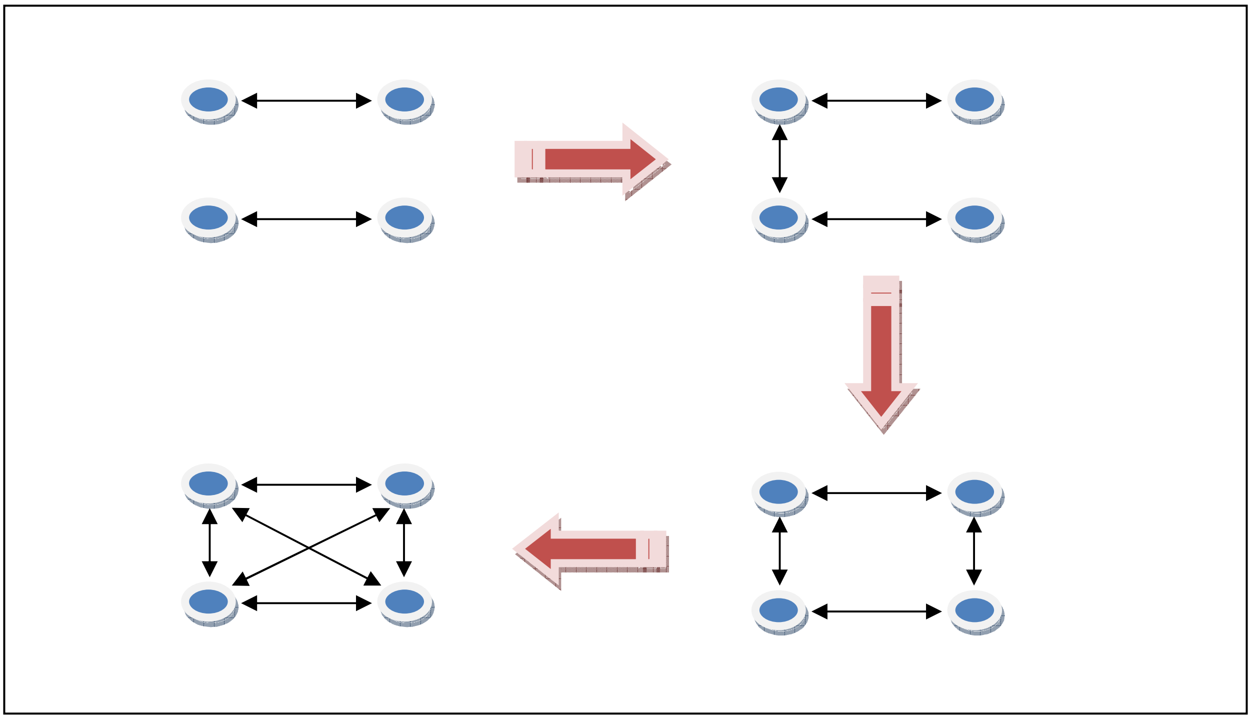
It is important to notice that in a model with two-sided link formation the concept of network stability lies on pairwise incentive compatibility, as opposite to cooperative models in which links are unilaterally built (Bala and Goyal, 2000:1186).

As far as network *efficiency* is concerned, again we adopt the definition of strongly efficient networks by J&W. Suppose that, together with g, we consider a second network g'. We say that g is a strongly efficient network if

$$\forall g' \subset g^N, v(g) \geq v(g') \qquad \textbf{(cond 4c)}$$

Additionally, notice that, in defining efficiency, we are taking into account the network as a whole, “rather than a Paretian notion” (J&W:47).

Figure 4.1 – Trade off efficiency/stability in the Co-author model (N=4)



As it will be further discussed in the next section, J&W found that the strongly efficient network in the co-author model consists of $\frac{N}{2}$ pairs, if N is an even number. Therefore, a strongly efficient network cannot be pairwise stable; suppose we have a network composed of 4 agents, each of whom has built one link, so that 2 pairs are formed.

The two unconnected players would like to establish a link between each other, because this allows them to increase their utility, provided that the other two agents do not form a link. Hence, the network evolves until becoming a line, but in this case the two extreme players would like to be connected to each other and so on. Therefore, the network becomes overconnected from an efficiency point of view. **Figure 4.1** helps us to understand how the trade off between efficiency and stability could emerge in the example provided.

4.3 The Co-author model by Jackson and Wolinsky

This section aims to illustrate in more details the model which will be taken as the theoretical starting point in the experiment described in section 4.4.

As it has been already pointed out (see section 4.2), J&W built the only model which specifically tries to address co-authorship phenomena; this is the reason why we design our experiment considering it as our theoretical pillar. The main innovation brought about the authors consists in the application of their model to specific "allocation mechanism of non-market goods" (J&W:46) in a *cooperative* setting. The authors provide two models which have the common feature that agents can only build two-sided links: the *connection* model and the *co-author* model. For our specific purposes we consider just the latter, because it is focused on explaining the way in which people make connection decisions in real networks, such as scientific co-authorship networks, specialised professional networks, firms organisation networks and so on.

Although this essay is not specifically written to address the empirical issue of scientific networks, it is worth to explain the J&W model by applying their model to networks of researchers who spend their available time on producing papers to be published on scientific Journals[57].

In the previous chapters of this thesis, we have already highlighted the fact that co-authorship is a constantly increasing trend amongst scientists.

[57] This is also the way in which J&W presented their model.

Let us consider a group of authors (researchers) whose main aim is to produce scientific knowledge in the form of published papers. According to the J&W co-author model, authors cannot write papers alone, but they can only collaborate with other researchers in doing so. Clearly, this assumption is restrictive, in particular if we consider that in real networks researchers do publish also alone. Nevertheless, it serves the purposes of the model.

Each author is a node in the network of potential co-authors and has a certain amount of time available to spend collaborating with others. As Goyal (2007:208) points out, "starting a new project allows access to the skills of a new partner and this is attractive, but a new project also takes time away from existing projects, which reduces their worth". Therefore, the crucial feature of the model is that it intrinsically generates *negative externalities* from the collaborations (that is, links). In fact, adding more links increases the number of projects/papers and the number of co-authors a researcher has, but at the same time, the quantity and the quality of the time they can invest in each project is reduced (providing that authors have time constraint – *i.e.*: a fixed amount of available hours).

J&W suggests two different versions of the co-author model; even if in our experiment a more specific version is considered, it is useful to briefly detail J&W *general model*.

Suppose g is a network of researchers. We define i, j as two different authors in the network, while n_i, n_j are the projects (or papers) author i and j are respectively working on. The utility (or productivity in this specific case) of author i is given by

$$u_i(g) = \sum_{j:ij \in g} w_i(n_i, i, n_j) - c(n_i) \tag{4.1}$$

where $c(n_i)$ represents the cost author i has to bear in order to maintain n_i connections (therefore, projects). Moreover, $w_i(n_i, i, n_j)$ is the utility author i gets when he/she is directly connected with j, and i and j are working on n_i, n_j projects respectively. Hence, one's utility is given by the sum of the utilities he/she gets from his/her connections minus the costs of maintaining them.

However, a more *specific version* of the model is provided by J&W, which is used in our experiment due to the fact that it better details the phenomenon under investigation.

In a network in which n_i is positive, agent i's utility function is given by

$$u_i(g) = \sum_{j:ij\in g}\left[\frac{1}{n_i}+\frac{1}{n_j}+\frac{1}{n_i n_j}\right] = 1+\left(1+\frac{1}{n_j}\right)\sum_{j:ij\in g}\frac{1}{n_j} \qquad \textbf{(4.2)}$$

where agent i's utility function is equal to 0 if he/she is not involved in any collaboration (see above). Also in this version of the model, n_i represents i's number of links and n_j represents j's number of links. Given an initial time endowment which is equal to 1, utility depends on the time all researchers involved spend in total on a specific project $\left(\frac{1}{n_i}+\frac{1}{n_j}\right)$ plus what J&W define as *synergy effect*, $\frac{1}{n_i n_j}$, which is a sort of benefit from the "production process" (J&W:56). As the number of project increases, the synergy effect decreases; hence, they are inversely proportional to each other.

Notice that, if we compare this version of the model with the previous one, we can realise that direct costs of connections do not enter the utility function in an explicit way. This is explained by the presence of negative externalities: adding new links represents an indirect cost for the authors, which diminishes the synergy effect, therefore the utility they get.

Recall that J&W's main purpose is to define efficiency and stability of networks when analysed through the co-author model. In particular, they found that if the network is composed by an even number of nodes (N) , the only strongly efficient network is represented by a graph of $\frac{N}{2}$ separate pairs (**Figure 4.2a**). Moreover, "a pairwise stable network can be partitioned into fully intraconnected components, each of which has a different number of members" (Proposition 2, *ibid*:56)[58] (**Figure 4.2b**).

As detailed in section 4.2, efficiency and stability of networks are often substitutes, but could occur following the J&W model. In fact, as J&W predicts, co-authors networks tend to be over-connected with regard to efficiency, because authors cannot fully internalise the negative externalities they endogenously generate in the process of network formation itself. **Figure 4.2** sketches some example of efficient and pairwise stable networks which could emerge from the co-author model.

As Goyal (2007:210) underlines, it is crucial to stress that the negative externalities which are generated by each author do affect his/her own productivity, but at the same time also the productivity of the co-authors. Nonetheless, authors are concerned with their own productivity;

[58] Please refer to J&W (1996:57-58) for complete formal proofs.

hence, they do not take into account co-authors productivity in their connection choices. It follows that “the private costs of additional links are smaller than the social costs” (*ibid*).

Moreover, pairwise stable networks are composed of *unequal* components. In fact, if we look at **figure 4.2b**, although both components are complete (that is to say, all nodes are connected to each other) and not efficient, they display a crucial difference in terms of nodes’ connections opportunities. In particular, more formally, from a Network Analysis perspective, they show an *unequal degree distribution*: the average degree[59] and the average (geodesic) distance between nodes[60] are the main differences between the two components in the graph in **figure 4.2**.

Figure 4.2 – Examples of efficient and pairwise stable networks in the co-author model

Figure 4.2a – Strongly efficient network

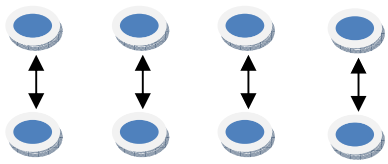
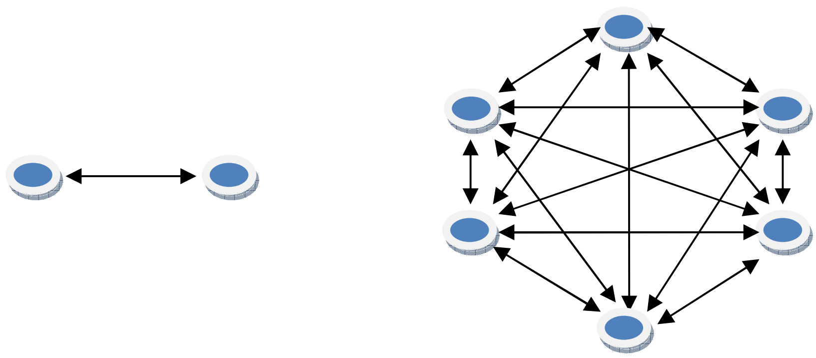


Figure 4.2b – Pairwise stable network (complete components)



In conclusion, it is worth noting that J&W co-author model presents a main limitation, which is implied by the fact that only negative externalities are taken into account, while empirical research has shown that informational asymmetries and externalities, together with knowledge (and expertise) heterogeneity do matter in real networks, including scientific networks for

[59] In general, in a graph, the *degree of a node i* is the “number of nodes adjacent to it” (Wasserman and Faust, 1994:100); therefore the *average degree of a node i* is defined as “a statistic that reports the average degree of the node in the graph” (*ibid*).
[60] In general, in a graph, the *(geodesic) distance between two nodes i* and *j* is “the length of a shortest path between them” (Wasserman and Faust, 1994:161).

instance. Although J&W account for the role of informational externalities in their connection model, the latter aspects are not introduced even though they could explain the emergence of hierarchical architectures in real networks. Therefore, as suggested by Goyal (2007:212), a richer model should be developed in order to integrate J&W co-author model; for example by assigning different value functions to each node, hence, incorporating agents heterogeneity (see section 4.5).

4.4 Efficiency and stability of Co-authorship networks: An experiment

Laboratory experiments might represent a useful tool to study the process of network formation. More specifically, as it has been already emphasised, this work aims to shed a light on the processes which lie behind co-authorship networks. Although there is a lack of literature contributions in this area, some attempts have recently been made (see section 4.1).

In this section, a laboratory experiment on co-authorship network formation will be presented, together with the main results obtained. Even if it represents just the first attempt to address some of the research questions proposed both by theoretical and empirical literatures, it could be considered as a useful starting point for future research.

The pilot experiment was built since its inception following, from a theoretical point of view, the *co-author model* by J&W; while the experimental design and the main results obtained by Vanin (2002) were considered as a benchmark from an empirical perspective. J&W model, as it has already been mentioned in sections 4.2 and 4.3, is the unique attempt to build a model specifically designed to study the co-authorship phenomenon in a *cooperative* framework; this is the main reason why it is used in this experiment as theoretical base. Therefore, despite the co-author model shows some weaknesses, it should be considered as the most suitable in order to study such processes. Moreover, Vanin (2002) designed a pioneer experiment which tries to test the applicability of the co-author model in a laboratory setting.

Section 4.4 is organised as follows. Firstly, the theoretical predictions and the experimental hypotheses which we are going to test are presented (section 4.4.1). Secondly, the experimental design and methodology are illustrated (section 4.4.2), followed by an analysis and a discussion of the main results which emerged from the experimental sessions, in order to isolate some possible behavioural constants (and exceptions) which might characterise and explain the emergence of co-authorship networks.

4.4.1 Theoretical predictions and Experimental hypotheses

The hypotheses we want to test through the laboratory experiment follow a collection of results obtained by different researchers, even if in most of the cases such results were gathered by testing different phenomena, which might be, nonetheless, adapted for our specific purposes. In fact, as often stated in the literature concerning network formation (Callander and Plott, 2005:1471; Kosfeld, 2004:21 and J&W, 1996:45-46 amongst others), the closest contributions are represented by theoretical models and experiments in the area of cooperative games which include communication structures.

As mentioned in section 4.1, one of our main research goals is to understand whether the trade-off between efficiency and stability of network is overcome when we let real people form (either exogenously or endogenously) networks in a controlled environment. Both cooperative (Bala and Goyal, 2000) and non-cooperative (J&W, 1996) models of network formation focus their attention on the evidence that stability and efficiency are in most of the cases substitutes, rather than complements. Moreover, convergence on strongly efficient networks which do not display pairwise stability constitutes one of the classical predictions, most of the time confirmed also by empirical evidence (Vanin, 2002:7-8). Nevertheless, the role of individual and cultural characteristics, their levels of cooperativeness and of inequity aversion are not given the attention they should deserve.

The set of hypotheses (supported by the theoretical predictions) which follows attempts to cover all these issues.

<u>Hypothesis 1</u>: *When negative externalities are not fully internalised by people, the process of network formation leads to inefficient network architectures due to the formation of redundant links. Therefore, over-connected non efficient networks should be observed. However, under the most favourable conditions, people do converge on efficient but not pairwise stable network structures.*

It is necessary to remind that the co-author model we are considering in our analysis implies, by definition, that subjects must deal with *negative* externalities; this could clearly represent a problem for the individuals who could struggle along the process of internalising these externalities which are endogenously generated by them in the dynamic process of adding new links, therefore new collaborations (Vanin, 2002:5). As explained in section 4.3, "(…) indirect connections (…) enter the utility function in a negative way as they detract from one's co-author time" (Jackson and Wolinsky, 1996:56). Hence, if, for example, we apply this assumptions to scientific co-authorship networks, it holds that increasing the number of scientific collaborations with different co-authors reduces researcher's own payoff: from a rational maximisation point of view, this should boost researchers not to build redundant collaborations.

However, it is worth noticing that empirical evidence has shown that efficient networks do emerge *under favourable conditions* (*e.g.*: people are allowed to communicate and all "channels" of communication are open and accessible). For instance, Vanin found that in two groups out of three in the application of the *connection model* and in all the groups in the application of the *co-author model*, individuals do converge upon efficient collaboration network structures, which, on the contrary, are not pairwise stable. This would provide some evidence of the trade-off efficiency/stability, even if the observations gathered by Vanin are not enough to derive conclusions. It is clear that evidence is quite controversial; it is also crucial to consider the fact that in real networks, including scientific co-authorship networks, people do not experience the most favourable conditions when making connection decisions; in fact, it is not always possible to communicate with all other agents and not all channels of communication can be available and/or open.

To sum up, why do people in controlled experiment generate redundant inefficient networks? This question brings us to the following hypothesis, which takes into consideration some of the possible explanations that might lie behind this incongruity.

<u>Hypothesis 2</u>: *Individual characteristics in terms of social preferences (cooperativeness and imitation of others) and fairness attitudes (equity and symmetry) can affect the decision-making process when people build networks.*

Falk and Kosfeld (2003:20-25) provide important evidence on the role that *social preferences* and *symmetry* concerns play in network formation processes. Even if they design a laboratory experiment which apply the Bala and Goyal model (2000), which is not the main focus of this analysis, they show how the differences in the one-way flow and two-way flow treatments can be explained in terms of individual attitudes. In particular, with regard to the role of social preference, they apply the model by Fehr and Schmidt (1999), which takes into account individual's inequity aversion. More specifically, aversion against adverse inequality (envy) in the payoff distribution and profitable inequality (guiltiness) are taken into consideration by introducing two corresponding coefficients in the players' utility function. As Lowenstein et al. (1989) point out, adverse inequality is stronger than profitable inequality; therefore, we expect individuals to care about others' payoff, particularly when others' utility is higher than his or hers. Clearly, this represents an aspect that should be considered when studying network formation processes in controlled environments and it could be useful to explain, as we will see in section 4.3, why participants to our experiment mutually agree to change imposed hierarchical structures in favour of symmetric architectures in terms of payoffs. According to Falk and Kosfeld, asymmetrical networks are not "fairness compatible" (2003:23), leading people to use symmetrical collaboration agreements as a sort of coordination device, regardless of efficiency. However, it is important to remark that fairness motives are not sufficient to explain people's deviations from network efficiency. For instance, the model developed by Charness and Rabin (2002) generated interesting experimental results concerning the trade-off between equity/inequity and efficiency; hence, a model which takes into account both equity concerns and payoff efficiency seems to better explain network formation.

As long as cooperation amongst people is concerned, it is important to underline that the participants of our experiment are asked to play the connection games in group of four individuals. Inevitably, we can expect that different individuals and different groups could opt for different network architectures depending on their level of cooperativeness. Therefore, it is also useful to attempt to relax the level of cooperativeness: as Deck and Johnson (2002) point out, the process of *individual* decision making could lead to inefficient outcomes that perhaps would not be reached through a fully cooperative process of network formation.

Finally, the potential influence that individual can exert on others is an additional factor which might play a role. As Kirchkamp and Nagel (2007) underline by describing their learning model applied to a cooperation network experiment using repeated prisoners' dilemma games, individuals operating in different environments tend to learn from others' experiences and therefore to imitate them. Since in our experiment we allow participants to discuss and agree about the connection decisions, we can expect that imitation and learning from others could play an important role in terms of influencing power.

In light of these observations, another way to look at the role that the cooperativeness level can play on network formation is to vary the degree of cooperation itself, by diminishing people's opportunity to cooperate and by emphasising, on the contrary, the importance of individuals' decisions (Vanin, 2002). The following hypothesis attempts to address this issue.

Hypothesis 3: *When the possibility of severing links without incurring additional costs is introduced, people deviate from building efficient collaboration networks.*

For instance, allowing people to discuss their connection decisions implies, even if indirectly, that people share their intentions and start a sort of bargaining process which can possibly be biased or influenced by others. In fact, one of the crucial assumptions in our experiment, derived by J&W model, is that both parties (in our case, collaborators) have to agree on the collaboration decisions in order for a link to be formed. In this way, we can expect that, in some cases, individual intentions can be blurred by a sort of mutual agreement constraint, which does not come into place in the case of link(s) severance decisions, that instead leave a room open to individuals' genuine intents. In general, putting emphasis on the single individual, rather than on the group of individuals, should provide people with the incentives to deviate towards utility maximising decision; hence, efficient networks should be more difficult to achieve.

Moreover, the only attempt of applying the J&W model made by Vanin follows a framework in which "the social structure changes endogenously for given patterns of interaction (or for a given value function)" (Vanin, 2002:3) that can be defined as a *hybrid approach*. Nevertheless, as Vanin himself observes (*ibid*:12-13), considering both *endogenous* (also in terms of patterns of interaction) and *exogenous* network formations could lead to different results. For instance, Riedl and Ule (2002), following the model by Vega-Redondo (2006), designed a series of repeated prisoners' dilemma games in the context of endogenous network formation. They obtained interesting results in terms of cooperation rates: in the endogenous network treatment they are significantly higher than in the exogenous one. Even if this is not directly related to our analysis, we can infer that, methodologically, the choice between these two different approach might crucially influence our results. It follows the next hypothesis.

Hypothesis 4: *We expect to observe significant differences in terms of efficiency when people are asked to endogenously build their own collaboration network compared to when they are asked to modify pre-existing imposed exogenous collaboration networks. Different imposed network architectures should boost people to make different connection decisions and in some cases to deviate from considering the utility maximising options.*

In the context of exogenous networks, we could infer that imposing structurally different networks could lead individuals to make significantly different connection decisions, which might generate different network architectures. Hence, we expect people to be influenced by the imposed network structures and to be more prone to deviate from the generation of efficient networks. Additionally, to my knowledge, there is no contribution in the literature which attempts to address the issue. Nevertheless, Bala and Goyal (2000:1197) observe that "convergence occurs *irrespective of* the size of the society or *the initial network*" (italics added).

Their observation contradicts the second part of hypothesis 4, but it could be related to the fact that, as observed in section 4.1, the Bala and Goyal model originated from different assumptions which could not be applied to our experimental results.

Having explained the hypotheses we want to test, we can now proceed by describing the criteria based on which our experiment has been designed.

4.4.2 Design and experimental methodology

In order to test our hypotheses, a controlled laboratory experiment has been designed. We start from the cooperative co-author model by J&W (1996) and we follow some aspects of the experimental implementation of the same model as designed by Vanin (2002).

First of all, it is crucial to remark that our experiment has to be considered as a *pilot experiment*. On one hand, time and budget constraints did not allow the author to run the experiment in a larger scale; on the other hand, since the topic has not been sufficiently investigated in the literature from an empirical point of view, it seemed more useful to test the feasibility of the experiment itself and the reliability of the very first results before further proceeding.

Before starting to describe the experimental procedures, it is useful to detail both the experimental conditions and the features which render our experiment innovative in comparison to the experiments already available in the literature.

With regard to the experimental conditions, we follow the assumptions of the J&W model; in particular, there are no direct costs of connection for the participants, due to the fact that the costs of connection are implicit in the assumption of negative externalities. In fact, since increasing the number of links reduces one’s own payoff, the indirect cost of building redundant collaborations is the decrease in the utility itself; as long as the number of collaborations is augmented, the time an individual can actually dedicate to each collaboration decreases, provided that (as we do) we assume an upper bound on the maximum number of connections (3 in our experiment) and we assign each researcher a determined time endowment (6 hours in our experiment).

Additionally, individuals are allowed to create only two-sided links (J&W, 1996), which imply the need for mutual agreement in order for a connection to be built. In fact, subjects in our experiment can decide to create as many two-person collaborations as possible, but the connection proposal has to be accepted by the recipient. On the contrary, in the game in which it is possible to severe collaborations (see below), the potential decision about deleting a link is unilateral: there is, in fact, no need for mutual agreement. This means that an individual can autonomously decide to severe one or more links, even if the partner does not agree on it. As in the case of connections building, severance of links is not costly.

As far as the innovations of this experiment are concerned, we follow some of the suggestions for future research by Vanin (2002:12-13) together with some adjustments in light of the hypotheses we stated above. Firstly, we have to specify that in his experiment, Vanin stresses the role of cooperation by letting people discuss about connection decisions. We followed the same path, even if the J&W model originally specifies only the two-sided links requirement. However, in contrast to Vanin, in one of our connection games we attempt to relax the cooperativeness amongst the participant by imposing exogenously the network structures, rather than impeding communication. In fact, in order to relax the degree of cooperation, Vanin (*ibid*) suggests two possible alternatives: either using the endogenous network (that is to say, people start from an empty network and they endogenously build their own architecture) as a starting point and then asking people to dynamically change it; or imposing different exogenous network structures and then studying the potential influence on the dynamics of the initial network and on the possible convergence to architectures which display stability and/or efficiency. In our experiment we adopt the latter approach.

Secondly, Vanin suggests to vary the size and the composition of the network. Unfortunately, we are not able to implement this modification in our experiment for the following reasons: on one hand, the availability of participants, and on the other hand, the necessity to introduce agents heterogeneity[61] in values and features which would have created some noise in our results, if not adequately supplemented with an additional game (or treatment) that, for time constraint, was not possible to add.

Finally, it is worth noticing that, in this experiment, when we talk about collaboration or co-authorship we are considering a sort of "willingness to collaborate", and not the process of collaborating itself, which could change some of the interpretations of efficiency and stability which are illustrated later on in the analysis.

4.4.2.1 Experimental parameters

Participants were provided with the following experimental parameters (adapted from Vanin, *ibid*), in order to compute their own payoff per each game they played during the experiment. We define $i =$ 1, 2, 3, 4 as participants; **table 4.1** shows the number of hours each participant is

[61] For a more detailed discussion about the role of agents' heterogeneity, please refer to section 4.5 of this chapter.

allowed to spend on each collaboration depending on the number of links he/she could build during the connection game(s).

Table 4.1– Experimental parameters: available hours per collaborations built

Number of *i* 's collaborations	*i* 's hours spent per collaboration
0 **(alone)**	**0** (do not work with others)
1 **(one collaboration)**	**6** (all your time is for that collaboration)
2 **(two collaborations)**	**3** (half time for each collaboration)
3 **(three collaborations)**	**2** (one third of time for each collaboration)

For example, if player 1 decides to build just one collaboration with player 2 (who in turn accepts his/her proposal), 1 spends all the time endowment on that collaboration. Instead, if he/she decides to create two collaborations with two different players, he/she would equally split the time endowment between the two collaborations.

Table 4.2 – Experimental parameters: Synergy effect values

i 's hours spent on the collaboration	*j* 's hours spent on the collaboration	Synergy effect
6	6	**6**
6	3	**3**
6	2	**2**
3	6	**3**
3	3	**1.5**
3	2	**1**
2	6	**2**
2	3	**1**
2	2	**0.6**

Moreover, according to the *specific* co-author model by J&W, subjects benefit of an additional value, the so-called *synergy effect*, added to the collaboration itself due to the fact that they are working together (in specific terminology, we would use *in co-authorship*). **Table 4.2** presents the specific values of the synergy effect depending on the number of hours the two partners spend on the same collaboration.

Recall from the J&W model that the synergy effect is given by the ratio of the initial time endowment over the product of the number of collaborations the two individuals are involved in. Therefore, for example[62], if agent 1 has only one collaboration with agent 2, but agent 2 has also a second collaboration with a third agent, say 3, both agents 1 and 2 will gain a synergy effect

[62] For a formal presentation of the model, please refer to section 4.3 of this chapter.

which is equal to 3 from the collaboration between each other. Therefore, the utility (or payoff) of agent number 1 from collaborating with agent 2 is given by

$$u_1(g) = \frac{6}{1} + \frac{6}{2} + \frac{6}{1 \times 2} = 6 + 3 + 3 = 12 \qquad \textbf{(4.3)}$$

Real values have been plugged into the original co-author model, following the experimental parameters above. The numerators represent the initial time endowment, the denominators in the first two ratios are the number of collaborations each agent is involved in (in this case, 1 and 2 respectively), while the denominator of the third ratio is just the product of the number of each agent's collaborations. Finally, the third ratio represents the synergy effect.

Moreover, the *total payoff* of agent 1 is computed as the sum of the payoffs he/she receives for each of his/her collaboration. For instance, suppose player 1 equally splits his/her time working with 2 different partners. Therefore, agent 1 allocates 3 hours per each collaboration. 1's first partner is collaborating only with 1; hence, he/she is allocating all the time (6 hours) for the collaboration established with 1. Instead, 1's second partner is involved in other 2 collaborations; hence, he/she can spend just 2 hours of his/her time endowment working with 1. It follows that 1's total payoff is given by the utility from the first collaboration $(= 3 + 6 + 3 = 12)$ plus the utility derived from the second collaboration $(= 3 + 2 + 1 = 6)$, implying a total payoff equal to 18 $(= 12 + 6 = 18)$.

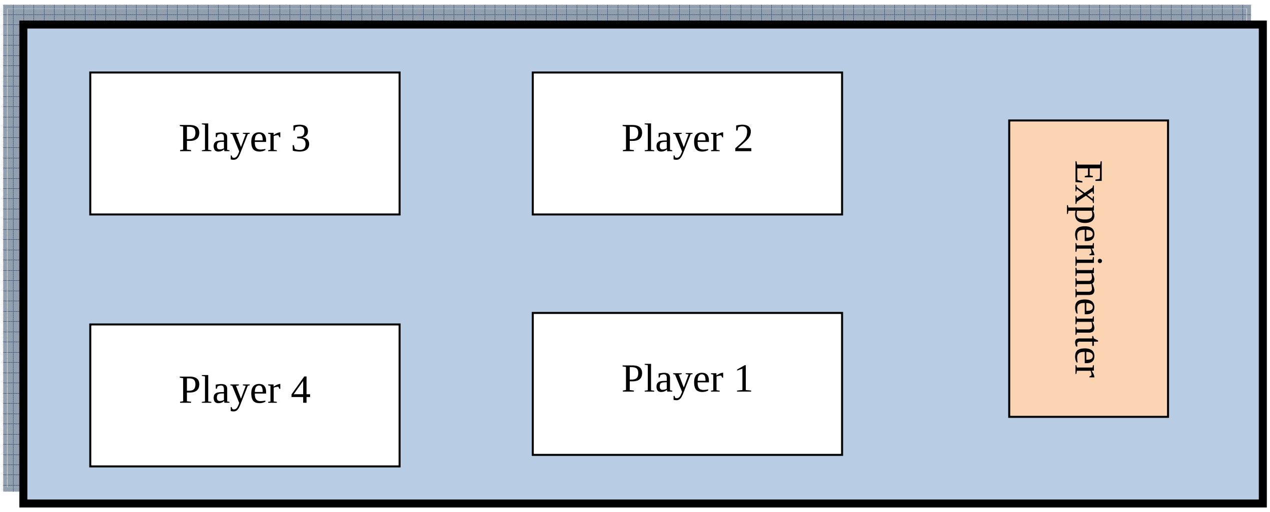
4.4.2.2 Experimental procedure

The experiment was run in Nottingham in July 2011 and a total of 16 people participated. All participants were currently international postgraduate students at The University of Nottingham (UK campus) and enrolled in programs from different academic departments[63]. None of the participants was specifically trained in network analysis and theory, and most of them did not have a background in Economics. Nonetheless, they are all MSc students and therefore highly-skilled people.

The participants were randomly divided in 4 different experimental sessions, each of which composed of a group of 4 participants. Each session lasted approximately 75 minutes.

[63] More specifically, 5 participants from Engineering and Manufacturing department, 4 from the department of Architecture; 3 students from the Business School, 2 from the School of Economics; 1 student from the department of Applied Linguistics and English studies and 1 from the School of Geography.

Figure 4.3 – Room arrangement and participants seats



Before entering the room where the experiment took place, each subject was asked to pick up a number from a box. The selected number was used to identify the single participant for the entire duration of the experiment and to randomly assign a seat to the participants, because of the possible influence of the spatial arrangement (**Figure 4.3**) on the experimental results.

The participants found on their desk the participation agreement form and the initial set of instructions, which were read aloud at the beginning of the experiment, followed by a 3-minute slot given to the participants to read back through them[64]. The set of instructions the subjects received during the experiment was specifically written using a neutral language[65], in order not to bias or influence participants' decision(s).

Two different connection games were implemented in this experiment[66]:

i. *Baseline Connection Game* (henceforth, *BCG*)

It is the exact replication of the experiment on the co-author model made by Vanin (2002). Participants start from an empty network and they are asked to make their connection decision(s), according to the J&W model's utility function and the experimental parameters (see

[64] Please note that a copy of the participation agreement form, the set of instructions used in the experiment, a decision sheet sample and the questionnaires delivered to the participants are provided in Appendices A-F.
[65] For instance, terms like "author(s)", "co-author(s)", "scientific collaboration(s)" and "network(s)" were replaced with terms with a more neutral connotation, such as, respectively: "agent(s)", "partner(s)", "project(s)" and "scenario(s)".
[66] Note that, we could have not called the two games (or cases) "treatments", because theoretically in BCG people *do* start from an *empty* network. Therefore, BCG could have been incorporated in ENCG. However, since we want to compare the results to the ones obtained by Vanin (2002) and to emphasise the difference between network dynamics and endogenous network formation, we opted for separating the two games.

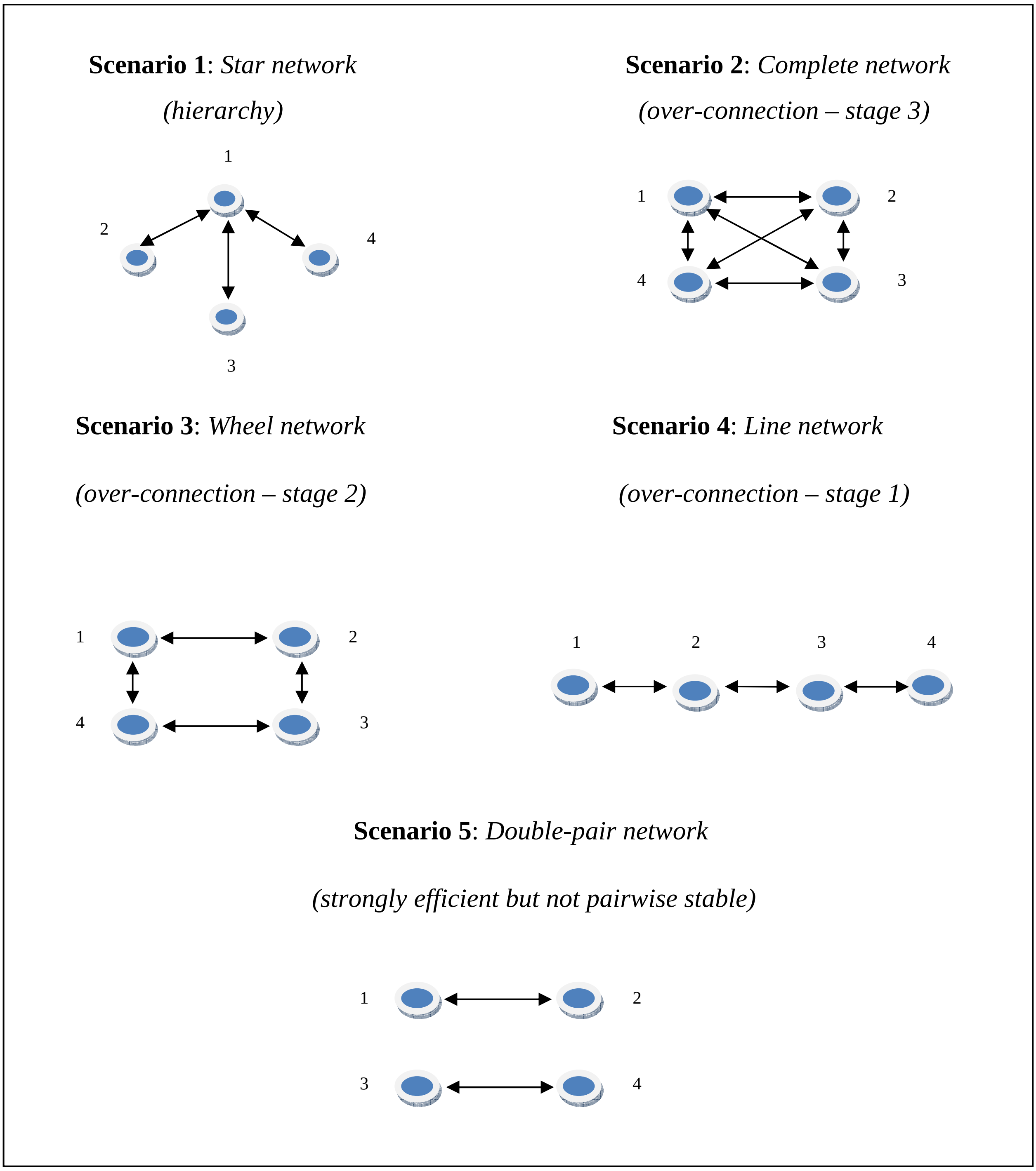
section 4.2.1). Note that the game is presented to the participants as the *connection game number 1*. The maximum number of links (that is, collaborations) per node (that is, participant) is 3. A node can decide not to build any link, therefore incurring no losses and no gains. Moreover, if subjects do not agree upon any network structure, they all receive a 0 payoff. Participants are given 5 minutes to think alone about their connection decision(s), followed by an open discussion with the other participants which lasts 10 minutes. When the time has expired, subjects fill their decision sheet according to the network they have agreed to build. Then, the experimenter collects the decision sheets.

ii. *Exogenous Networks Connection Game* (henceforth, *ENCG*)

This game represents the main innovation compared to Vanin (2002) and, to my records, there is no specific experiment like this in the literature. Note that this game is presented to the participants as *connection game number 2*. Subjects are shown 5 different network structures (exogenously given) one at a time and they are asked to make connection decision(s) including: *a)* do not change any of their pre-existing collaborations; *b)* add one or more collaboration(s) up to a maximum of 3 collaborations, and *c)* severe one or more of the existing collaboration(s). Subjects are given the first network structure and they have firstly 1 minute to think alone about their connection decision(s), followed by 5-minute open discussion[67].

[67] Participants had more time both for thinking alone and for discussing in BCG than in ENCG. This was necessary in order to let participants familiarising with the experiment.

Figure 4.4 – Imposed exogenous networks in the order presented in ENCG



Moreover, as in BCG, if subjects do not agree upon any network structures and they do not explicitly decide to keep the pre-existing network (by re-writing the links as given by the experimenter), they all receive a 0 payoff[68]. When their time is up, subjects are asked to fill their

[68] This solution was implemented instead of assigning the payoff according to the initial imposed network because of the possibility that, in particular when people get tired, they just stop putting effort on the task and consequently

decision sheets according to the network built, the experimenter collects them and delivers the second network architecture. The game ended when the decision sheets of the fifth network were collected by the experimenter.

The five networks shown to the participants were selected according to the basic concepts of NA[69] and following two different criteria: network efficiency and network hierarchy. It is important to specify that each exogenous network gives participants the possibility of reaching the strongly efficient network architecture, given the assigned experimental conditions and parameters. **Figure 4.4** shows the networks and the order in which they were presented to the participants, in order not to suggest subjects the most efficient (but not pairwise stable network structure) which should emerge according to the theoretical model of J&W.

As it is possible to notice, we have adopted the graph representation which is typically used in the field of NA[70]: circles represent nodes (that is, in our case, the experiment participants) together with the identification number, and arrows represent connections (that is, established collaborations). Even though this type of representation is quite specialised, we thought it could have helped the subjects (who were not trained to network studies) to easily understand and interpret the connection scheme. Notice that, since we are taking into consideration the final network (that is to say, what matters is the network the participants eventually agree upon), all links (therefore, graphs) are *non directed*, or, using the traditional terminology *non directional* (Wasserman and Faust, 1994:72). This means that it does not matter, for our purposes, who proposes the collaboration to whom, but just that a collaboration is established between two persons.

If we analyse in more details the five different scenarios, we need to consider the utility each participant would have received in the limit case the experimental group decided not to change the network structures. This can help us to interpret the experimental results later on.

The first scenario represents the most hierarchical network (namely, *star* network) participants could have agreed upon (always considering the experimental parameters and conditions).

leave the network as it is and distort the results. Hence, the 0 payoff would incentivise them to put more effort on the game.

[69] For further information about Network Analysis theory and techniques, see, for example Wasserman and Faust (1994) and Scott (2000). A brief overview of the indices of NA used in this thesis is also included in Appendix G.

[70] See footnote 69.

Subject 1, who plays the role of the star[71], is directly connected to all other participants, who, in turn are not connected to each other but to the star only. By computing the utilities of the subjects applying **equation 4.1**, we can see that player 1 would gain 30 while the utility of players 2, 3 and 4 is equal to 10. Therefore, even if we are not considering the typical measures of centrality used by Network Analysis[72], we can infer the level of hierarchy which this network displays.

The second scenario represents the so-called *complete* network: all participants are directly connected to each other. This type of network has to be considered as the most over-connected one. Subjects could be connected to each other with no need to build redundant links. For instance, if agent 1 establishes a link with agent 2, who is already connected with agent 3, agent 1 would be indirectly connected also with agent 3. Using Network Analysis terminology, there is a path of degree 2 between agents 1 and 3. However, from an utility point of view, the scenario displays symmetry amongst the agents, because they all receive a payoff which is equal to 4.6. Nevertheless, it is not an efficient network clearly; as we will point out below, participants could opt for a payoff-symmetric solution which maximises individuals' utility, given the experimental conditions. However, from a stability point of view, it is a pairwise stable network.

The third scenario represents a *circle* or, using Bala and Goyal (2000) terminology, a *wheel* network. Also this network displays redundant connections, but provides the agents with a higher (symmetric) utility of 15.

The fourth scenario is the *line* network. Even though the first graphical impression would suggest that we are considering a payoff-symmetric network, this is not the case. Agents 2 and 3 have 2 collaborations each, while agents 1 and 2 (the extreme nodes) build just one collaboration each. This implies that the most connected agents receive a payoff of 19.5, while agents 1 and 2 of 12. Moreover, from an efficiency point of view, there exists another possibility for the agents to maximise their utilities, which would imply convergence upon the network shown at last.

Finally, the fifth scenario represents the *double-pair* network, which is the strongly efficient one. In fact, by building only one collaboration with another agent, all participants receive a (symmetric) payoff which is equal to 18. Nonetheless, as all the networks described above, it is

[71] Please note that we imposed player 1 to be the star in all groups. This choice is consistent with the fact that players were randomly assigned an experimental ID number at the beginning of the experiment and this allowed us not to have to randomly re-assign the role of the star per each group.

[72] For more details, please refer to footnote 69.

not pairwise stable[73], because agents 2 and 3 (who are unconnected), for example, would be better off by creating a collaboration between each other, provided that also agents 2 and 4 do not decide to build a connection between each other. Consequently, the network becomes a line (as shown in scenario 4), but then the extreme nodes (2 and 4) would be better off to connect to each other, creating a wheel (scenario 3) in doing so.

We can notice that, just by looking at the exogenously imposed networks, the trade-off between stability and efficiency clearly emerges.

The two connection games described above define the two different parts in which the experiment was divided, each one associated with specific additional instructions which were again read aloud by the experimenter at the beginning of each experimental game.

A *within-subjects design* was implemented, that is to say, all participants undertook all tasks. This choice is justified by the fact that we believe the decisions taken by the participants in BCG do not substantially affect the ones made in the next game, while if we decided to implement ENCG at first and BCG in the second place we would possibly have had some issues, due to the fact that, even if indirectly, the participants would have acquired experience (through the open discussion in ENCG) about most of the potential network structures they could have converged on.

An issue related to the choice of applying a within-subjects design is the problem of providing the subjects with the correct economical incentives, further worsened by budget constraint. Although this is an issue frequently debated in the literature (Starmer and Sugden, 1991; Cubitt et al., 1998), it is widely recognised that a reliable solution to experiments which ask subject to undertake multiple tasks is to apply a *random lottery incentive scheme* (*ibid*); and we followed this line. In fact, participants were told at the beginning of the experiment that, once the experiment finished, firstly one of the two games would have been randomly selected and then one of the four participants. In case ENCG was selected, before proceeding with the random choice amongst subjects, one of the five scenarios was selected. Then, the participant was paid accordingly to the payoff he/she was able to get in the specific game (and scenario in case of ENCG) to the exchange rate of £ 0.30.

[73] Note that, as explained in section 4.2 of this chapter, the notion of pairwise stability we are adopting is the one proposed by Jackson and Wolinsky (1996:48), even if the authors do not consider (and do not model) any formation process behind the concept of pairwise stability.

After the general instructions were read and the possible questions answered by the experimenter, the participants received a questionnaire which had the purpose to help them to understand the games, and in particular to become confident with the way in which the payoffs were computed. Once all participants gave the correct answers (checked by the experimenter), the BCG was implemented, followed by ENCG.

When both games were played the experiment ended and the experimenter delivered a questionnaire to be filled out by each participant. The questionnaire was designed to collect biographical information and to compute an index of cooperativeness per each participant and each group. In this way it was possible to interpret the results by controlling for individual characteristics and for attitudes towards cooperation[74].

4.4.3 Results and discussion

We can now proceed by firstly describing and then analysing and interpreting the results obtained from the experiment per each game. Finally, we will make a comparison between our implemented BCG and the results obtained by Vanin (2002), together with a comparison between our BCG and ENCG. In conclusion, we will attempt to interpret the results also in light of the personal and cultural characteristics of the participants and their displayed level of cooperativeness by creating an index of cooperativeness extracted by the questionnaire delivered at the end of the experiment.

4.4.3.1 Results: Baseline Connection Game (BCG)

Not surprisingly, the 4 groups opted for the same network solution; in fact, they all agreed and converged upon the strongly efficient network structure, that is, they created 2 pairs[75]. **Figure 4.5** shows the exact connection choices made by the participants per each group.

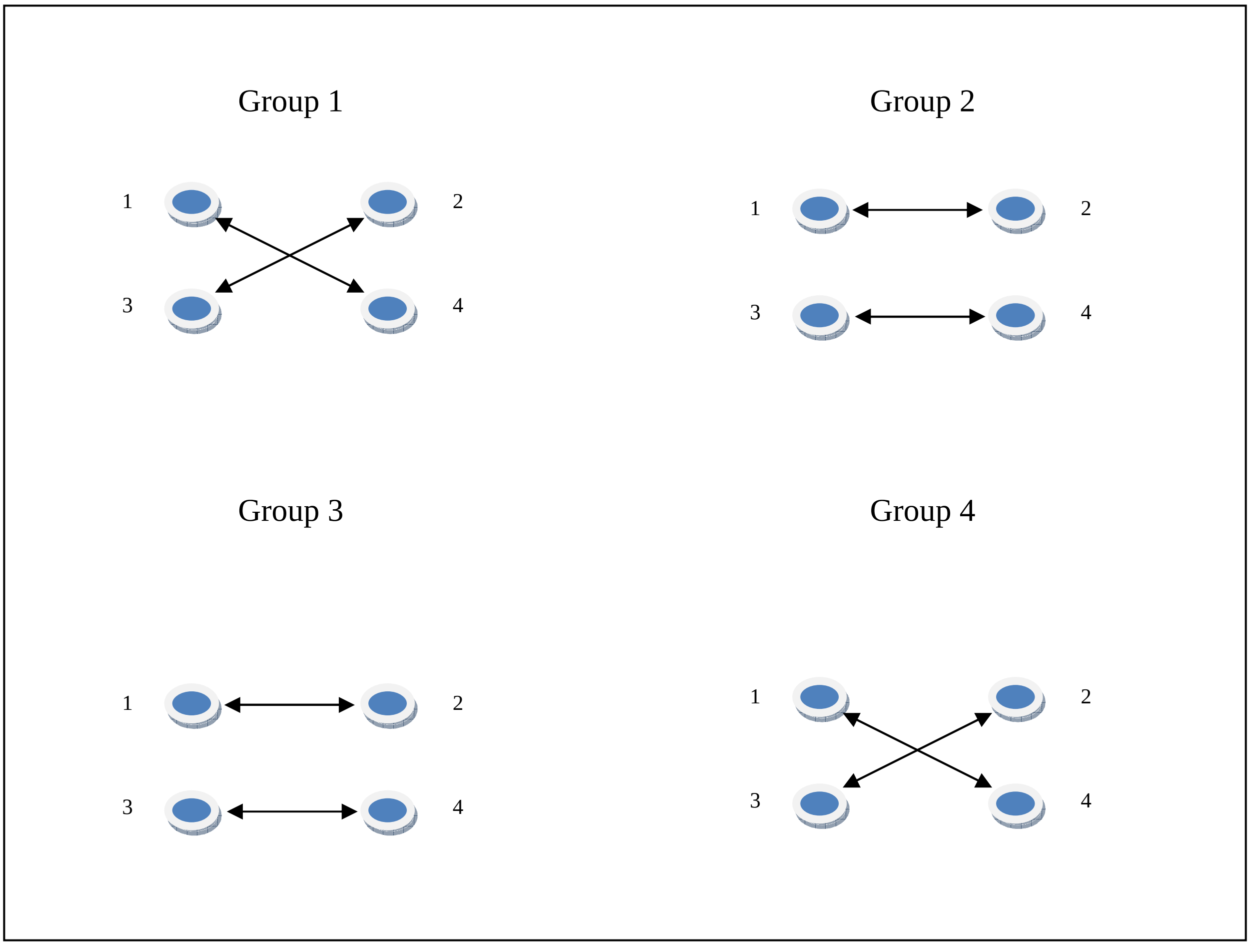
According to the predictions and the results obtained by the analogous experiment made by Vanin (*ibid*:11-12), the opportunity of discussion amongst participants boosted the level of cooperation up to completely internalise the externalities which, by definition, enter the J&W utility function in a negative way. Listening to participants' discussions made the experimenter realise that, despite an initial confusion, the role played by (depending on the group) one or two

[74] See paragraph 4.3.3 for further information about the structure of the questionnaire, which can be found in appendix F.

[75] Please note that, in both connection games, it does not matter for our purposes the partner each participant chooses. Since in our experiment agents are homogenous in values and features, creating a collaboration with subject 1 or 2, for example, does not affect our results.

participants on clearly depicting the possible connections scenarios, was crucial in order for the other subjects to agree on the network to be built.

Figure 4.5 – Achieved networks per each group in BCG



Creating a 2-pair network allowed participants to overcome the efficiency problems that would have been generated by adding new links to the collaboration network.

Nevertheless, if it is true that the groups converged on the strongly efficient network, they did not reach a pairwise stable configuration. As Vanin (*ibid*:11) correctly observes, “forming a new link, while increasing the payoff of the newly connected subjects, would reduce the payoff of the remaining ones”.

However, it is important to specify that alternative solutions to the 2-pair network were considered as possible configuration by the participants; this element, in fact, became more evident in ENCG, where different structures emerged.

All participants received a payoff which is equal to 18, according to the experimental parameters. Hence, according to the implemented random incentive scheme, the randomly selected participant (in case BCG was selected as appropriate game) gained £ 5.40.

4.4.3.2 Results: Exogenous Networks Connection Game (ENCG)

In this game, results are more interesting, in the sense that one group (number 4) deviated from the pattern displayed by the three other groups. Therefore, we might infer that, by exogenously imposing different network configurations and by allowing participants to dynamically change them, other factors came into place and affected the final outcomes.

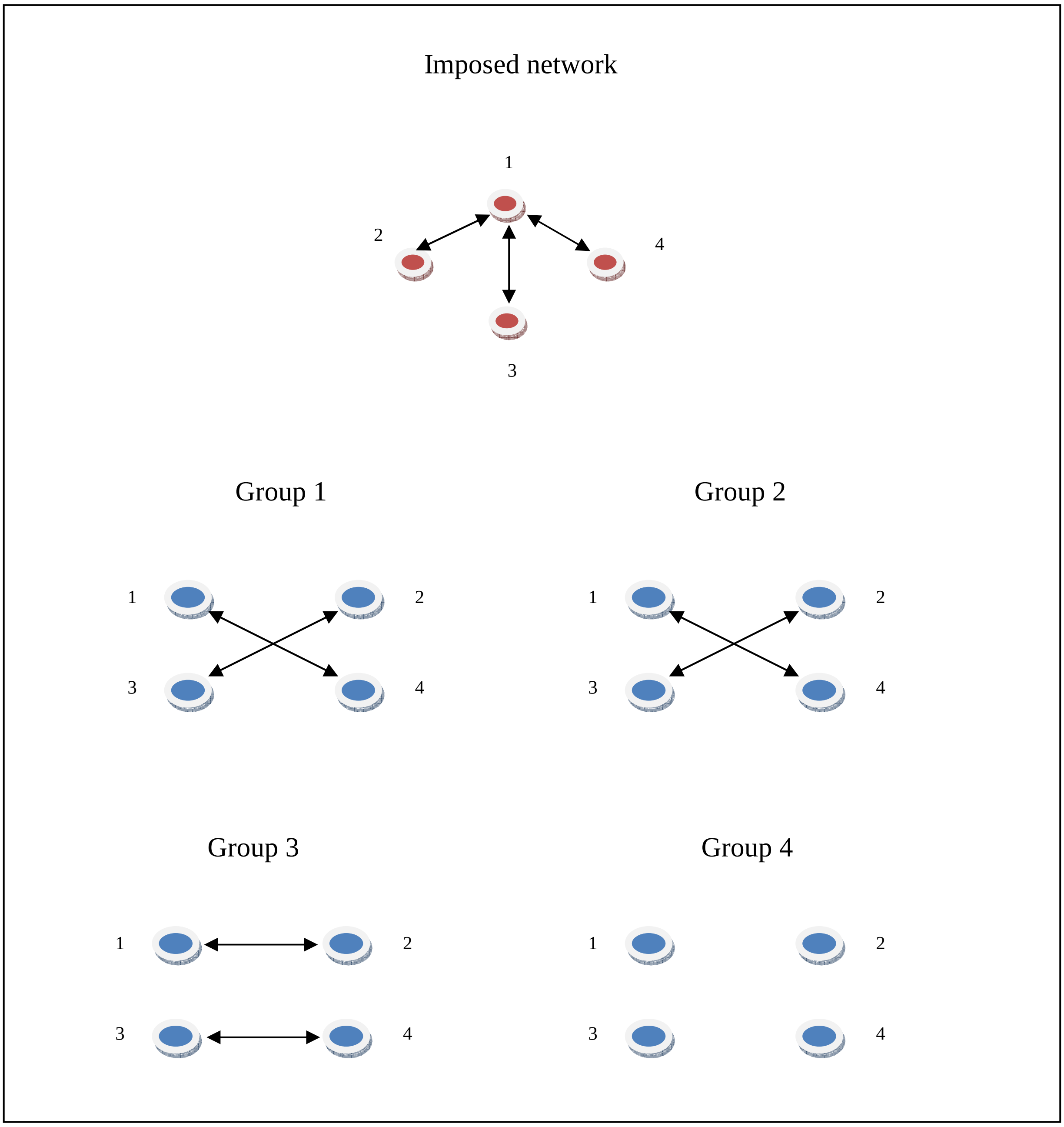
Since we are interested in understanding which variables had really determined the deviations from the strongly efficient network, it is useful to split our analysis in order to isolate the different variables at stake according to: *a)* firstly, the 5 exogenously imposed structures; *b)* secondly, the participants' individual and cultural features and level of cooperativeness.

i. Scenario 1: *Star network*

Figure 4.6 shows the connection decisions made by each group.

Groups 1, 2 and 3 decided to modify the initial hierarchical network structure to converge upon the strongly efficient network. On the contrary, group 4 was not able to reach an agreement at all: neither modifying the network nor keeping the imposed one. It is important to remark that we cannot consider group 4's decisions as an agreement of forming an *empty* network. From the discussion between the participants a complete disagreement about which connections had to be built emerged. Therefore, subjects wanted to form some connections, but the imposed hierarchical structure clearly affected their decision, up to the point that they could not figure out the strongly efficient network as the best architecture, in spite of the fact that in BCG they converged on the double pair network.

Figure 4.6 – Achieved networks per each group in ENCG –scenario 1 (star)



Additionally, it is useful to remember that, even if according to most of the literature contributions[76] the empty network has to be considered as an efficient configuration, we decided not to consider it as such. From a mathematical and network analysis point of view it is, but it is not crucial or it does not add any value when we want to empirically analyse real networks.

[76] On this issue, please refer to, for example, Bala and Goyal (2000) and Goyal (2007) amongst others.

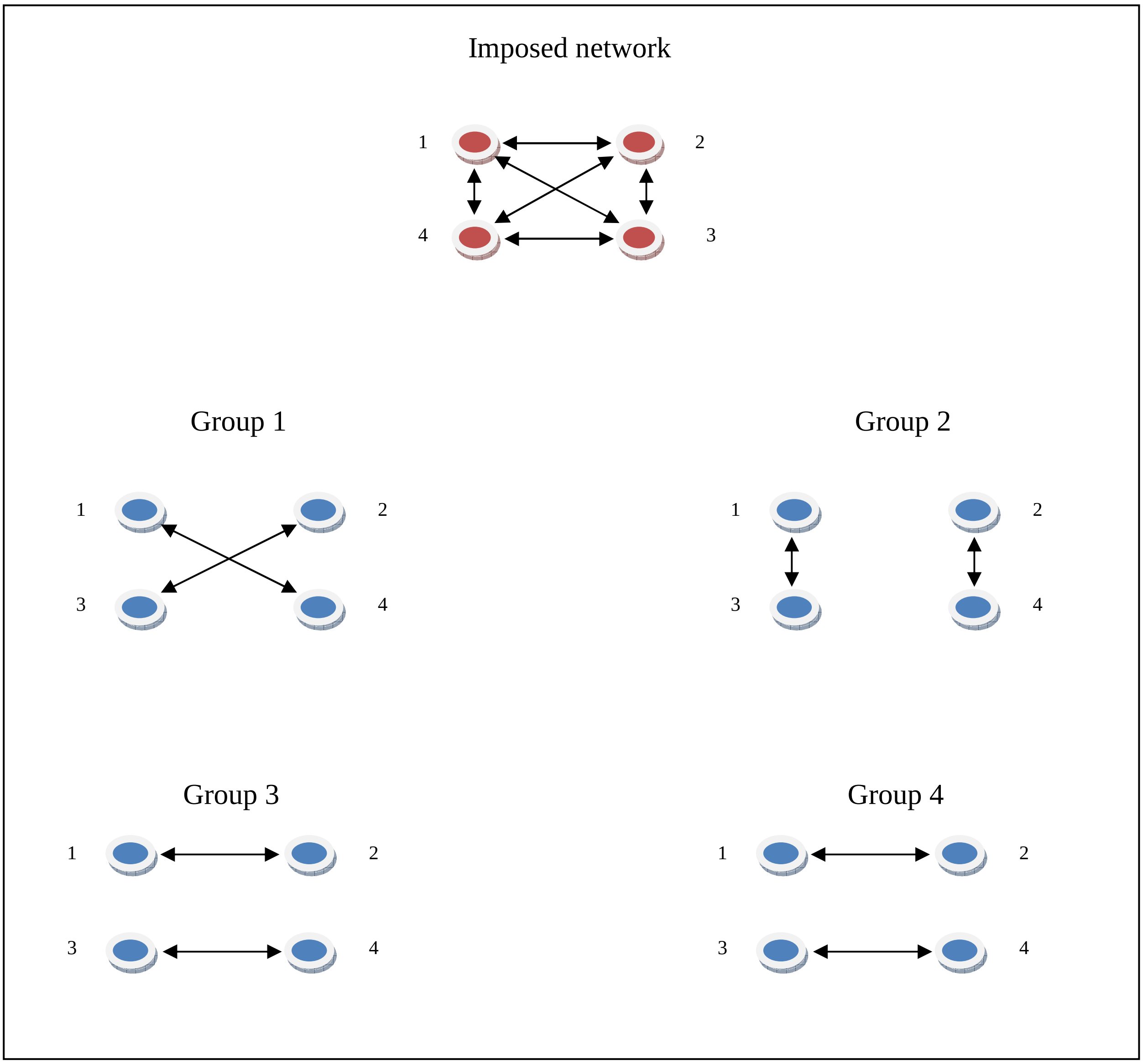
This interpretation was also confirmed by the discussions which occurred in groups 1-3; they were not able to reach at first a common decision on the possibility of converging to efficiency. In particular, group 2 was really in doubt whether to form a network in which everybody had two collaborations each or to stick to the double-pair structure. Only after the discussion they did so. As it will be underlined later on, this behaviour might be explained by looking at the level of cooperativeness of the members of group 2. Furthermore, the star network was the structure that created more disagreement amongst participants. This can be inferred by the audio of the discussions, which is the only element we have. Hence, we should state that creating a hierarchy amongst participants might influence connection decisions; this could be mainly due to the payoff disparity amongst subjects: recall that the star would have earned 30, while the other 3 agents just 10. On the contrary, agreeing on a 2-pair structure gives all participants a (symmetrical) payoff of 18.

ii. Scenario 2: *Complete network*

The complete network did not generate particular problems for the groups, but again for group 2, which struggled during the discussion but eventually converged upon the strongly efficient network. Moreover, we could infer that perhaps in group 2 subjects had an egalitarian first impression of the network, due to the graphical representation of the structure.

On the contrary, it seems quite interesting that group 4 found relatively easy to reach an agreement towards convergence when the complete network was shown. Again, we will attempt to provide some reasons to these behaviour by looking at the personal characteristics of the individuals. **Figure 4.7** represents the final network as agreed in the 4 groups.

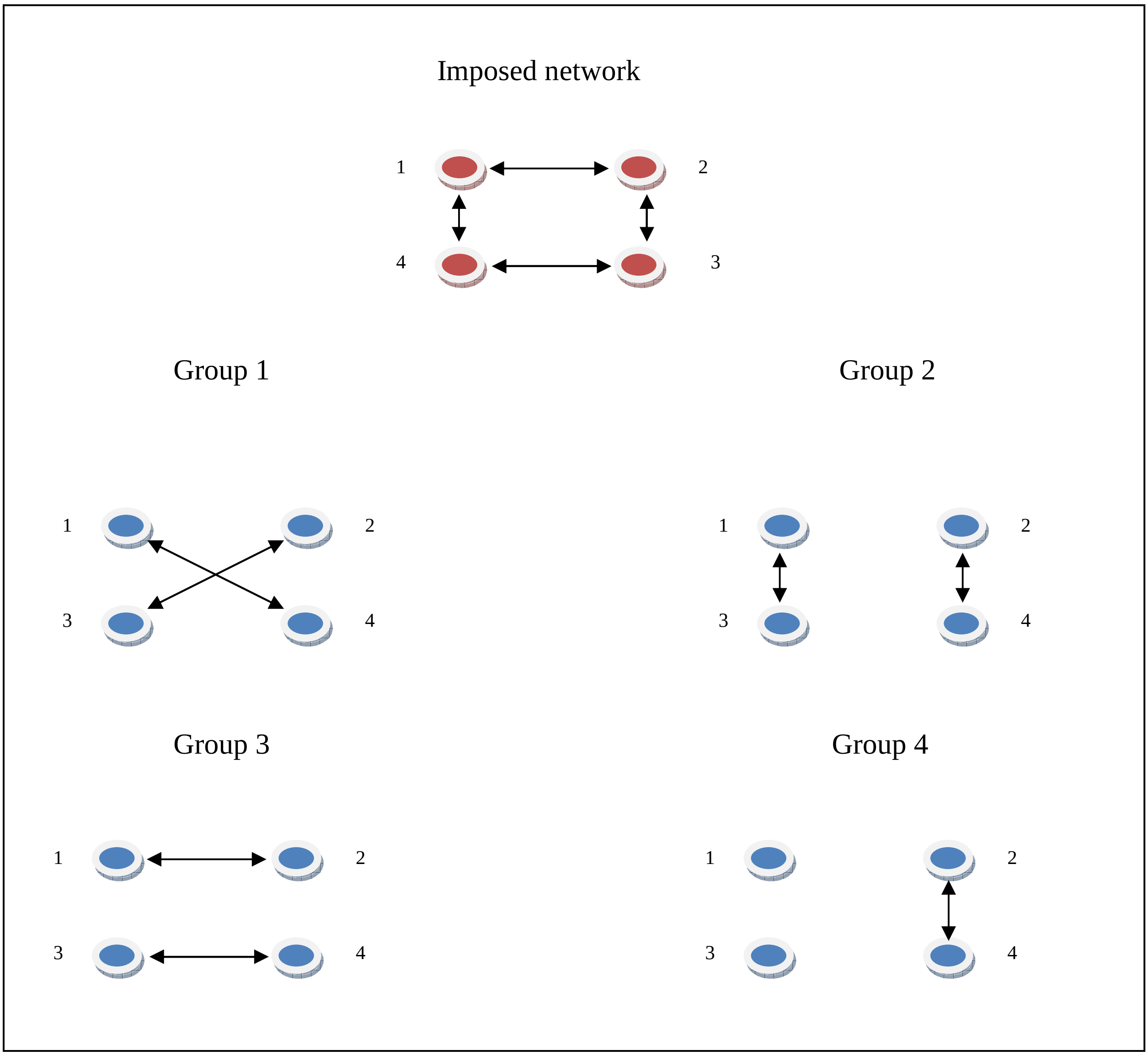
Figure 4.7 – Achieved networks per each group in ENCG –scenario 2 (complete)



iii. Scenario 3: *Wheel network*

Again, even if the wheel represents a clearly over-connected structure, participants in groups 1-3 mutually agreed to converge on the strongly efficient and fair (from a payoff perspective) network architecture. This was not the case in group 4 where only one pair was built, that is, one connection between subjects 2 and 4, while agents 1 and 3 decided, in a utility maximising myopic way, not to build any link. In fact, this network structure gave the connected individuals a payoff of 18, while the unconnected players received a 0 payoff. Even if group 4 seemed to randomly agree on connection decisions, if we compare it with the other three groups, we will notice (see section 4.3.3) that it is characterised by a different level of individual cooperativeness, which could explain the different results. **Figure 4.8** shows the agreed networks.

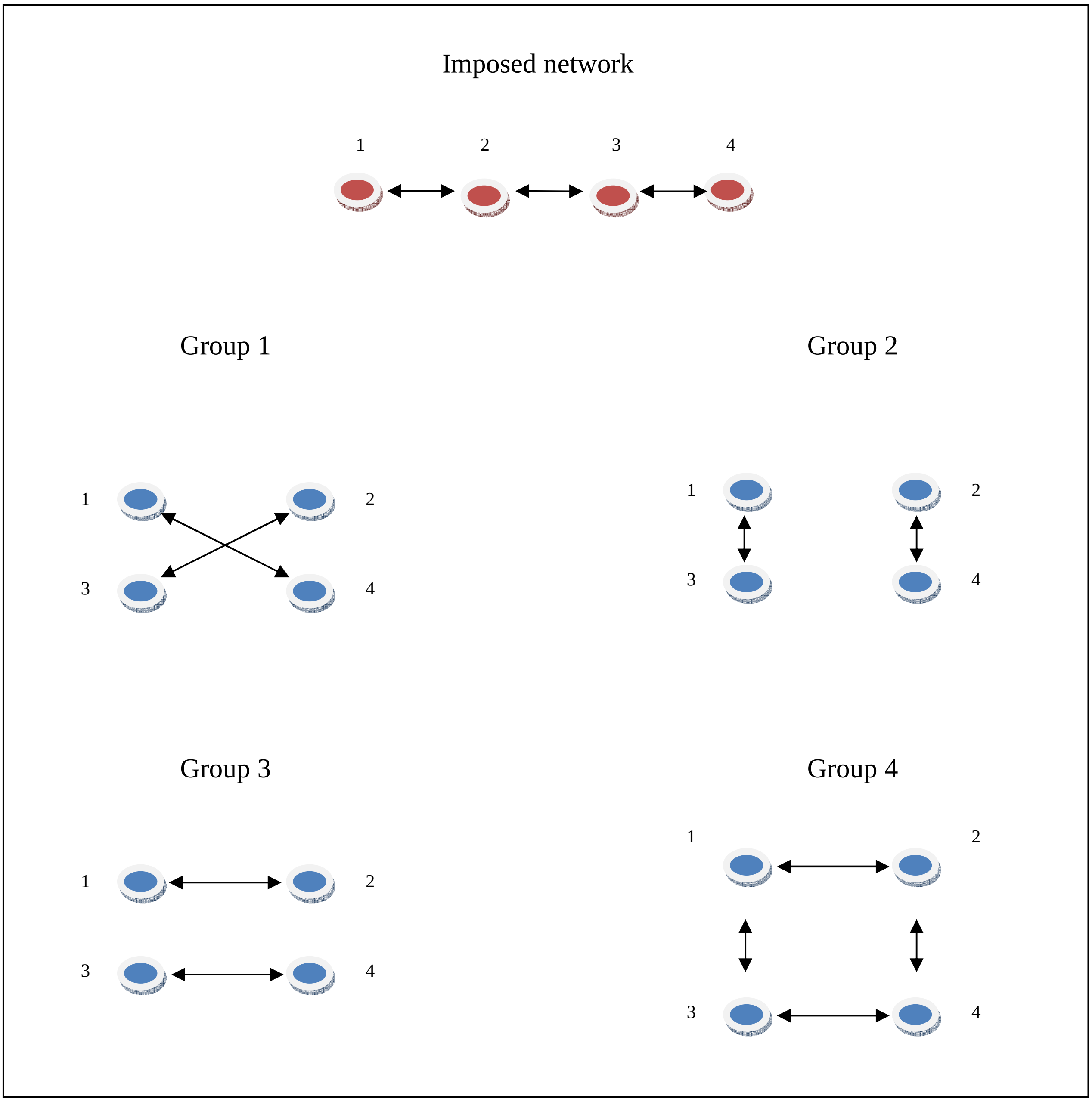
Figure 4.8 – Achieved networks per each group in ENCG –scenario 3 (wheel)



iv. Scenario 4: *Line network*

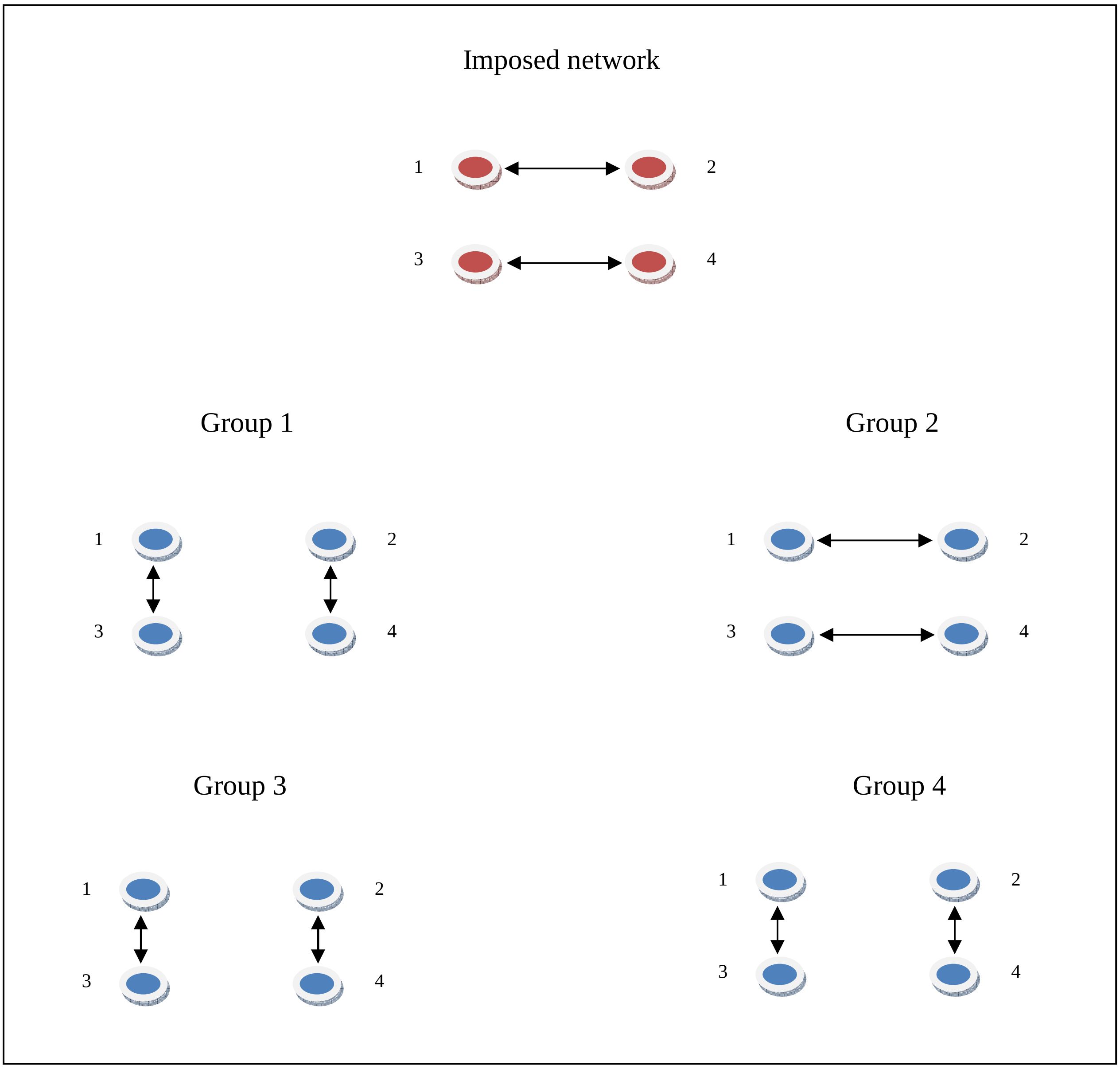
In this scenario, groups 1-3 were not influenced at all by the line structure. We can think that, once all participants have really understood that the double-pair is the most efficient network, they always converge upon it in a sort of *inertia*.

Figure 4.9 – Achieved networks per each group in ENCG –scenario 4 (line)



However, group 4 deviated and decided to build two collaborations per each group member; thus, creating a wheel, which includes more redundant connections than the line network does. Instead, if we consider the payoff symmetry, a wheel is a symmetrical structure, while the line is not. Therefore, we can assume that group 4 members were influenced by the initial network structure, but cared about payoff equality at the same time. In particular, all participants received a payoff which was equal to 15. **Figure 4.9** depicts the achieved network structures per group of participants.

Figure 4.10 – Achieved networks per each group in ENCG –scenario 5 (double-pair)



i. Scenario 5: *Double-pair network*

In the last scenario, the strongly efficient network was presented to the participants as exogenously imposed. With no surprises, all groups agreed not to change the architecture and all participants received a payoff of 18. For sake of completeness, we report in **figure 4.10** the confirmed network architectures.

We can now proceed with some general observations before providing the reader with more details about the interpretation of the results in light of the personal characteristics and attitudes towards cooperation of the participants.

First of all, we need to clarify that the number of observations we gathered is too small to make systematic inferences. This is why we have already stressed previously in this section that our experiment has to be consider just a pilot experiment, a first attempt to experimentally evaluate co-authorships phenomena.

The results we obtained partially confirmed the validity and reliability on one hand of the J&W model of co-authorship, even if, as we will underline in section 4.5, with some limits due to its intrinsic assumptions; and on the other hand the empirical results obtained by Vanin. In fact, with the exception of group 4, Vanin's predictions are confirmed also by the results we obtained: subjects converge upon the strongly efficient network, which is not pairwise stable though; hence, people internalise the negative externalities generated by adding new connections (that is, collaborations). We can conclude that the expected tendency towards building redundant connections is not displayed, at least in our sample.

However, how can we explain the behaviour and the connection decisions of group 4? The results could induce us to think that in some way the exogenous networks presented in ENCG did have an influence, together with other group-specific factors. Even if it is just one group out of 4, given the available few observations, we need to consider and analyse it in details, because we do not know which trend would be displayed if we integrate the experiment with a more sufficient number of sessions (hence, observations).

The next paragraph serves this purpose.

4.4.3.3 Results: The cooperativeness index

In order to better explain the results, and in particular the disparities between the connection decisions made by group 1-3 and group 4, it is necessary to take into account participants' individual and cultural characteristics on one hand, and their level of cooperativeness on the other hand.

With concern to the individual characteristics, as we have already explained, all participants are postgraduate international students at The University of Nottingham (UK campus). Even though they are enrolled in different programs, we can assume that they are all highly skilled and coming from different departments should not make a difference between them. However, 2 participants are students of Economics, therefore trained to rational thinking. We could believe this could have made a difference, but we need to consider that one of the 2 students was part of group 1 and the other of group 2, and in group 3 (where participants always converged on the double-pair structure) participants were from Architecture, Manufacturing and Geography and they reached the same network. Therefore, the specific educational background did not seem to influence the obtained results.

Table 4.3 – Participants' biographical data per group

Participant ID number	Gender	Age	Experimental Group	Department/School	Country of origin	PDI	avg PDI per group	IDV	avg IDV per group	Years in the UK
1	F	3	1	Architecture	Puerto Rico	95*	73	11*	31	1
2	M	3	1	Eng and Manufacturing	Venezuela	81		12		1
3	M	3	1	Eng and Manufacturing	Spain	57		51		1
4	M	3	1	Economics	Mexico	81		30		1
5	M	3	2	Architecture	Taiwan	58	50	17	54.75	1
6	F	4	2	Appl. Ling. and Engl. Studies	Japan	54		46		1
7	M	3	2	Business School	The Netherlands	38		80		1
8	M	3	2	Economics	Italy	50		76		1
9	F	4	3	Architecture	Chile	63	57	23	47.75	1
10	M	3	3	Eng and Manufacturing	Italy	50		76		1
11	M	2	3	Architecture	Iran	58		41		1
12	M	4	3	Geography	Spain	57		51		1
13	M	4	4	Engineering	Chile	63	67.3	23	31.3	1
14	M	3	4	Business School	Morocco	70**		46**		1
15	M	3	4	Engineering	Mexico	81		30		1
16	F	3	4	Business School	Iran	58		41		1

** Puerto Rico is not included in Hofstede's (2001) list of countries; therefore we opted for using the values of Panama because we believe it is the most similar country to Puerto Rico amongst the latin-american ones.*

*** Estimated values*
Legend: Age categories: ***1*** *= ‹ 20,* ***2*** *= 20-23,* ***3*** *= 24-27,* ***4*** *= 28-31,* ***5*** *= › 31*
Gender: ***M*** *= male,* ***F*** *= female*

The same conclusion can be drawn in relation to the gender of the participants. Even if they were randomly assigned to sessions (hence, to groups), every group was composed of 3 men and 1 women. Therefore, we do not have any variance between the groups and a comparison cannot be made.

The same observations can be done with concern to age and years spent in the UK.

Table 4.3 summarises participants' biographical and cultural data.

Instead, we can consider the country of origin as a measure of participants' cultural background. **Table 4.3** reports the values per participant and the appropriate average index per group of two indices elaborated by Hofstede (2001)[77] on a database of employees' values firstly built by IBM (1967-1973 on a sample of 70 countries) and then extended by Hofstede until 2001 covering 74 countries[78]. These measures are computed from a set of questions presented to workers in the form of Likert scale items (*ibid*). In particular, two cultural indices were selected:

i. the *Power Distance* Index (PDI), which is a measure of inequality but "from below, not from above" (http://www.geert-hofstede.com/index.shtml). Hofstede adopts the definition of power distance originally elaborated by Mulder (1977), which consists of "the degree of inequality in power between a less powerful Individual (I) and a more powerful Other (O), in which I and O belong to the same (loosely or tightly knit) social system". In other words, it is a proxy of the extent to which power inequality between people in a society is accepted and expected to be as a normal feature of the society itself[79];

ii. the *Individualism Index* (IDV), which measures the extent to which in a society "the ties between individuals are loose: everybody is expected to look after him/herself and his/her immediate family" (http://www.geert-hofstede.com/index.shtml)[80].

These two measures can help us interpreting the results we obtained. In particular we can look at the correlation between the number of times each group achieved the strongly efficient network

---

[77] For more details, please refer also to Hofstede, G. and Hofstede, G.-J. (2004).

[78] For additional information concerning Hostede's indices please visit http://www.geert-hofstede.com/index.shtml *[viewed 2nd Sept 2011]*, and http://www.gertjanhofstede.com/ *[viewed 2nd Sept 2011]*, in addition to the references provided in the bibliography.

[79] For an analytical definition of the PDI index and they way in which it is computed and adapted from the original IBM scores, please refer to Hofstede, 2001:85-91.

[80] For an analytical definition of the IDV index and they way in which it is computed and adapted from the original IBM scores, please refer to Hofstede, 2001:214-219.

configuration[81] and the average PDI per group. Their relationship is negative (-0.355); therefore, the fact that inequality is more widely accepted in the societies where some of our experimental subjects come from is inversely correlated to the convergence towards strongly efficient network architectures. On the contrary, IDV and the number of times the groups achieved the double-pair network is positively correlated (0.55). Hence, individualism is positively correlated to network (and payoff) efficiency. Even if it has to be recognised that we are considering mere correlations, the results above give us a good overview about the trend that could be found out in an extension of our experiment to a larger scale of observations.

Table 4.4 – Set of items (questions) of the Likert Scale

Question ID code	Item
a	My collaborators usually act in favour of the interests of our cooperative work
b	I think I usually act in favour of the interests of my cooperative work more than my collaborators do
c	My relationship with my collaborators can be defined as "mutually gratifying"
d	I usually trust the work done by my collaborators
e	I usually trust my collaborators to do things I cannot do by myself
f	I share my knowledge with my collaborators whenever possible
g	I am usually open to being influenced by my collaborators' ideas
h	I enjoy discussing issues with my collaborators when we are working together

Adapted from: Rindfleisch (2000)

Additionally, if we consider scenario 3 when the wheel network was shown to the participants, group 4 decided to create just one pair and the other two individuals were isolated in the network. In particular, agents 2 and 4 built a collaboration; they come from Iran and Morocco respectively and their countries have a IDV score of 46 and 41 which are higher compared to the scores of the other group members' countries (23 and 30).

As far as the level of cooperativeness of each participant is concerned, we computed an index extracted from the answers the subjects gave to 8 question about their perception on themselves as collaborators and about the cooperative attitude of their collaborators in real life contexts (*e.g.*: job experiences, university group coursework). **Table 4.4** lists the set of questions or *items* (Rindfleisch, 2000) which had to be answered using a *Likert scale* (from 1 – strongly disagree – to 7 – strongly agree –).

[81] Recall that, out of 6 possibilities, groups 1, 2 and 3 achieved the strongly efficient configuration 6 times, while group 4 just 2 times.

As we can notice from **table 4.5**, if we consider the average score per items all over the groups, questions *f* and *h* obtained a high average score (5.81 and 6.06 respectively), meaning a wide agreement on the items: participants seem to enjoy discussing and sharing their knowledge with collaborators.

Nonetheless, items *b* and *e* received the two average lowest scores (4.12 and 4.43 respectively): on average, participants do not think they act in favour of their collaborative work more than what their collaborators actually do, but, at the same time, they do not trust the collaborators on doing tasks they cannot do by themselves.

Additionally, considering the average cooperativeness index per group, it is worth noting that while groups 1-3 have an average index about 5.3, group 4 has an average index of 4.03. Therefore, people in groups 1-3 seem to have a stronger attitude toward cooperation, while participants in group 4 has a weaker one. Hence, we can infer that participants' cooperativeness might have played an important role in our experiment, in particular when, in ENCG, we had imposed exogenously given network structures. Furthermore, the variances amongst individuals who are part of the same group and the variances amongst the average cooperativeness level of the four groups can support us on explaining some of the results obtained in ENCG. For example, group 2 struggled to converge upon the strongly efficient network when both star and complete networks were shown. If we look at the variance of the cooperativeness index among group members, we can notice that they have the highest variance (0.72) when compared to the ones of the other groups. Therefore, we can infer that this group was less "homogenous" in terms of cooperativeness level, even when compared to group 4, which, although has the lowest average index (4.03), presents the second highest variance among individuals (0.57). This could be one of the reasons why group 2 partially deviated from the behaviour displayed by groups 1 and 3, where the variance between individuals is lower (0.44 and 0.01 respectively). Although we do not have enough observations in order to achieve general results, we cannot ignore this aspect and further research has to be done on this issue.

Table 4.5 – Index of cooperativeness per participant and per group

		Likert Items											
Group ID number	**Participant ID number**	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**Average Index**	**Index variance among agents**	**Group Average**	**Index variance per group average**
1	**1**	7	4	5	5	7	7	6	6	**5.87**	**0.74**	**5.53**	**0.44**
	2	7	6	7	6	5	7	5	7	**6.25**			
	3	7	3	5	5	4	6	3	5	**4.75**			
	4	6	4	6	6	3	6	5	6	**5.25**			
2	**5**	7	6	4	7	6	7	6	7	**6.25**		**5.37**	**0.72**
	6	6	6	6	5	5	6	6	6	**5.75**			
	7	4	4	4	4	4	4	4	6	**4.25**			
	8	7	1	7	4	2	7	7	7	**5.25**			
3	**9**	1	1	7	7	6	7	6	7	**5.25**		**5.37**	**0.01**
	10	7	5	6	5	3	5	6	6	**5.37**			
	11	5	6	6	6	7	5	3	6	**5.5**			
	12	5	4	6	6	6	5	4	7	**5.37**			
4	**13**	4	5	4	5	4	5	5	5	**4.62**		**4.03**	**0.57**
	14	1	1	1	3	3	7	4	7	**3.37**			
	15	4	5	3	3	3	3	3	3	**3.37**			
	16	7	5	5	3	3	6	3	6	**4.75**			
	avg score per item	**5.31**	**4.12**	**5.12**	**5**	**4.43**	**5.81**	**4.75**	**6.06**				

Finally, as expected, the correlation between the average cooperativeness index per group and the number of times each group achieved the strongly efficient network configuration is positive (0.99).

If we take into account *hypothesis 2*, it is possible to state that individual characteristics *per se* did not affect the experimental results[82], while cultural backgrounds and social preferences might have had an impact on them. In fact, with regard to personal features, as we have underlined above and as it emerged from the audio of the sessions, participants were often influenced by leader(s) of the group, who they "imitated". The leader(s), therefore, played a fundamental role on convergence onto the strongly efficient network, at least in BCG, where the network was endogenously built by the participants. Moreover, attitudes towards cooperativeness are important when we take into consideration the network evolution: showing subjects different pre-imposed network configurations to dynamically modify had an influence on the group composed of the "most selfish" participants[83] (that is to say, the participants with the lowest average index of cooperativeness).

As far as fairness attitudes are concerned, in the analysis of the results above we have showed that participants cared about payoffs equity and symmetry, both in BCG and ENCG. The most significant example is provided, again, by group 4; when they were showed the line network (which is asymmetric), they opted for a wheel network, which is not payoffs maximising but symmetric. For further research purposes it would be useful to apply a theoretical model *à la* Fehr and Schimdt (1999) in order to systematically test this relationship in a larger sample.

In light of this evidence, we can conclude that *hypothesis 2* cannot be rejected.

4.4.3.4 Results: Final observations

We can now proceed on considering whether the hypotheses 1, 3 and 4 we made at the beginning of this section were satisfied or not according to the experimental results.

With regard to *hypothesis 1* concerning negative externalities, we had controversial evidence. In fact, partially in line with the results obtained by Vanin (2002), in BCG all groups fully internalised the negative externalities generated by the same assumptions of the co-author model;

[82] Nevertheless, we need to specify that group 4 was the only group in which any of the participants knew each other. Even if we cannot consider acquaintance as a personal feature of the individuals, we can consider it as a sort of "relational" characteristic itself, which could have had played a role.

[83] It is interesting to notice, for example, that one of the participants in group 4 suggested four times to the other participants not to collaborate at all.

on the contrary, group 4 deviated in most of the cases from convergence on the strongly efficient but not pairwise stable configuration in ENCG. Additionally, it is important to notice that we let people playing the connection games under the most favourable conditions, that is, allowing them to communicate and discuss the collaborations structure amongst each other. What happens if we relax this assumption and we put emphasis on the role of individual decision making? We will provide a partial answer and a suggestion for improvement section 4.5; for the moment, we can just state that, with regard to hypothesis 1, we had controversial evidence: in the majority of the observations and sessions people did internalise the negative externalities but we cannot ignore the peculiarities of group 4 which did not overcome this problem and converged upon over-connected non efficient networks. Hence, we can only partially accept it.

The same conclusions can be inferred in the case of *hypothesis 3*; people have partially deviated from the efficient networks when, coping with exogenously imposed networks, they were given the possibility of severing collaborations by incurring no costs. Although the no-cost condition was revealed as a strong assumption, we can only partially accept hypothesis 3, because, in BCG, where no possibility of deleting links was given, all groups converged on the strongly efficient network, while in ENCG group 4 frequently deviated.

Moreover, the same arguments can be used to provide support to the non rejection of *hypothesis 4*, according to which we should have observed differences in terms of efficiency when people are asked to endogenously build their own collaborations compared to when they exogenously modify a pre-existing network, even if we cannot conclude that the difference is *significant* (as stated in the relevant hypothesis). Finally, we observed that different imposed network structures made group 4 deviating from efficiency, also from a payoffs perspective. In that way, payoffs were not maximised, while, again, symmetry and equity concerns were always taken into account by all groups.

4.5 Conclusion and Research Agenda

In this study an experimental application of the co-author model of J&W has been presented. The main purposes of this chapter were to analyse the way in which collaborative networks emerge and the potential trade-off between networks pairwise stability and efficiency. Although our experiment is conceived as a pilot, mainly because of the small sample size, some observations drawn from our analysis could be considered of interest for future research.

Firstly, it is possible to confirm one of the main results obtained in the analogous experiment designed by Vanin (2002): even if we let people form their own collaborative network *under the* *most favourable conditions* (*e.g.*: communication amongst participants), it is not clear whether people systematically reach the strongly efficient network configuration or not. The behaviour of the participants in one of the experimental groups clearly confirmed this tendency.

Secondly, evidence of the trade-off between pairwise stable configurations and efficiency is provided: most of the time participants converged and agreed on building a strongly efficient architecture, which, on the other hand, is not pairwise stable, given the structure of costs and benefits the participants were provided with.

Thirdly, we obtained controversial evidence with respect to the differences between the condition in which people built their own network endogenously and the one in which they were provided with exogenously imposed configurations. However, star, line and complete networks appeared to be the imposed architectures which have mostly influenced the participants, also in one of those experimental groups that always agreed on the strongly efficient network eventually. These results might be interpreted in light of the fact that people, when making their connection decisions in a collaborative environment, do not care only about utility maximisation, but they are also concerned with fairness motives and with payoff symmetry amongst individuals.

Finally, through a simple statistical analysis and the elaboration of both an individual and a group index of cooperativeness, we showed the importance of investigating the role of people's cultural features and degrees of cooperativeness. We provided a first insight into the crucial role these factors could play on affecting individuals' connection decisions. However, it was not possible to take into account individual characteristics (*e.g.*: gender and years spent in the UK) because our sample did not show variance with respect to them.

Nonetheless, it is useful to underline some limitations which emerged from the study, in order to provide some suggestions for a further development of this analysis.

With concern to the game in which participants were shown different imposed network structure (ENCG), we must specify that the order in which they had been presented could have affected groups' decisions. Due to the fact that we had only 16 participants, it was not possible to vary the order of the networks per different groups; in fact, in order to make a systematic comparison, we would have needed a sufficient number of groups in which networks were displayed, for instance, from the strongly efficient network to the star; another set of groups in which the order was the opposite and so on. Additionally, it would be interesting to check whether letting people play *only* ENCG would give the same results or not; on one hand, we could incur the risk that subjects would not be sufficiently confident with the game, but on the other we could isolate the role played by imposed exogenous structures more clearly. Further developments of this experiment should take these aspects into consideration.

Moreover, due to budget constraint, it was not possible to use PC lab when running the experiment; hence, the “pen and paper” solution was adopted. Nonetheless, using a software could have helped to “clean” the results, for different reasons. *i)* the experimenter would be able to control for the actual time people take to make a decision, both individually and in terms of groups; *ii)* the potential systematic influence of one member of the group (*e.g.*: a leader) could be detected and some inferences could be drawn from his/her behaviour and from the other members’; *iii)* the degree of cooperativeness could be reduced and the role of individual genuine intentions emphasised in a more systematic and effective way. For example, if we let people playing at PC stations, on one hand, they would not know the real identity of the members of their same group: we could control for factors which might generate experimental noise (*e.g.*: physical aspect, mutual acquaintance or a sort of “first impression” effect); and on the other hand, they could be allowed to make several links proposals to be accepted or rejected by anonymous other members contemporaneously. This process could last until the experimenter (by adequately programming the software) randomly stops the game (Di Cagno and Sciubba, 2010).

Furthermore, from some comments made by the member of one of the groups (number 2), it emerged how important would be to design the experiment with more rounds. In this group, participants explicitly considered the opportunity, when a hierarchical network as the star was shown, of a sort of “rotation” of the person who was going to get the highest payoff. Statements like “*I allow you to get the highest payoff now, but in the next round you give me back the favour: I get the highest and you take the lowest payoff*” clearly suggest a modification of the experiment in this way. In fact, people had to exclude the possibility of gaining the highest payoff in rotation, because the experimental setting and conditions did not imply this possibility: participants did not know whether in the following round of the game they would have been provided with the same network or with a different one. However, given the experimental parameters and the random lottery incentive system, agreeing on the double-pair network was the best solution in terms of payoff maximisation.

In addition, this variation could help to explain some phenomena which occur in real collaborative networks. For instance, in the case of scientific co-authorship networks, different name ordering rules exist, based on which the authors of a paper written in co-authorship are listed; these rules mainly depend on the scientific area. Nonetheless, for example in economics, there is not such a specific rule and scientists could decide to alternatively be the "first author" of a paper; position that in some cases could assign more prestige to the scientist[84].

Finally, it is useful to underline some of the limits embedded in the J&W co-author model itself, which could have influenced the results we obtained.

In fact, the model assumes zero payoff when no links are built amongst agents. Moreover, the model imposes a time endowment, which people can spend on working in collaboration with other individuals, but not alone. This is a restrictive assumption: in real networks people could intentionally decide to work alone and to spend all their time endowment in it. Therefore, an improvement of the model should take into consideration the possibility of being an isolated agent in the network, and incurring appropriate costs while gaining benefits as well.

Additionally, with regard to the time endowment mentioned above, the assumption that people equally split it between the collaborations they build seems to be far from reality. We could think that on one side, different projects require different effort levels, hence less or more time; and on the other side, people can autonomously decide to invest more or less time on different projects, or different stages of a project require diverse amounts of resources, including time.

In conclusion, it is fundamental to underline that J&W co-author model and the experiment illustrated in this essay deal with agents' homogeneity. This implies that the same value function is assigned to all individuals; we did not assume that participants were given different knowledge values, different expertise and so on, while in real networks this is the case. For example, our experiment could be enhanced following the experiment of Goeree et al. (2009) where subjects are divided in different "types of individuals" (low, high and normal); the authors, even if they follow the theoretical model by Bala and Goyal (2000), found that individuals' heterogeneity do affect the emergence of, for instance, hierarchical networks. Assuming that people differ with respect to some relevant criteria makes the controlled environment more realistic without loosing experimental control and avoiding possible confounding factors and noise in the data. Moreover, some of the architectures proposed in our experiment could be revealed as stable. As Cowan and Jonard (2004) suggest, a theoretical model in which a vector of knowledge types is considered could provide a deeper insight into the way in which networks form and evolve. Alternatively,

[84] For example, suppose four authors named A, B, C and D are going to publish four papers together in 6-year time. Since each single paper is written by the same four authors and it will be (hopefully) cited by other papers as "A et al.", they could decide to rotate the first author amongst the four papers, equally splitting the "prestige" deriving from the publication.

another solution that could be implemented consists in including a coefficient which depends on the type of the link formed (*e.g.*: a link is built between a high type and a low type individuals and a coefficient is assigned to the value of their collaboration, which is different from the one assigned to a link between two low-type agents, for instance).

The suggestions proposed above are of fundamental importance in order to improve the experiment illustrated in this essay and they could be useful to provide a more exhaustive understanding of the mechanism behind the processes of network formation in a collaborative environment.

APPENDIX A – EXPERIMENTAL INSTRUCTIONS ***(General, Game 1 and Game 2)***

INSTRUCTIONS

You have been asked to participate in an economics experiment. The instructions you are about to read are self-explanatory. No questions will be answered once the experiment starts. If you have any questions, you should read back through these instructions and ask the experimenter before we start. From now on I ask you not to talk at all during this experiment (except for when you are allowed to) and to pay attention to the instructions.

In this experiment you will be asked to play two different connection games. Each time you will receive appropriate instructions. In both games you will be playing with three other persons, and they will be the same for the entire experiment.

Each of you is denoted by an identification number which was randomly assigned to you at the beginning of the experiment; I kindly ask you to show your identification number on your desk during the experiment. You will keep this number in both games.

In both games you have the opportunity to build connections with the other agents, in order to work together on a project. You may decide to create several two-person links, that is to say connections involving you and one other agent working on the same project. Deciding to build connection(s) implies that both you and your partner incur some costs in terms of time, but also gains in term of the benefits deriving from the collaboration(s). You may also decide not to build any link and in that way you will not incur neither costs nor gains.

Any collaboration is established only if both agents agree upon. For example, if you propose an agent to work together, he/she has to accept your offer in order for the connection to be built. In the two games you will have the possibility to establish a maximum of three collaborations.

You are given an initial time endowment of 6 hours. If you decide to collaborate with other agent(s), you split your time endowment equally among all collaborations you are involved, according to the following table.

Number of YOUR collaborations	***Hours YOU spend per collaboration***
0 *(alone)*	***0*** *(do not work with others)*
1 *(one collaboration)*	***6*** *(all your time is for that collaboration)*
2 *(two collaborations)*	***3*** *(half time for each collaboration)*
3 *(three collaborations)*	***2*** *(one third of time for each collaboration)*

Additionally, for each collaboration, you and your partner benefit from a *synergy effect*, which is an increase in the payoff due to the collaboration itself. The synergy effect depends on how many hours you and your partner decide to spend on the collaboration. The following table shows the synergy effect per each possible combination:

Hours YOU spend on the collaboration	*Hours your PARTNER spends on the collaboration*	*SYNERGY EFFECT*
6	*6*	***6***
6	*3*	***3***
6	*2*	***2***
3	*6*	***3***
3	*3*	***1.5***
3	*2*	***1***
2	*6*	***2***
2	*3*	***1***
2	*2*	***0.6***

For any given collaboration, you and your partner receive the same payoff, given by:

Agent's payoff = Number of hours you spend on the collaboration

+ Number of hours your partner spends on the collaboration

+ Synergy effect

Your **total payoff** is just the sum of the payoffs you receive for each of your collaborations.

Example:

Suppose you are involved in only 1 collaboration and your partner is involved in 3 collaborations (including the one with you). Therefore, you spend 6 hours and your partner just 2 hours on the collaboration. Hence, your payoff for that collaboration is 10 = 6 + 2 + 2.

If you decide to establish more than 1 collaboration, you just have to sum up the payoffs you receive in each collaboration. Suppose that

Collaboration #1 payoff = 10

Collaboration #2 payoff = 6

Collaboration #3 payoff = 4.6

Therefore, your total payoff is 20.6 = 10 + 6 + 4.6.

At the end of the experiment, one of the two games and one of the four participants will be randomly selected by the experimenter. The selected participant will be rewarded according to the payoff he/she received in the selected game. 1 point will be equal to 30 pence.

Do you have any questions?

Now please complete the questionnaire you are given, in order to practise calculating the payoffs. When you have finished, please raise your hand and the experimenter will check your answers. The experiment will start once all participants have successfully answered the questionnaire.

GAME NUMBER 1 – YOUR TASK

In this game you are asked to decide if you want to collaborate in order to work on one or more projects with other agents. If so, you have to decide with whom you want to collaborate and to find the necessary agreement in order to establish the collaboration.

You are given 5 minutes to think alone about your decision(s). When your time is up, the experimenter will stop you and will ask you to start an open discussion with the other agents, in order to reach an agreement on the connection decisions. The open discussion will last 10 minutes.

Recall that a collaboration is established if and only if both agents agree upon and your payoff is calculated according to the connection(s) you will be able to build during the open discussion.

Your payoff is calculated in the way explained in the instructions. If you cannot reach an agreement with the other agents, the payoff of all participants will be equal to 0.

At the end, please fill out your decision sheet according to the collaborations you have built and give it back to the experimenter.

Please take a look at the decision sheet you are given and be sure you fully understand how to fill it out.

Do you have any questions?

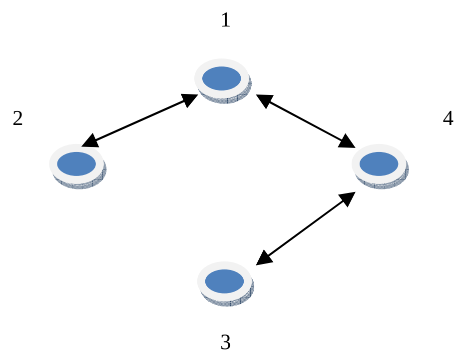
GAME NUMBER 2 – YOUR TASK

In this game 5 different scenarios representing collaborations between agents will be shown you one at a time. You are free to decide:

- ✓ not to change any of your collaborations and leave your collaborations as shown in the network;
- ✓ delete one or more of your collaborations;
- ✓ add one or more collaborations (up to a maximum of 3 collaborations).

Also in this game you need to find the necessary agreement if you want to establish new collaboration(s), whereas you do not need any approval by your partner in order to severe a collaboration.

For example, suppose you are agent number 1 and you are shown the following collaboration scenario:



You could decide not to modify any of your collaborations or to delete both collaborations (with agents 2 and 4); or to add, for example, a new collaboration with agent number 3.

You are given 1 minute to think alone about your decision(s). When your time is up, the experimenter will stop you and will ask you to start an open discussion with the other agents, in order to reach an agreement on the connection decisions. The open discussion will last 5 minutes. Recall that a collaboration is established if and only if both agents agree upon, but you could autonomously severe one or more collaborations. Moreover, remember that your payoff is calculated according to the connection(s) you will be able to build during the open discussion.

Your payoff is calculated in the way explained in the instructions. If you cannot reach an agreement with the other agents, the payoff of all participants will be equal to 0.

At the end, please fill out your decision sheet according to the collaborations you have built, give it back to the experimenter and wait until the experimenter shows you the second scenario and so on.

The experiment ends when you give back the decision sheet for the fifth scenario.

Please take a look at the decision sheets (one per each scenario) you are given and be sure you fully understand how to fill it out.

Do you have any questions?

Appendix B – Sample of the Decision Sheets (Game 1 and Game 2)

GAME 1

DECISION SHEET

Agent number 1

Recall that 1) you are not forced to fill out all the tables. You only need to complete as many collaboration tables as the number of collaborations you could build during the game (if any). You always have to fill out the TOTAL PAYOFF TABLE though; 2) your collaboration(s) is established according to the decisions agreed during the open discussion.

✓ **Collaboration tables**

A - Collaboration with agent number:			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
....	....	....	....

B - Collaboration with agent number:			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
....	....	....	....

C - Collaboration with agent number:			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
....	....	....	....

✓ **Total payoff table**

Your payoff in collaboration A	**Your payoff in collaboration B**	**Your payoff in collaboration C**	**Your TOTAL PAYOFF**
....	....	....	....

GAME 2

DECISION SHEET – SCENARIO 1

Agent number 1

Recall that 1) you are not forced to fill out all the tables. You only need to complete as many collaboration tables as the number of collaborations you could build during the game (if any). You always have to fill out the TOTAL PAYOFF TABLE though; 2) your collaboration(s) is established according to the decisions agreed during the open discussion.

✓ Collaboration tables

A - Collaboration with agent number: ……………………….			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
……………………………	……………………………	……………………………	……………………………

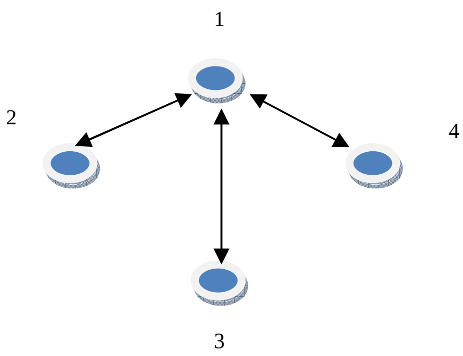
B - Collaboration with agent number: ……………………….			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
……………………………	……………………………	……………………………	……………………………

C - Collaboration with agent number: ……………………….			
Hours you spend on this collaboration	Hours your partner spends on this collaboration	Synergy effect	Your payoff for this collaboration
……………………………	……………………………	……………………………	……………………………

✓ Total payoff table

Your payoff in collaboration A	**Your payoff in collaboration B**	**Your payoff in collaboration C**	**Your TOTAL PAYOFF**
……………………………	……………………………	……………………………	……………………………

SCENARIO 1



Please feel free to use this space to write your own notes.

PARTICIPATION AGREEMENT

Economics experiment on decision-making

By signing up this form you accept to participate in this economics experiment.

Please note that the experimenter will treat all responses in strict confidence during this experiment and no personal information will be passed onto third parties.

All responses will be saved in a format which protects your and other participants' anonymity. Experimental data will be stored using personal identification numbers to identify participants.

All experimental results will be reported in a manner which relates to the whole experimental sample of respondents. Readers will not know who participated in the experiment, and therefore will not be able to infer any personal details about any participant.

Please note, the experimenter has received formal approval to undertake this experiment, which is compliant with University of Nottingham's **'Code of Research Conduct and Research Ethics'** –

http://www.nottingham.ac.uk/ris/local/research-strategyandpolicy/Code_of_Conduct(Version_3_January_2010).pdf

Participation will take approximately 60 minutes, and one randomly selected participant can earn up to 9 pounds depending on his/her decisions during the game.

I have read and understood this Participation Agreement. I agree with the terms outlined above, and wish to participate in this experiment.

Name ……………………………………….

Signature ……………………………………….

Date ……………………………………….

APPENDIX E – QUESTIONNAIRE: PAYOFFS

QUESTIONNAIRE

Please answer the following questions using the instructions concerning the calculation of the payoffs. Once you have finished, please raise your hand and the experimenter will check your answers and clarify your doubts, if any.

1. Suppose you decide to establish 1 collaboration with another agent who wants to collaborate only with you. Therefore, you have 1 collaboration and your partner has 1 collaboration.

Your payoff <u>for that collaboration</u> is: ____

2. Suppose you decide to establish 1 collaboration with another agent who wants to work only with you. Instead, you also want to collaborate with another agent. Therefore, you have 2 collaborations and your partner has 1 collaboration.

Your payoff <u>for that collaboration</u> is: ____

3. Suppose you decide to establish 1 collaboration with another agent who wants to collaborate with you, but he also wants to establish another collaboration with another agent. You also want to collaborate with another agent. Therefore, you have 2 collaborations and your partner has 2 collaborations.

Your payoff <u>for that collaboration</u> is: ____

4. Suppose you decide to establish 1 collaboration with another agent who wants to collaborate with you, but he also wants to work with other 2 agents. You also want to collaborate with other 2 agents. Therefore, you have 3 collaborations and your partner has 3 collaborations.

Your payoff <u>for that collaboration</u> is: ____

5. Suppose you equally split your time working with 2 different agents. Therefore, you allocate 3 hours per each collaboration. Your first partner is working only with you; hence, he/she allocates all the time (6 hours) for the collaboration he/she has established with you. Instead,

your second partner is involved in other 2 collaborations; hence, he/she can spend 2 hours working with you.

Your <u>TOTAL</u> payoff is: ____

<u>Answers:</u>

1. *6+6+6=18*
2. *3+6+3=12*
3. *3+3+1.5=7.5*
4. *2+2+0.6=4.6*
5. *(3+6+3)+(3+2+1)=12+6=18*

APPENDIX F – QUESTIONNAIRE: BIOGRAPHICAL INFORMATION AND COOPERATIVE BEHAVIOUR

QUESTIONNAIRE

Please answer the following questions about your biographical information and your attitudes. Your identity will not be revealed neither to the other participants nor to the experimenter.

Thank you very much for your collaboration!

1. Gender: □ female □ male

2. Age: □ ‹ 20 □ 20 – 23 □ 24 – 27 □ 28 – 31 □ › 31

3. MSc/PhD Program and year: ____

4. Department/School: ____

5. Country of origin: ____

6. Years spent in the UK (if not British): ____

7. To what extent Do you agree with the following statements? (Please answer accordingly to your job experience a/o university experience – *e.g.*: group projects/courseworks)

a. My collaborators usually act in favour of the interests of our cooperative work.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

b. I think I usually act in favour of the interests of my cooperative work more than my collaborators do.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

c. My relationship with my collaborators can be defined as “mutually gratifying”.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

d. I usually trust the work done by my collaborators.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

e. I usually trust my collaborators to do things I cannot do by myself.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

f. I share my knowledge with my collaborators whenever possible.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

g. I am usually open to being influenced by my collaborators' ideas.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
□	□	□	□	□	□	□

h. I enjoy discussing issues with my collaborators when we are working together.

Strongly disagree						Strongly agree
1	2	3	4	5	6	7
□	□	□	□	□	□	□

APPENDIX G – NA INDICES OF CENTRALITY OF A NODE IN A NETWORK

To be reminded here that, broadly speaking, the indices of centrality of the networks allow the analyst to define the hierarchical position that any node in a network play in terms of its strategic power within the network itself.

	INDEX	AIM	COMMENT
Degree centrality	$C'_G = \frac{C_G}{(n-1)} = \frac{\sum_{i=1}^{n} l(a_i, a_k)}{(n-1)}$ l = direct link between two adjacent nodes (*ai*, *ak*) C_G = un-normalised degree centrality	It measures the number of links per each single node in the network.	The index is normalised according to the network magnitude.
Closeness centrality	$C'_P(a_k) = \frac{n-1}{\sum_{i=1}^{n} d(a_i, a_k)}$ d = geodesic distance (the smallest one) between two nodes (a_i, a_k)	It measures the node's strategic role in a network: it is related to its ability to reach each other node efficiently and autonomously.	The index is normalised according to the network magnitude.
Betweenness centrality	$C'_I(a_k) = \frac{2C_I(a_k)}{(n-2)(n-1)}$ $\frac{(n-2)(n-1)}{2}$ = maximum value of $C_I(a_k)$	It measures the potential strategic role which a node could play in the scientific collaboration network, through connecting sub networks which otherwise would be disconnected from each others.	The index is normalised according to the network magnitude. The un-normalised index is computed as follows: $b_{ij}(a_k) = \frac{g_{ij}(a_k)}{g_{ij}}$ $b_{ij}(a_k)$ = proportion of geodesic distances (the smallest ones) between every two nodes in the network (i, j and $i<z>j$ $g_{ij}(a_k)$ = geodesic distance between i and j (where we also find z) g_{ij} = the shortest path between i and j

References

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics*, 74: 47-97.

Aumann, R. and Myerson, R. (1988). Endogenous Formation of Links Between Players and of Coalitions: an Application of the Shapley Value. In: Roth, A.E. (eds.), *The Shapley Value.* Cambridge, England: Cambridge University Press, pp. 175-191.

Bala, V. and Goyal, S. (2000). A Noncooperative Model of Network Formation. *Econometrica*, 68: 1181-1229.

Barabási, A.-L. (2002), *Linked. The New Science of Networks*, Plume Books, USA

Barabási, A.-L., Jeong H., Néda, Z., Ravasz E., Schubert, A. and Vicsek, T. (2002). Evolution of the Social Network of Scientific Collaborations. *Physica A*, 311: 590-614.

Berkowitz, S.D. (1988). Markets and Market-area: Some preliminary formulations. In: Wellman, B. and Berkowitz, S.D. (eds.), *Social Structures: A Network Approach*. Cambridge, England: Cambridge University Press, pp. 261-303.

Bertrand, M., Luttmer, E.F.P. and Mullainathan, S. (2000). Network Effects and Welfare Cultures. *The Quarterly Journal of Economics*, 115: 1019-1055.

Boyle, G. (2008). Pay Peanuts and Get Monkeys? Evidence from Academia. *The B. E. Journal of Economic Analysis & Policy*. 8: art. 21.

Breiger, R.L. (1981). The social class structure of occupational mobility. *American Journal of Sociology*, 87: 578-611.

Breiger, R.L. (1990). *Social Mobility and Social Structure*. Cambridge, England: Cambridge University Press.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. (2000). Graph Structure in the Web. *Computer Networks*, 33: 309-320.

Burt, R.S. (1988). The stability of American Markets. *American Journal of Sociology*, 94: 356-395.

Cainelli, G., Maggioni, M.A., Uberti, E. and De Felice, A. (2010). The strength of strong ties. Co-authorship and productivity among Italian economists. *Marco Fanno working papers series*, n. 0125, Dipartimento di Scienze Economiche "Marco Fanno".

Cainelli, G., Maggioni, M.A., Uberti, E. and De Felice, A. (2012). Co-authorship and productivity among Italian economists. *Applied Economic Letters*, forthcoming.

Callander, S. and Plott, C.R. (2005). Principles of Network development and evolution: an experimental study. *Journal of Public Economics*, 89: 1469-1495.

Carayol, N., Roux, P. and Yildizoğlu (2008). In search of efficient network structures: the needle in the haystack. *Review of Economic Design*, 11: 339-359.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3): 817-869.

Cowan, R. and Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28:1557-1575.

Crane, D. (1972). *Invisible Colleges. Diffusion of Knowledge in Scientific Communities*. The University of Chicago Press, Chicago & London.

Cubitt, R.P., Starmer, S. and Sugden, R. (1998). On the Validity of the Random Lottery Incentive Scheme. *Experimental Economics*, 1: 115-131.

Daley, D.J., Gani, J. (2000). *Epidemic Modelling*. Cambridge University Press. Cambridge.

David, P. A. (2008). 'The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution'. *Capitalism and Society*, 3: art. 5.

De Solla Price, D. J. (1986). *Little Science, Big Science… and beyond.* Columbia University Press.

Deck, C. and Johnson, C. (2002). Link Bidding in a Laboratory Experiment. *Review of Economic Design*, 8(4): 359-372.

Di Cagno, D. and Sciubba, E. (2010). Trust, trustworthiness and social networks: Playing a trust game when networks are formed in the lab. *Journal of Economic Behavior and Organization*, 75: 156-167.

Dutta, B. and Mutuswami, S. (1997). Stable Networks. *Journal of Economic Theory*, 76(2): 322-344.

Erdős, P., Rényi, A. (1959). On Random Graphs. *Publicationes Mathematicae*, 6: 290–297.

Fachamps, M., van der Leij, M. & S. Goyal (2006). Scientific Networks and Co-authorship. *Economics Series Working Papers*, n. 256, Oxford, University of Oxford.

Falk, A. and Kosfeld, M. (2003). It's all about Connections: Evidence on Network Formations. *mimeo, University of Zurich*.

Fehr, E. and Schmidt, K.M. (1999). A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics*, 114: 817-868.

Frey, B. S. (2003). Publishing as Prostitution? Choosing Between One's Own Ideas and Academic Success. *Public Choice,* 116:205-223.

Frey, B. S. (2009). Economists in the PITS?. *International Review of Economics,* 56:335-364.

Goeree, J.K., Riedl, A. and Ule, A. (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67: 445-466.

Goyal, S. (2005). Strong and Weak Links. *Journal of European Economic Association*, 3:608-616.

Goyal, S. (2007). *Connections. An Introduction to the Economics of Network*. New Jersey: Princeton University Press.

Goyal, S. (2009). *Connections. An Introduction to the Economics of Network*. New Jersey: Princeton University Press.

Goyal, S., van der Leij, M.J. and Moraga, J.-L. (2006). Economics: An Emerging Small World?. *Journal of Political Economy*, 114(2): 403-432.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 81:1287-1303.

Granovetter, M. (1974). *Getting a job: A Study of Contacts and Careers.* Cambridge, MA: Harvard University Press.

Granovetter, M. (1983). The strength of Weak Ties. A Network Theory revisited. *Sociological Theory,* 1(6): 1360-1380.

Greene, W.H. (1994). *Econometric Analysis*. Pearson Education.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47: 153–61.

Hoekman, J., Frenken, K. and van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, 43(3): 721-738.

Hofstede, G. (2001). *Culture's Consequences. Comparing Values, Behaviors, Institutions and Organizations Across Nations.* Thousand Oaks CA: Sage Publications.

Hofstede, G. and Hofstede, G.-J. (2004). *Cultures and Organizations: Software of the Mind.* New York: McGraw-Hill USA.

Hollis, A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, 8: 503-530.

Holzer, H. (1987). Hiring Procedures in the Firm: Their Economic Determinants and Outcomes. *NBER Working Paper*, n. 2185.

Jackson, M.O. (2003). A Survey of Models of Network Formation: Stability and Efficiency. In Demange, G., Wooders, M. (eds.) *Group Formation in Economics; Networks, Clubs and Coalitions*, pp. 11-56. Cambridge University Press, Cambridge.

Jackson, M.O. (2009). *Social and Economic Networks*, Princeton University Press, Princeton.

Jackson, M.O. and Watts, A. (2001). The Evolution of Social and Economic Networks. *Journal of Economic Theory*, 106(2): 265-295.

Jackson, M.O. and Wolinsky, A. (1996). A Strategic Model of Social and Economics Networks. *Journal of Economic Theory*, 71: 44-74.

Johnes, J. (2006). Data Envelopment Analysis and its Application to the Measure of Efficiency in Higher Education. *Economics of Education Review*, 25: 273–288.

Johnson, C. and Gilles, R.P. (2000). Spatial Social Networks. *Review of Economic Design*, 5:273-300.

Kirchkamp, O. and Nagel, R. (2001). Naïve Learning and Cooperation in Network Experiments. *Games and Economic Behavior*, 58(2): 269-292.

Kosfeld, M. (2004). Economic Networks in the Laboratory: A Survey. *Review of Network Economics*, 3(1): 20-42.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34:1-14.

Lazerson, M. (1993). Factory or Putting-Out? Knitting Networks in Modena. In: Grabher, G. (eds.), *The Embedded Firm: On the Socieconomics of Industrial Networks*. New York: Rutledge, pp. 203-226.

Leifer, E.M. and White, H.C. (1987). A structural approach to markets. In: Mizruchi, M.S. and Schwartz, M. (eds.), *Intercorporate Relations: The Structural Analysis of Business*. Cambridge, England, Cambridge University Press, pp. 85-108.

Lévi-Strauss, C. (1963). *Structural Anthropology*. New York: Basic Books.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences*. Volume 7. Sage Publications, Thousand Oaks, CA.

Lowenstein, G.F., Thompson, L. and Bazerman, M.H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57:426-441.

Maggioni, M.A. and Togni, L. (2012). Micromotives and Macrobehavior in Scientific Formation: Towards an Empirical Estimates of Jackson and Wolinsky (1996). *CSCC working paper series*.

Maggioni, M.A. and Uberti, E. (2009). Knowledge networks across Europe: which distance matters?. *The Annals of Regional Science*, 43(3): 691-720.

Maggioni, M.A. and Uberti, E. (2011). Networks and geography in the economics of knowledge flows, *Quality and Quantity*, 45(5):1031-1051.

Maggioni, M.A., Nosvelli, M. and Uberti, T.E. (2007). Space versus Networks in the Geography of Innovation: A European Analysis. *Papers in Regional Science*, 86(3): 471-493.

Mantovani, M., Kirchsteiger, G., Mauleon, A. and Vannetelbosch, V. (2011). Myopic or Farsighted? An Experiment on Network Formation. *Fondazione Enrico Mattei Working Papers*, n. 596.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Springer, The Netherlands.

Montgomery, J. (1991). Social Networks and Labour Markets Outcomes: Towards an Economic Analysis. *American Economic Review*, 81: 1408-1418.

Moreno, J.L. (1946). Sociogram and Sociomatrix: A note to the paper by Foryth and Katz. *Sociometry 1*, 348-349.

Mulder, M. (1977). *The Daily Power Game*. Leiden, Netherlands: Martinus Nijhoff.

Myerson, R. (1977). Graphs and Cooperation in Games. *Mathematics of Operations Research*, 2: 225-229.

Nagpaul, P., Roy, S. (2003). Constructing a Multi-objective Measure of Research Performance. *Scientometrics*, 56:383–402.

Newman, M.E.J. (2001). The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences USA*, 98(2): 404-409.

Nishiguchi, T. (1994). *Strategic Industrial Sourcing*. New York: Oxford University Press.

Riedl, A. and Ule, A. (2002). Exclusion and Cooperation in Social Network Experiments. *mimeo, University of Amsterdam*.

Rindfleisch, A. (2000). Organizational Trust and Interfirm Cooperation: An Examination of Horizontal Versus Vertical Alliances. *Marketing Letters*, 11(1): 81-95.

Rivellini, G., Terzera, L. (2010). Co-authorship in Italian Worshops on Population Studies: An Analysis with a Network Approach. *Connections*, 18: 57-74.

Rousseau, S., Rousseau, R. (1997). Data Envelopment Analysis as a Tool for Constructing Scientometric Indicators. *Scientometrics*, 40:45–56.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White D.R. (2009). Economic Networks. The New Challenges. *Science*, 325:422-425.

Scott, J. (2000). *Social network analysis. A Handbook.* London: Sage Publications.

Shrum, W., Genuth, J. and Chompalov, I. (2007). *Structures of Scientific collaboration.* MIT Press (MA).

Slikker, M. and van den Nouweland, A. (2000). Network Formation Models with Costs for Establishing Links. *Review of Economic Design*, 5: 333-362.

Starmer, C. and Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *The American Economic Review*, 81(4): 971-978.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26:24-36.

Togni, L. (2009). La Geografia dei Geografi: il ruolo di collaborazioni e acknowledgments in una comunità scientifica. *CSCC working paper series*, 0209.

Togni, L. (2011). An Experiment on Stability and Efficiency of Scientific Co-authorship Networks. *mimeo*, The University of Nottingham.

Togni, L. (2011). Does Co-authorship matter for Scientific Productivity? Evidence from Geography's Top Journals. *CSCC working paper series*, 0111.

van Dolder, D. and Buskens, V. (2009). Social Motives in Network Formation: An Experimental Study. *Accepted paper series of the International Conference on Game Theory for Networks (2009)*, pp. 593-602.

Vanin, P. (2002). Network Formation in the Lab: a Pilot Experiment. *mimeo*, Universitat Pompeu Fabra.

Vega Redondo, F. (2007). *Complex Social Networks*. Cambridge University Press, New York.

Vega-Redondo, F. (2006). Building Up Social Capital in a Changing World. *Journal of Economic Dynamics and Control*, 30(11): 2305-2338.

Vuong, QH (1989). Likelihood Ratio Tests for Model Selection and Non-tested Hypotheses. *Econometrica*, 57: 307-333.

Wagner, C. S. (2008). *The New Invisible Colleges: Science for Development*. Brookings Institution Press.

Warning, J. (2004). Performance Differences in German Higher Education: Empirical Analysis of Strategic Groups. *Review of Industrial Organization*, 24:393–408.

Wassermann, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge, England: Cambridge University Press.

Watts, A. (2001). A Dynamic Model of Network Formation. *Games and Economic Behavior*, 34: 331-341.

Watts, D.J. (1999). *Small Worlds. The Dynamics of Networks between Order and Randomness*. New Jersey: Princeton University Press.

Watts, D.J. and Strogatz, S.H. (1998). Collective Dynamics of “Small World” Networks. *Nature*, 393(6684): 440-442.

Weibull, J.W. (2000). Testing Game Theory. *Stockholm School of Economics*, Working Paper n. 382.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadverted consequences?. *Scientometrics*, 62: 117-131.

White, H.C. (1981). Where do markets come from?. *American Journal of Sociology*, 87: 517-547.

White, H.C. (1988). Varieties of Markets. In: Wellman, B. and Berkowitz, S.D. (eds.), *Social Structures: A Network Approach*. Cambridge, England: Cambridge University Press, pp. 226-260.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. (2006). *Introductory Econometrics: A Modern Approach*. International Edition, Thomson.