UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche “Marco Fanno”

THE STRENGTH OF STRONG TIES: CO-AUTHORSHIP AND PRODUCTIVITY AMONG ITALIAN ECONOMISTS

GIULIO CAINELLI
University of Padova

MARIO A. MAGGIONI
University Cattolica del Sacro Cuore of Milano

T. ERIKA UBERTI
University Cattolica del Sacro Cuore of Milano

ANNUNZIATA DE FELICE
University of Bari

December 2010

“MARCO FANNO” WORKING PAPER N.125

The strength of strong ties:
Co-authorship and productivity among Italian economists

Giulio Cainelli°, Mario A. Maggioni[+]*, T. Erika Uberti*, Annunziata de Felice^

° *Dipartimento di Scienze Economiche 'Marco Fanno', Università degli Studi di Padova e CERIS-CNR, Milano*
* *DISEIS and CSCC, Università Cattolica del Sacro Cuore, Milano*
^ *Dipartimento per lo studio delle società mediterranee, Università degli Studi di Bari*

> Most intuitive notions of the 'strength' of an interpersonal tie should be satisfied by the following definition: the strength of a tie is a (probably linear) combination of *the amount of time*, the emotional intensity, the intimacy (mutual confiding), and *the reciprocal services* which characterize the tie. Each of these is somewhat independent of the other, though the set is obviously highly intracorrelated. Discussion of operational measures of and weights attaching to each of the four elements is postponed to future empirical studies
>
> M. Granovetter (1973):1361, emphasis added

Abstract: Increased specialization and extensive collaboration are common behaviours in the scientific community, as well as the evaluation of scientific research based on bibliometric indicators. This paper aims to analyse the effect of collaboration (co-authorship) on the scientific output of Italian economists. We use Social Network Analysis to investigate the structure of co-authorship, and econometric methodologies to explain the productivity of individual Italian economists, in terms of '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 indexes and clustering coefficients).

Keywords: co-authorship, scientific productivity, Italian economists, social network analysis.

[+] Corresponding author: DISEIS, Università Cattolica del Sacro Cuore, L.go Gemelli 1, I-20123 Milano, ITALY, tel.: ++39 02 72343951; *email*: mario.maggioni@unicatt.it.

I. INTRODUCTION

The world of scientific research has undergone dramatic change, which originated in the realm of the 'hard sciences' and extended rapidly to the social sciences and especially the field of economics. Economics frequently acts as the bridge between the 'hard' sciences and 'social' science. Increased specialization and extensive collaboration are becoming more and more common behaviours in the scientific community, and the evaluation of scientific research using bibliometric indicators is being adopted by government bodies and funding agencies across the world (Kalaitzidakis *et al.* 1999).

These phenomena are not independent; indeed, there is an extensive stream of literature (see, among others: Sauer, 1988; Barnett *et al.*, 1988; Piette and Ross, 1992; Laband and Piette, 1995; Hudson, 1996; Laband and Tollison 2000, 2006) showing their increased interdependency since the increased pressure to publish on academics has spurred a higher propensity to co-author papers through to a series of demand-side and supply-side channels.

The aim of this paper is threefold: first, to describe the structure of the Italian academic economist community through an in depth analysis of its scientific output; second, to analyse the collaborative behaviour of this community; third, to study the effects of this collaboration on scientific productivity.

In Section II we briefly review the existing knowledge on the determinants and effects of co-authorship in the social sciences; Section III describes the purpose built, original database used in our analysis; Section IV introduces social network analysis (SNA) methodologies and describes a number of relational and positional indexes used in the subsequent econometric analysis. Section V presents the results of three econometric exercises to show the role of attributional, relational and positional variables in determining the productivity of Italian economists. Section VI concludes the paper.

II. CO-AUTHORSHIP: CAUSES AND CONSEQUENCES

Co-authorship is an increasing phenomenon in economics. Laband and Tollison (2000) show that in 'three prominent economics journals: the *American Economic Review*, the *Journal of Political Economy*, and the *Quarterly Journal of Economics'*, the percentage of co-authored papers grew steadily between 1950 and 1994, less than 10% in the 1950s and around 70% in 1994. Their data perhaps overemphasize the phenomenon, but are also consistent with the results of other studies. For instance, McDowell and Melvin (1983), for a sample of eight major economic journals, found the percentage of co-authored papers was about 30% in 1976; Hudson (1996), based on a similar sample, found that the average number of co-authored papers in 1993 was over 50%; Sutter and Kocher (2004) analyse 15 economics journals with the highest average impact factors, for the period 1977-1997 and found 44% of papers on average were co-authored; Medoff (2003) surveyed 'thirty-one top economics journals', and found the percentage of papers with two authors, in 1990, was approximately 40%; and Cosme Costa Vieira (2008), for a sample of 168 journals available in the 'economics' class in the ISI database, found 47% of papers were multi-authored (two or more authors) between 1986 and 1996.

This phenomenon of co-authorship is strictly related to the increasing practice of universities', government bodies' and funding agencies' evaluating scientific research output through 'objective' bibliometric indicators and to the need more generally, to 'publish with a high impact factor or perish', a policy that is pursued at all levels of academia.

The most general explanations of this connection rely on the belief that co-authorship increases both the quantity[1] and quality[2] of research output in the form of published papers (Lee *et al.* 2005). Thus, as the 'publish or perish' policy diffuses across national and disciplinary borders, the appeal of co-authorship increases among researchers. There is a substantial, but not very prominent stream of literature devoted to empirical and theoretical analysis of the determinants and consequences of co-authorship, which has studied this issue extensively in order to identify the determinants of this behaviour.

While most of the theoretical literature (see among others Butler, 2007) concentrates on complementarities in skills and attitudes, Fafchamps *et al.* (2006, p. 8) underline that 'if research output depends only on ability, collaboration is most likely between authors of a similar level of ability (assortative matching) but with non-overlapping competences (complementarity in competences)'. However 'collaboration between high and low ability authors can arise if the low ability author provides more effort. In this manner the time-constrained high ability author can produce more research while the low ability researcher produces better quality output' (*ibid*). In this process, each author faces a 'matching' problem in finding his co-author, and this problem is exacerbated by the fact that one of a co-author's qualities is based on an *ex ante* evaluation of an unobservable variable (his/her effort). However, since 'collaborating with someone reveals valuable information about their ability and motivation, it follows that a

[1] The empirical evidence for both arguments is mixed. In terms of quantity, McDowell and Smith (1992) use cross-sectional data on academics and regress the number of articles produced by an individual (with co-authored articles discounted by the number of authors) on the percentage of co-authored articles and find no significant result. Durden and Perri (1995) use time series data for annual economics publications, over 24 years, and find that the total number of publications is positively related to the number of co-authored publications. Hollis (2001) uses the same data, and regresses total publications on the proportion that is co-authored, and finds no significant relationship. He then employs panel data for 339 individuals (US and Canadian AEE members) and finds that, if publications are discounted by number of authors, 'adding one more author is associated with a per capita reduction in output of between 7% and 20%' (Hollis, 2001, p. 527).

[2] In terms of quality also, the empirical evidence is mixed. Laband and Tollison (2000) document an acceptance rate of 12% for collaborative papers submitted during the mid-1980s to the *Journal of Political Economy*, compared to 10% for single-authored papers. Others, such as Laband (1987), measure quality by citation frequency and report that this index is significantly higher for co-authored articles, while Barnet et al. (1988) use the positioning of the article in the journal as the signal of quality and find no support for this argument.

referral about a researcher *i* is particularly informative when it is provided by a previous coauthor of *i*. Referral by a coauthor can thus be construed as a vetting process, stating whether a coauthor is competent and can be trusted to do his or her share of the work' (*ibid*., p. 12).[3]

The scale and scope of the empirical literature indicate that a brief survey of the different determinants and facilitators of co-authoring highlighted over 20 years of research, would be helpful.

- Specialization: 'the explosion of knowledge in economics, as the sheer growth of knowledge resulted in increased efficiencies of specialization and co-authorship relative to working alone' (McDowell and Melvin, 1983, p. 156).

This argument has been developed by several other authors:

- Multi-disciplinarity: since it is often fruitful to bring different perspectives to the study of a single issue, this is easily achieved by a 'multi-disciplinary configuration of the research team'[4] (Sigelman, 2009, p. 508). Further: 'As authors work in areas outside of their major areas of specialty they tend to engage in co-authorship to a larger extent than the authors who published within presumably more familiar areas in terms of interest or expertise' (Piette and Ross, 1992, p. 281).
- Technological complementarities: the need to master very different analytical tools increases the importance of different expertise in order to manage numeric simulation packages, econometric estimation applications, and huge data bases[5] (Hudson, 1996).
- Synergy: the gains from collaborative work may be the result of a sort of synergy where 'multiple contributors develop ideas that none would have developed on his or her own. Synergy differs from skill complimentarily in the sense that it can exist between individuals with very similar skill sets' (Hudson, 1996, p. 157).

- Opportunity cost of time: the increased emphasis on research output and the use of publication output as a criterion for promotion, increases the opportunity cost of time for typical researchers in economics. Every other activity than doing research and writing papers, tends to be shelved or minimized. This means that acknowledgement is no longer a sufficient 'reward' for pre-submission review by a colleague in the field and 'this increased "price" often takes the form of co-authorship' (Barnett *et al*., 1988, p. 540).

[3] These cites theoretically support the inclusion of a series of network analysis indexes in the empirical part of this paper, to take account of both the attributional features of each author and his/her relational and positional characteristics.

[4] 'Interdisciplinary research also should be characterized by high rates of co-authorship by the same reasoning. The Piette-Ross insight, coupled with economists' steadily increasing colonization of other disciplines during the latter half of the twentieth century, may explain a significant portion of the increase in the incidence of co-authorship' (Laband and Tollison, 2000, p. 640).

[5] 'It may be cheaper for an individual to acquire new capacity (human capital) to produce through formal collaboration (merger) with someone who already has the requisite human capital than to acquire the needed knowledge de novo, personally' (Laband and Tollison, 2000, p. 639-40).

- Risk diversification: 'the editorial review process contains a random element that many would agree is large (...) and a given review may motivate a rejection, revision, or acceptance depending upon the editor's judgement concerning reader interest, the size of the journal's backlog, or a host of other potential factors (…). Thus, the author of a paper faces considerable uncertainty (...) A natural response is to diversify against this risk by co-authoring papers. Through co-authorship, one is able to increase the total number of papers submitted within a given period of time, thereby reducing the variance of the random element inherent in the review process. Thus, even if the value of co-authored papers is discounted exactly by the number of authors, and if there are no synergistic or quality effects in co-authoring, there will still be incentives to collaborate' (Barnet et al., 1988, p. 540).
- Assigned value of co-authored papers: department 'chairmen ordinarily "assign a weight" to coauthored papers that exceeds l/*n* (with *n* being the number of authors), presumably to encourage collaborative research' (Liebowitz and Palmer, 1983, quoted in Sauer, 1988, p. 857). Many universities and funding agencies promote and reward collaboration in the belief that it has a positive impact on research productivity (Laband and Tollison, 2000).
- Social interactions and pressures: collaboration may be chosen based on consumption/leisure reasons. Working with co-authors 'offers opportunities for friendship and camaraderie' and is a way to escape academic isolation (Medoff, 2003, p. 607; Acedo *et al.*, 2006; Holder, 2000). Having co-authors acts also as a motivation to keep to self-imposed deadlines.

We can add also that innovations in information and communication technologies (ICT) have made collaboration, even at distance, easier and less costly[6] although Sutter and Kocher (2004) based on a gravity model of the co-authorship patterns in US departments, and based on the top 15 economics journals, find that the coefficient of the variable for geographical distance is non-significant for all the periods investigated (1977, 1982, 1987, 1992, 1997).[7]

Finally, it must be acknowledged that co-authorship has some negative effects, otherwise every paper would be written in collaboration, and the number of collaborators per paper would be infinite.

- Compromise: 'An individual author working with a group will have to agree to a certain approach, certain text, even certain conclusions that that person might not enunciate in the same way if working alone. Because multiple authorship inevitably involves compromise,

[6] 'The development of technology has made collaboration more accessible across time and space. Overnight mail, photocopiers, computers, fax machines, email, and teleconferencing make long-distance collaboration considerably less daunting and time consuming. In essence, the invisible college of the 1960s and 1970s has been replaced by the "virtual college," or, more appropriately, the "virtual research center," of today' (Fisher *et al.* 1998, p. 847-848).

[7] For as the effects of ICT on co-authorship patterns, Butler (2007) shows that (at least for a subsample of American Economic Association members working in the JEL fields D8, G2 and J3) the rate of internet penetration moved from almost nothing in 1995 to around 60% in 1997, and to almost 100% in 1999; while Maggioni *et al.* (2009) show that the average distance between co-authors working on the issue of 'industrial clusters', increased continuously in the period 1969-2007.

my own intuition is that it tends to reduce risk taking in academic papers. The result may be more technically proficient papers than in the past, but at the cost of the imaginative leap forward that starts economics in a new direction or gives fresh impetus to an old subject area' (Hudson, 1996, p. 157).

- Organization and communication costs: 'multi-authored papers impose costs of organization and communication that may lead to diseconomies of scale. These are probably greater if all the collaborators are equally involved with all parts of the research and all parts of the paper. Developments in technology in recent years may have reduced the threshold at which these problems occur, but at some level they surely continue to exist' (*ibid.*, p. 157-8).
- Reward structure: 'Any net advantage of collaboration may disappear altogether if some individuals combine even though the sum of what each could achieve working alone exceeds their combined efforts. This may occur if an economist can achieve a greater gain in academic reputation from multi-authored rather than single-authored papers' (*ibid.*, p. 158).

From the above, it is evident that we need to consider both 'positional' and 'relational' variables in estimating the determinants of scientists' productivity. To achieve this, we built an original database in order to calculate positional and relational indexes for use as regressors in our econometric exercise.

III. THE ORIGINAL DATABASE

Our analysis is based on an original dataset built by matching two different data sources: (i) the Italian economists population drawn from the official database of the Italian Ministry of Universities and Research (*Ministero dell'Università e della Ricerca* – MIUR), managed by Cineca and (ii) the Econlit database of the American Economic Association.

Data from the MIUR-Cineca database refer 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).[8] The official MIUR definition of economics[9] includes six disciplinary groups; for the empirical analysis we classified these groups into four sub-disciplines: *Economics*, *Econometrics*, *Public Economics* and *Others*.[10] Table 1 presents the

[8] These positions in Italian are: *Professore Ordinario, Professore Straordinario, Professore Associato Confermato, Professore Associato, Ricercatore Confermato, Ricercatore*. There is another position of *Assistente Ordinario* which is between a lecturer and a professor whose academic duties are similar to a senior lecturer. Since this position is increasingly disappearing (percentage is only around 1%), we code this as SL. See Appendix A1 and Cainelli et al. (2006) for more details.

[9] The law is contained in the *Decreto Ministeriale 4 ottobre 2000* and published in the *G.U. n. 249 del 24 ottobre 2000 - supplemento ordinario 175*.

[10] *Economics* also includes Political Economy and Economic Policy (i.e. corresponding to the disciplinary sectors SECS-P/01 *Economia Politica*, SECS P/02 *Politica Economica*), *Econometrics* refers to the disciplinary sector SECS-P/05 Econometria, *Public Economics* to SECS-P/03 *Scienza delle Finanze*, and *Others* is a miscellaneous

distribution of academic positions and scientific fields in the Italian academic population of economists.

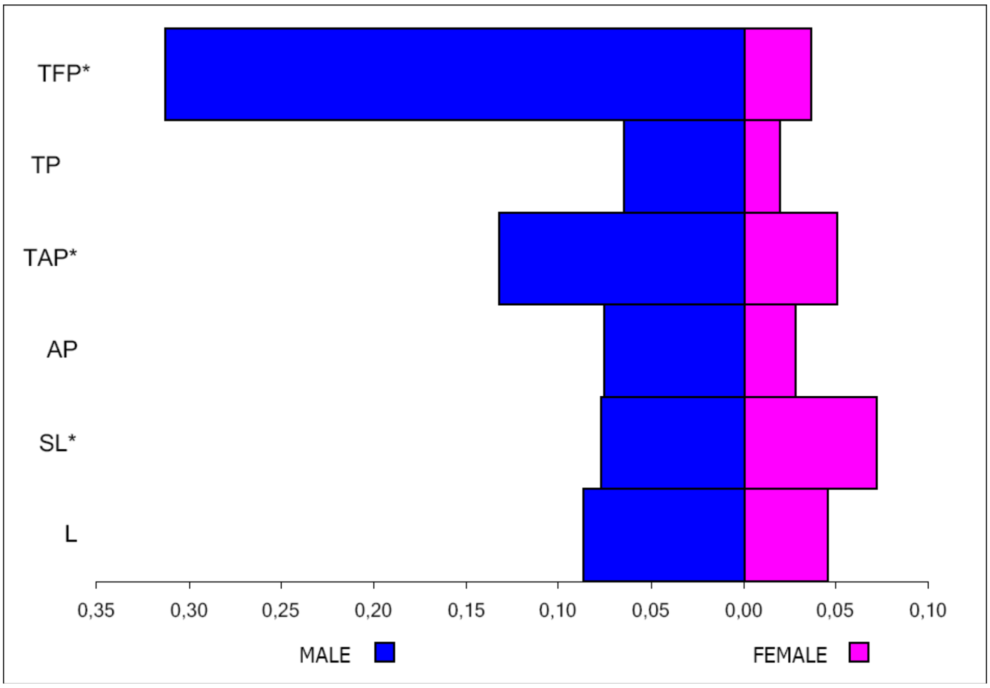
Table 1: Academic Position and Scientific Field
% percentage (by rows) and total number at 31 December 2006

SCIENTIFIC FIELD	L	SL*	AP	TAP*	FP	TFP*	Total
Economics	13.9	15.0	10.6	18.5	8.0	34.0	**1,153**
Econometrics	13.1	13.1	9.8	13.1	13.1	37.7	**61**
Public Economics	13.5	16.0	4.5	14.0	8.0	44.0	**200**
Others	8.7	13.6	14.6	22.8	10.2	30.1	**206**
Total	**214**	**241**	**167**	**296**	**137**	**565**	**1,620**

Note: (*) these academic positions are tenured
Source: our calculations based on MIUR-Cineca database at 31 December 2006

The population pyramid (Figure 1) shows a high percentage of male economists in the top positions (TFP: more than 30% of the total) and female economists mostly occupying the position of senior lecturers. This unequal distribution, gender biased and disproportionally large at the top, is typical of the gender-based distortion in the Italian labour market and reflects the effects of the last two university reforms which modified the selection procedures along an insider-outsider dynamics which privileged the advancements of careers of insiders at the expenses of entry of outsiders in the academia.

Figure 1: Distribution of Italian economists by academic position and gender



Source: our calculations on MIUR-Cineca database at 31 December 2006

disciplinary sector which includes SECS-P/04 *Storia del Pensiero Economico* (History of Economic Thought), and SECS-P/06 *Economia Applicata*, a mix of regional economics, transport economics, and industrial organization.

Having identified the population of Italian economists as defined by MIUR-Cineca, we determine their scientific production from the information in the Econlit database, categorized according to Econlit "product" groups – journal articles (JA), collective volume articles (CVA), books, working papers and dissertations – recorded for the period 1969-2006. These records were downloaded between August 2007 and February 2008, and painstakingly corrected for errors in people's names and double entries. Since our focus is scientific collaborations among Italian economists, we identify scientific productivity in terms of total number of JA published in the period being analysed (1969-2006), which is 8,679 JA.

Before describing the Econlit database we need to highlight some limitations. First, Dolado *et al.* (2003) suggest that it has missing information on authors, and contains some errors and omissions with respect to publications, number of pages, etc. Second, especially in the earlier years, the geographical coverage was limited to certain areas and most diffused international publishers, restricting enormously the entries of country-specific works (which are an important part of scientific research in economics). This limitation no longer applies: the database is changing continuously over time, and include a much broader set of journals and publishers worldwide. However, this introduces a bias in the dataset. Thus, in our comments relating to co-authorship networks it must be remembered that this database is dynamic. Finally, the database does not include any evaluation or publication weight (i.e. impact factor or similar) for scientific "product". Use of Econlit data means, we look at the quantitative profile of the internationally-visible scientific production of Italian economists, but we do not measure the "scientific value" of their publications. Although this might be considered a significant shortcoming, we do not believe it to be as serious as the dominant faction in the current debate on evaluation might suggest[11].

IV. RELATIONAL STRUCTURE OF CO-AUTHORSHIPS

This section introduces the SNA indexes and analytical tools (§IV.1), and their application to study the structure and evolution of the scientific collaboration behaviour of Italian economists (§IV.2) and finally we present SNA results (§IV.3).

IV.1. From two-mode networks to one-mode networks

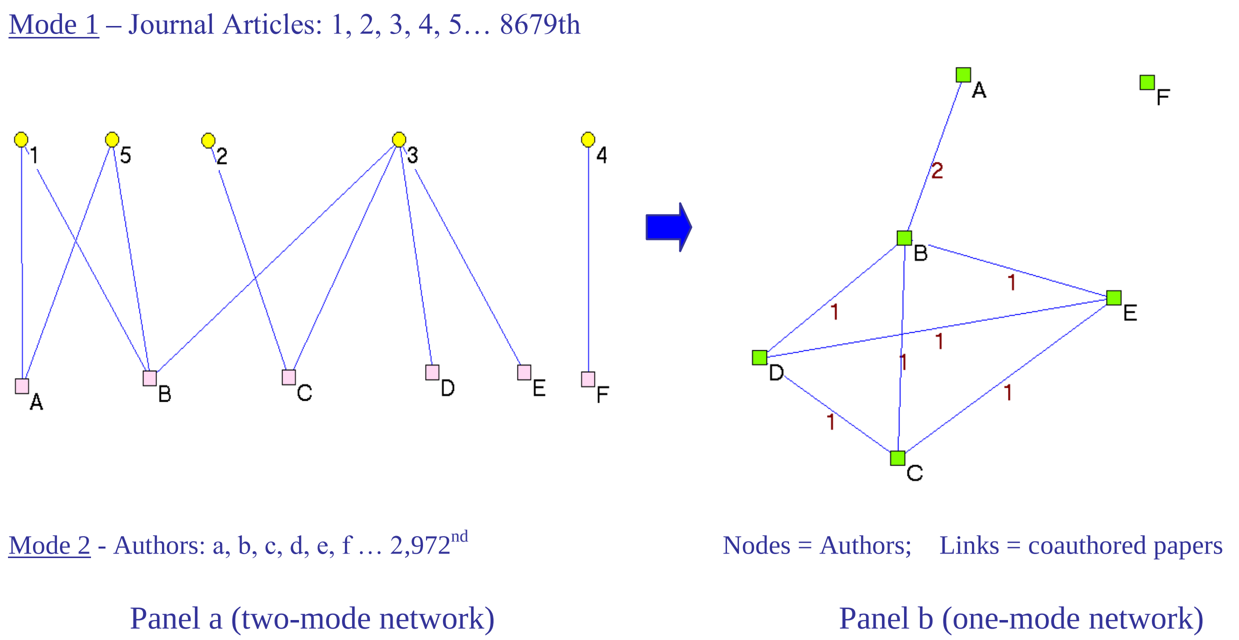
SNA is a scientific method of analysis that investigates the structure of the relations between social units of analysis, using graph theory, mathematics and statistics tools. Key SNA concepts are *actors* (i.e. players or nodes in the *network*); *relational ties* (i.e. the links connecting the actors) and *groups/subgroups* (i.e. the subsets of the actors and the relations among them).

[11] We are in fact convinced that the issue of how to weigh publications, and particularly the use of impact factors or citation indexes, should still be considered an open question for economics, as suggested by the results of the evaluation of European economics departments carried out by the *European Economic Association* (see Neary *et al.*, 2003).

Having defined the sample of actors and relations to be studied, SNA distinguishes the type or *mode* of the networks, i.e. the 'number of sets of entities on which structural variables are measured' (Wasserman and Faust, 1994, p. 35). Most SNA studies relate to 'one-mode networks' (i.e. where all entities or nodes belong to the same set of actors); however, some deal with *two* sets of social entities[12] (and the relations connecting one set to the other). These are called 'two-mode networks' or 'affiliation networks'.

In this paper we build a two-mode network, where one set of nodes, i.e. mode 1, is composed of journal articles (*JA*) and the other set of nodes, i.e. mode 2, is the papers' authors (*AU*) (see Figure 2, Panel a). We transform this into a 'one-mode network' in which *AU* are the nodes and the papers are the links between co-authors (see Figure 2, Panel b).

Figure 2: One-mode vs two-mode representation of co-authorship networks



More formally Panel a is a two mode network defined as **X** = (*JA*, *AU*, *R*, *w*), where the two disjoint sets are respectively *JA*, i.e. 8,679 articles, and *AU*, i.e. 2,972 authors included in the analysis, the scientific collaborations are denoted as $R \subseteq JA \times AU$ and the mapping $w: R \rightarrow \mathbb{R}$ represents a weight, i.e. the number of co-authored articles. The resulting affiliation matrix, **X**, is rectangular, i.e. 8,679 *JA* times 2,972 *AU*, with links originating from different authors and targeting the same paper representing cases of co-authorship. In order to emphasize the structure of co-authorship among economists we transformed the two-mode network into a one-mode

[12] These sets can include 'actors' and 'social events', i.e. members and administrative boards, authors and papers, citations and patents, etc. (Wasserman and Faust ,1994; Borgatti and Everett, 1997; Doreian *et al.*, 2004).

network (depicted in Panel b) where the network is defined as $\mathbf{X_1} = (AU, R_1, w_1)$, representing the collaborations, R_1, among authors AU, i.e. a 2,972 times 2,972 squared matrix where nodes are authors, and w_1 represents the number of scientific collaborations. We then performed a series of SNA analyses on the one-mode network for the whole period 1969-2006 and – given that co-authorship changes radically over time – we investigated the networks for four different historical periods (1969-1976; 1977-1986; 1987-1996 and 1997-2006).

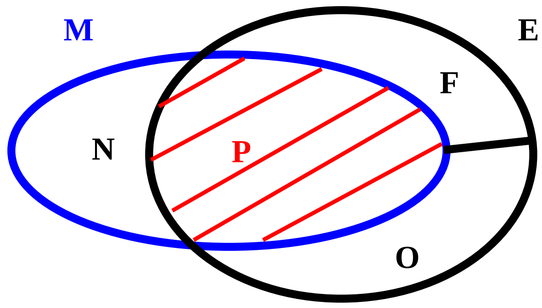
Before discussing the structural features of collaborative networks, we describe how the network of 2,972 authors is defined. Figure 3 synthesizes the final population of economists investigated, defined using the snowball sampling procedure, from 1,620 Italian economists selected from the MIUR-Cineca database.

First, we identify the entire population of Italian economists according to the MIUR-Cineca database (**M**) as defined in Section III, i.e. the 1,620 Italian economists[13] in an official academic position at the end of 2006. For each individual we identify the record indexed in the Econlit database, including details of year of entry and full JA records. This snowball procedure means that the number of individuals involved in writing JA increases with respect to the initial population **M** and we can identify a new Econlit population, **E**, that includes individuals belonging to set **M**, and all their co-authors (if any) with every affiliation,[14] for a total of 2,972 individuals.

The intersection between sets **M** and **E** produces a set **P**, which is composed of 1,317 'Italian' Academic Economists who wrote at least one JA indexed in Econlit,[15] and a subset **N** composed of 262 'Italian' academic economists with no JA entries in Econlit. The difference between **E** and **M** identifies the sum of two subsets: **O + F. F** is composed of 806 'foreign' economists with at least one coauthored JA recorded in Econlit, involving one or more Italian economists (included in **P**); **O** is composed of 849 'other Italian' coauthors with the economists in set **E**, which do not belong to **P** because: (i) they are affiliated to non university institutions (e.g. Bank of Italy, CNR-National Research Council, ISTAT-National Statistical Office, foreign and international institutions); or (ii) they belong to non economic sub-disciplines (i.e. business management, statistics, etc.).[16]

[13] The 19 foreign economists affiliated to Italian universities in the economics fields, in this paper are considered to be 'Italian economists'.
[14] The affiliations can be worldwide: this selection includes individuals with both foreign and Italian affiliations (i.e. individuals from other scientific sectors not included in those defined in Section III, or institutions outside academia, i.e. Banca d'Italia, CNR, ISTAT, etc.).
[15] We should stress that 41 individuals in the **M** population have entries in the Econlit database that do not refer to JA.
[16] Since we are interested in investigating differences in the collaborative behaviour of Italian economists in scientific collaborations with native academics and collaboration with foreigners, and since researchers are often quite mobile (and may change affiliations in the course of their careers), we attribute the residual population to sub-sets **F** or **O** based on nationality.

Figure 3: Set representation of the populations in the dataset



This 'partial snowballing' sampling procedure to build our database allows us to identify the structure of the scientific collaboration of the Italian academic economists (in set **P**); we do not investigate this structure for the residual **F** and **O** populations.[17]

IV.2. SNA indexes and network topology

Since our interest is in co-authorship in the production of scientific output, we treat the data on the scientific publication output of Italian economists transforming a two mode network into a one-mode network as described in Section IV.1, and calculate some indexes that measure the structure of the whole network and the positions and relational roles of individual nodes to be used as inputs for the econometric analysis in Section V.

In order to synthesize the general features of the overall network, we calculate its density value (d), degree centralization (C_{deg}), clustering coefficient (CC), average path length (APL), diameter (δ), and average degree (av_deg).

Density (d) is defined analytically as the ratio between the total number of actual links (L) and all possible potential links among all nodes in the network (n):

$$d = \frac{2L}{n \times (n-1)} \tag{1}$$

This network index ranges from 0 (i.e. the network is disconnected) to 1 (i.e. the network is complete; i.e. all possible links are present) and represents the completeness of the network (Wasserman and Faust, 1994).

Since it is not possible to compare densities for networks of different dimensions, we follow Maggioni (1995) and compute a relative density[18] value (d_r):

$$d_r = \frac{L-(n-1)}{\frac{n \times (n-1)}{2} - (n-1)} \tag{2}$$

---

[17] The size and percentage coverage of a network is an open issue in SNA, see Ter Wal and Boschma (2009) and Maggioni and Uberti (2010) for further details.

[18] For connected networks, the range is between 0 (i.e. the network is minimally connected, meaning that the removal of just 1 link would render it disconnected) and 1 (i.e. the network is complete, meaning all possible links already exist).

Degree centralization index (C_{deg}) is a measure of the variance in the degree centrality values of the nodes in a given network (Freeman, 1979), expressed analytically as:

$$C_{deg} = \frac{\sum_{i=1}^{n} [C_{deg}(a^*) - C_{deg}(a_i)]}{n^2 - 3n + 2} \tag{3}$$

where $C_{deg}(a^*)$ and $C_{deg}(a_i)$ respectively are degree centrality index (i.e. the number of a node's direct links) of the most central node, and degree centrality of a generic node *i*. As in the case of *d*, the index ranges from 0 (i.e. all nodes have the same degree centrality index) to 1 (i.e. there is one node that connects the entire network). This index, therefore, measures the hierarchization in the network and the presence (or absence) of pivotal node(s), i.e. a node(s) with a direct relation to most of the other nodes in the network.

While these two network indexes occur frequently in early applications of SNA to sociological analysis, *CC*, *APL* and δ were recently introduced by mathematicians, physicists and computer scientists (Strogatz, 2001; Albert and Barabasi, 2002; Newman, 2001 and 2003). These indexes (and the underlying degree distribution) are computed to classify large and complex structures (i.e. networks characterized by thousands of nodes) with reference to standard ideal types (e.g. random, regular, scale-free or small world structures).

The clustering coefficient of node *i* (CC_i) is the fraction of the existing links connecting a node's neighbours with one another, out of the maximum possible number of such links. Its average across all nodes in a network (*CC*) summarizes the extent to which the nodes in a graph tend to group together (Watts, 1999). More formally:

$$CC_i = \frac{\Lambda_i}{v_i} \quad \text{and} \quad CC = \frac{\sum_{i=1}^{n} \frac{\Lambda_i}{v_i}}{n} \tag{4}$$

where Λ_i is the number of edges in the neighbourhood of node *i*, v_i is the total number of possible edges of node *i* and *n* is the total number of nodes in the network. CC_i and *CC* vary between 0 (i.e. no neighbour of any vertex is adjacent to any other neighbour of vertex *i*) and 1 (i.e. a node's neighbours are also neighbours of each other).

Average path length (*APL*) is a measure of the degree of separation between two nodes along the shortest path of their intermediaries, i.e. the average number of intermediaries between all pairs of actors in the network (Uzzi and Spiro, 2005). If *APL* is low, it indicates that the actors in the network are close together and flows across the network are easy. Another measure of average distances in the network is diameter (δ), which measures network connectedness and corresponds to the maximal distance between any pair of nodes, i.e. the longest geodesic path, and therefore will be a positive number greater than 1. In a disconnected network, δ is equal to infinity, but it can be computed for connected components.

To classify the topological structure of a network it is usual to compare the indexes of the 'actual' or 'real' network, to an equivalent (i.e. with the same number of nodes and edges as the real network) random network (Watts, 1999; Uzzi and Spiro, 2005)[19],[20]. A 'small world' network is a network in which most nodes are not neighbours of one another, but are separate only by a small number of steps.[21] Watts (1999) synthesizes some of the structural features of the topology comparing a given network's *APL* and *CC* with the same indexes calculated for an equivalent random network.

In this paper we check whether the network of Italian economists is structured according to a 'small world' typology by computing the 'small world quotient' (Uzzi and Spiro, 2005), Q_{sw}, as follows :

$$Q_{sw} = \frac{\dfrac{CC_a}{CC_r}}{\dfrac{APL_a}{APL_r}} \tag{5}$$

where subscripts *a* and *r* respectively indicate the *actual* and the equivalent *random* networks' *CC* and *APL*, in terms of average degree and density. The greater the Q_{sw}, (and particularly if the quotient is greater than 1), the closer the structure of the network to a 'small world' structure.[22]

In order to investigate the evolution of the co-authorship network, we also study the relationship between network structure at time *t* and its evolution over subsequent periods. Albert and Barabasi (2002) find that the evolution of scale-free networks[23] generally follows a 'preferential attachment' process where central nodes become increasingly central because new nodes tend to establish proportionally more links with more central nodes.

In the following analysis we calculate $R(k)$, the relative probability of establishing new co-authorships, as follows:

[19] It is useful to recall that a large random network (Erdős and Renyi, 1959), is characterized by a binomial degree distribution, while *CC* – which depends on the size of the network – is equal to the ratio of average degree and number of nodes, and *APL* and δ depend on the size of the network structure and its average degree (Albert and Barabasi, 2002).

[20] Similarly it is possible randomly to remove some nodes in order compare the main component of the real network with the equivalent random ones (Maggioni and Uberti, 2009).

[21] In a 'small world' structure, the path between two nodes in very large network, may be extremely short due to the existence of bridging agents. This means that even in very large and locally clustered networks efficient and fast information diffusion is possible.

[22] In the paper we compute Q_{sw} for the whole network (from 1969 to 2006) as well as for the last two sub-periods when a main component emerged in the network structure.

[23] These networks are characterized by a very skewed degree distribution with very few pivotal nodes at the centre, and a large number of peripheral nodes

$$R(k) = \frac{proportion\ \ of\ \ new\ links\ to\ nodes\ of\ \ degree\ k}{proportion\ \ of\ \ nodes\ with\ degree\ k_i} \tag{6}$$

If there is no preferential attachment, R(k) is equal to 1, and describes a growth process in which new authors enter the co-authorship network by establishing links to existing authors randomly. If R(k) is greater than 1, then growth follows a preferential attachment process because more central authors are more attractive to new coauthors.

For our node indexes are concerned, we can compute three centrality indexes, degree, betweenness and closeness centrality, and the clustering coefficient for each node, as defined above.

Degree, betweenness and closeness centralities are calculated for every node, as defined by Freeman (1979). The degree centrality[24] index is a measure of the number of direct links connecting a given node, which in our case is the number of co-authors of each individual economist, which measures the local centrality of the economist.

Betweenness centrality is a measure of the number of times a vertex occurs on all geodesic (i.e. shortest) paths within a network connecting every nodes to every other nodes. This index identifies the strategic value of a node and its potential ability to control the relations of a network. Analytically:

$$C_b(a_i) = \sum_{j<k}^{n} \frac{g_{jk}(a_i)}{g_{jk}} \tag{7}$$

where $g_{jk}(a_i) / g_{jk}$ is the estimated probability of connecting two nodes and the numerator represents the number of geodesic distances connecting nodes *j* and *k* containing node *i*, while the denominator does not necessarily contain node *i* (Wasserman and Faust, 1994).

Closeness centrality is an inverse function of the geodesic distances from one node to every other node in the network – hence it depends on both direct and indirect links – and reflects the efficiency of the communication channels of a given node with the rest of the network. Analytically:

$$C_c(a_i) = \left[\sum_{j=1}^{n} d(a_i, a_j) \right]^{-1} \tag{8}$$

where $d(a_i, a_j)$ is the number of lines in the geodesic path linking nodes *i* and *j* (Wasserman and Faust, 1994).

[24] On the basis of the degree centrality index of each node, we computed a network index, average degree (av_deg), as the mean value of the degrees (i.e. direct links) of each node in the network, in order to enable comparison of the collaborative behaviours of Italian academic economists in different time periods.

IV.3. SNA of Italian economists

The second aim of this study requires the analysis of co-authorship patterns using SNA techniques, in order to detect whether and how co-authorship behaviour has changed over time, and whether the structure of co-authorship is similar to one of the ideal typical network topology.

First, we can highlight that collaboration is evolving over time, and is becoming a more common phenomenon of Italian economists' scientific behaviour. For example, the number of co-authored JA has increased more than the number of single authored JA. In 1969, 20 out of 24 articles recorded in Econlit were single authored; in 2003, the numbers were 272 out of 540 while in 2006, 328 out of 567 articles were multi-authored.

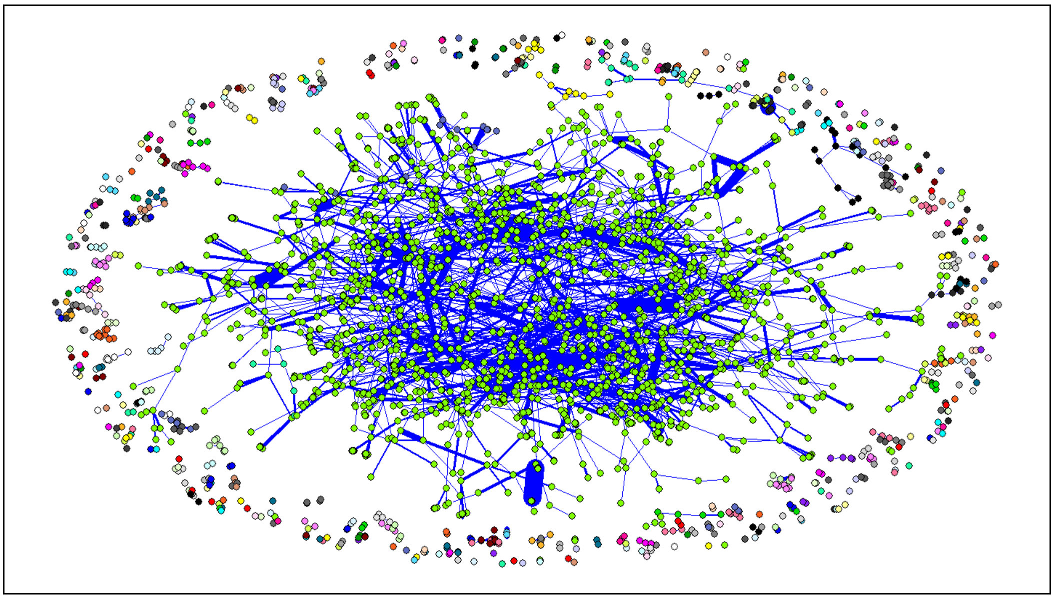
The structure of the whole period network is very complex (see Figure 4 and Table 2). In a network of 2,972 economists there is a main component, i.e. the largest group of connected nodes, which includes 2,061 economists, several of whom wrote more than one article (represented by the thicker lines between nodes), while the number of isolated nodes (i.e. economists always writing alone) is quite small, just under 10%. There are several sub-groups of different dimensions in the whole network, which represent small communities of economists (i.e. discipline-bounded, geographically delimited).

Table 2 shows the results for a network dichotomized according to a threshold value equal to 1. This procedure identifies collaborations for at least one JA, which means that multiple collaborations with the same co-author do not determine the structure of the network.

The density value shows that the network is quite sparse: density value is very low for both the whole network and the main component. The centralization measure indicates a fairly non-hierarchical network, with no stars in a pivotal role, in the whole network or in the main component. Diameter δ indicates that although the network is quite complex, the biggest distance between two connected economists is 30, which is a relatively small value.

The average degree values show that Italian economists collaborate with just over two people in the whole network, and around three people in the main component.

Figure 4: The co-authorship network of Italian economists (1969-2006)



Note: colours identify different sub-groups.

There are some other interesting features that emerge when we compare different groups of economists, especially subset **P** with subsets **F** and **O,** the groups of foreign and 'other Italian' economists. Firstly, group **P** economists, in this work, represent 44% of the whole network and 37% of the main component.

If we consider the international openness of Italian economists in terms of the nationalities of their co-authors, we find some interesting results. Foreign economists accounts for 38% of the network and this percentage is 46% for the main component. This suggests an overrepresentation of foreign economists in the most connected sub-community of co-authors.

Table 2: Basic SNA statistics of the whole network

threshold value = 1

INDEX	Whole network	Main Component
N	2,972	2,061
d	0.000824	0.001469
APL	8.292	8.296
δ	30	30
CC	0.569	0.552
C_{deg}	0.0103	0.0146
Isolated nodes	283	…
Isolated diads	113	…
Isolated triangles	43	…
Sub-networks with 4-9 nodes	40	…
Sub-networks with 11-15 nodes	4	…
Av_deg	2.447	3.026
Sd	2.841	3.144
Min Degree	0	1
MAX Degree	33	33

Since cooperation behaviour is changing dramatically over time, we can identify four distinct periods from 1969 onwards, which demonstrate how network features are changing (see Table 3 for basic SNA statistics).

In the first period (1969-1976) no particularly cohesive network structure can be identified, and the network is mostly disconnected. In the second period (1977-1986), the network structure is very disconnected, and we can identify several isolated nodes (i.e. economists writing alone). In both these networks the Italian economist community is not very open to collaboration with foreigners and is mainly representative of the selected scientific sectors. In both networks average degree is less than 1, and more habitual co-authorship among the Italian economists community does not emerge until the late 1980s. The structure of the networks changes radically in the last two periods: (1987-1996) and (1997-2006). A main component emerges, while the number of isolated authors remains fairly constant; average degree increases and APL and diameter increase dramatically.

The values for network density (both absolute and relative) are very low and are continuously decreasing over time.[25] This does not contradict the increasing values for average degree since, even if scientific collaborations increase over time, the increase in these values is less than the increase in the number of possible links in a growing network.[26]

Table 3: SNA statistics for different periods

threshold value = 1

INDEX	Period 1 (1969-76)	Period 2 (1977-86)	Period 3 (1987-96)		Period 4 (1997-2006)	
N	159	580	1,094		2,424	
size of MC	0	0		214		1380
MC/N	-	-	19.56%		56.93%	
Size of P	126	445	692	89	1.176	570
Size of F	25	60	227	68	619	389
Size of O	8	75	175	57	629	421
d	0.0070	0.0014	0.0012	0.0120	0.0010	0.0022
dr	-0.0056	-0.0021	-0.0006	0.0027	0.0001	0.0008
APL	1.328	1.751	8.180	8.533	8.282	8.295
δ	3	5	21	21	20	20
CC	0.458	0.606	0.535	0.531	0.596	0.564
C_{deg}	0.0221	0.0108	0.097	0.0447	0.0098	0.0167
Isolated nodes	92	293	295	…	263	…
n. isolated diads	34	55	88	…	114	…
n. isolated triads	5	16	43	…	46	…
Av_deg	0.553	0.786	1.364	2.561	2.328	3.041
Min degree	0	0	0	1	0	1
Max degree	4	7	12	12	26	26
St. dev.	0.774	1.147	1.498	1.932	2.459	2.830

[25] Relative density is negative, signalling the existence of isolated nodes.

[26] We should remember that the denominator of density increases non-linearly ($n*n-1$) with network size n.

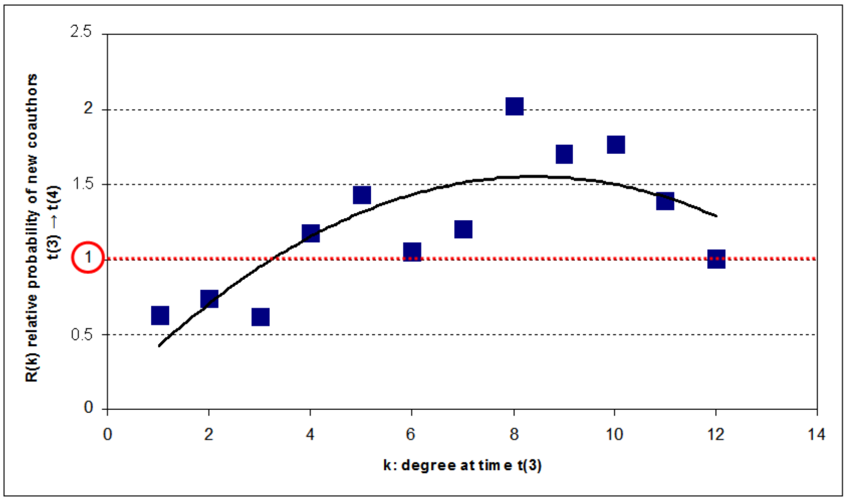
Table 4 presents the Q_{sw} index to detect the presence of a small world topology in the network structure of the Italian economist community. The value index (Q_{sw}) means that this network has the features typical of a small world topology. Since the network is changing over time, and collaborative behaviour is increasing, we calculate this index for all the periods in which a main component is present[27]. All indexes are higher than 1, confirming the existence of a 'small world phenomenon' in the network of Italian economists.

Table 4: The 'Small World' of Italian economists

	Total (1969-2006)	3rd period (1987-1996)	4th period (1997-2006)
N.	2061	214	1380
APL_a	8.296	8.533	8.295
CC_a	0.552	0.531	0.564
APL_r	6.739	5.349	6.383
CC_r	0.001	0.007	0.001
$\boldsymbol{Q_{sw}}$	**448.40**	**47.55**	**434.00**

Finally, to identify the network topology we investigate the presence or not, of preferential attachment. Figure 5 describes the existence of a non-linear dynamics of preferential attachment in the main component of Italian economists, passing from period 3 to period 4. As already explained, a value greater than 1 signals the existence of a preferential attachment dynamics in the growth of the network, with more central nodes receiving a more than proportional share of new links. This is the case for the Italian economist community in the two most recent decades. Authors with more than four co-authors in the 3rd period tend to show an increase in the number of their co-authors. This may be due to a number of reasons, including that: (i) previous co-authorship signals both a willingness and ability of an author to cooperate with other scientists; (ii) scientists tend to look for this type of person when looking for potential co-authors.

Figure 5: Preferential attachment in the main component of the Italian economist network



[27] Hence not for periods 1 and 2 (see table 3).

The depiction in Figure 5 suggests that the increasing returns to co-authorship are not linear; in particular, when the number of co-authors reaches 8 the relative probability of additional co-authors begins to decrease and becomes equal to 1 for twelve 12 co-authors in period 3. This may be explained by an excessive level of transaction and communication costs for a subgroup of more than 12 co-authors, and a preference for stability in co-authorship relations, as described in Section V. We also calculate the distribution of degrees, and find that it is very skewed, and very different from a normal distribution, which confirms the wide variation in the collaborative behaviour of Italian economists and suggests that the network topology is more similar to a scale-free than a random network.

V. THE ECONOMETRIC ANALYSIS

We use different econometric methods to identify empirically the determinants of the probability to publish, and of the scientific productivity of economists working in Italian universities. As stated in the introduction, and discussed extensively in Section II, we expect that these determinants will have three distinct dimensions:

- an *attributive* dimension related to the individual characteristics of the researcher (measured by variables for gender, academic position, tenure, location);
- a *relational* dimension representing the researcher's connection to his/her scientific community (measured by variables such as the propensity for co-authoring, propensity to cooperate with foreign researchers);
- a *positional* dimension, which relates to the role of the researcher in the scientific community (measured by network variables such as betweeness, closeness centrality index and clustering coefficient).

In particular: in Section V.1 we identify the *attributional* determinants of the probability to publish using a probit model; Section V.2.1 investigates the *relational* driving forces of an individual researcher's scientific productivity, through an Instrumental Variables (IV) estimation strategy; Section V.2.2 focuses on the impact of the *positional* characteristics of the researcher within the scientific community using OLS methods and analysing the sole main component of the network.

V.1. The determinants of the probability to publish

To identify the *attributional* determinants of the probability to publish, we estimate a maximum-likelihood *probit* model as follows:[28]

$$\Pr\left(publish_i = 1/X\right) = \Phi\left(X_i^{'}\beta\right) \tag{9}$$

[28] We actually estimate a *dprobit* model. Rather than reporting coefficients, a *dprobit* reports the marginal effects, i.e. the change in the probability for an infinitesimal change in each independent, continuous variable and, by default, reports discrete change in the probability of the dummy variables.

where Φ is the cumulative distribution function of the standard normal distribution, $publish_i$ is a dummy variable that is equal to 1 if the economist has at least published one journal article (JA) in Econlit in the period 1969-2006, and 0 otherwise, and X indicates the regressors. These variables, constituted by a set of attributional dummy variables, are added stepwise in the econometric specifications, and are available for all the economists in our 2006 population. They include: (i) a dummy for gender (*Gender*) equals to 1 if the academic is male and 0 otherwise; (ii) a dummy (*Tenured*) if the economist has a tenured academic position; (iii) four dummies for economist's disciplinary groups (*Economics*, *Public Economics*, *Econometrics*, and *Others*[29]); (iv) a dummy indicating whether the economist works in a Faculty of Economics and 0 otherwise (*Fac_economics*); and (v) five geographic dummies (*North West*, *North East*, *Centre*[30], *Islands* and *South*) describing the location of the economist's institution. Finally, we introduce a dummy (NT*Lecturer*), that is equal to 1 if the economist is a non-tenured lecturer[31] and 0 otherwise.

Table 5 reports the main findings of the analysis. The results suggest that the probability that the economist will publish is positively influenced by gender, geographic location of the university, and scientific sub-sector. These findings shows that men tend to be more productive than women and economists working in Northern universities have a higher probability to publish than those located in Sothern universities; tenure has a negative impact on the probability to publish at least one Econlit JA. Finally, the dummy indicating the position of "untenured lecturer" – as expected – is negative and statistically significant. This is because the youngest member of the Italian academic community may be will awaiting publication for already submitted papers. In this first econometric exercise, this variable is the only one that allows us to control for the "age" of an individual economist.

[29] Used as the reference and not included in the regression.
[30] Used as the reference and not included in the regression.
[31] In order to take into account the "beginning of academic career" effect.

Table 5 – Determinants of the probability to publish

Dependent variable: at least one JA in 1969-2006, 0 otherwise

	[1]		[2]		[3]		[4]	
	dF/dx	*t-values*	*dF/dx*	*t-values*	*dF/dx*	*t-values*	*dF/dx*	*t-values*
Gender	0.086**	3.80	0.086**	3.81	0.086**	3.80	0.075**	3.41
Tenured	-0.041**	-2.01	-0.038*	-1.88	-0.039*	-1.94	-0.163**	-6.10
Economics	…	…	0.064**	2.22	0.066**	2.27	0.080**	2.78
Econometrics	…	…	0.114**	2.36	0.112**	2.30	0.092*	1.81
Public Econ.	…	…	0.006	0.18	0.008	0.24	0.027	0.80
Others	…	…	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Fac_Economics	…	…	…	…	0.038**	1.96	0.036*	1.90
North West	…	…	…	…	…	…	0.042*	1.76
North East	…	…	…	…	…	…	0.055**	2.13
Centre	…	…	…	…	…	…	Ref.	Ref.
South	…	…	…	…	…	…	-0.039	-1.33
Islands	…	…	…	…	…	…	-0.170**	-4.05
NTLecturer	…	…	…	…	…	…	-0.345**	-7.14
N. Obs.	1,620		1,620		1,620		1,620	
Pseudo R^2	0.011		0.018		0.020		0.070	
Obs. P	0.812		0.812		0.812		0.812	
Pred. P	0.815		0.817		0.817		0.834	

Note: the regressions also include a constant term. Standard errors are robust to heteroskedasticity.
Legend: ** significant at 5%; * significant at 10%

V.2. The determinants of the scientific productivity

In this section we focus on the roles first of the relational and then the positional variables in the scientific productivity of Italian economists. In both cases, an economist's scientific productivity[32] – the dependent variable in these models – is measured as the total number of Econlit JA divided by his/her "seniority". Since we do not have information on the actual age of the economists in our sample, we use the proxy variable of "seniority", computed as the difference between the year of his/her first Econlit publication and 2006.[33]

V.2.1. Role of the relational variables in scientific productivity: full sample

In terms of *relational* driving forces, Italian economist's individual productivity is explained by two types of variables: (i) propensity to co-author and (ii) propensity to cooperate with "foreign" researchers. The propensity to co-author (*prop_coauth*) is measured as the proportion of co-authored articles on the total number of articles in the Econlit database. This variable ranges between 0 and 1, where 0 is no collaboration and 1 indicates that all papers are co-authored.

[32] Note that we do not correct for number of co-authors of an article, which means we do not capture the individual contributions' of each economist in order not to introduce arbitrary corrections in our regressions. To introducing corrections related to the number of authors would give more weight to papers in certain research fields such as economic and econometric theory, history of economic thought, economic history, etc. where the number of co-authors tends to be systematically lower than in applied economics, for example.

[33] E.g. for an individual *j* whose first article is an Econlit indexed journal was in 1985, "seniority" is calculated as 2006-1985 = 21.

Foreign measures the proportion of collaboration with economists not affiliated to an Italian university, i.e. set **F** defined in Section IV.1, and is a proxy for the economist's level of "external" connections. *Foreign* is calculated as the proportion of non-MIUR economists in total co-authors.[34] Note that this variable captures the degree of cooperation with economists affiliated to foreign institutions or co-authors from Italian non-academic institutions (such as Banca d'Italia, ISTAT, CNR, etc.).

We control for *attributional* variables shown to be significant in the previous econometric exercise (i.e. gender, tenure, geographical location, see Table 5). We control also for age class. Since the 1970s, the university system in Italy has experienced major changes to the career advancement system, research assessment, propensity for internationalization, etc., we try to capture these institutional changes through four time dummies to indicate the date of the first Econlit article (1997-2006; 1987-1996; 1977-1986 and 1969-1976): *Age97_06*, *Age87_96*, *Age77_86*, and *Age69_76*. These dummies control for the different "publication regimes" applying to the economists in our sample.

The econometric analysis involves two econometric problems: (i) sample selection and (ii) endogeneity. To address these simultaneously, we estimate a two-stage model. In the first stage, we eliminate the selection bias problem. Selection bias may occur because the average characteristics of the economists in the sample of publishing authors differ from those of the whole population. Without addressing this selection effect, the statistical association between scientific productivity and co-authorship will be inferred incorrectly because the impact of cooperation might be confounded with the coefficients determining the selection. To model the selection mechanism to enable adjustment of the parameter estimations in the structural equation, we adopt a Heckman (1979) procedure. This method uses all the available observations to estimate a *probit* model of the selection indicator. The residuals of this regression are then used to construct a selection bias control factor λ_i: the inverse Mill's ratio which, for each individual, can be computed as:

$$\lambda_i = \frac{\phi_i(X)}{\Phi_i(X)} \tag{10}$$

where X denotes the covariates used in the selection process, ϕ_i is the density probability function, and Φ_i is the cumulative probability function. This factor, which accounts for the effects of all the unmeasured characteristics related to the selection variable, is then introduced into our structural equation as an extra explanatory variable. Operationally, we estimate the inverse Mill's ratio from the fourth specification reported in Table 6.

At the second stage, we estimate a *structural equation* given by:

[34] Calculated to distinguish each person's identity.

$$prod_i = \alpha_o + \alpha_1 prop\_coauth_i + \alpha_2 foreign_i + X_i^{'}\beta + \theta\lambda_i + \upsilon_i \qquad (11)$$

where $prod_i$ is the scientific productivity of each economist *i*, $prop\_coauth_i$ and $foreign_i$ denotes the respective propensity to co-author and to cooperate with foreign researchers; *X* indicates other covariates such as *Gender*, *Tenured*, and controls such as *North West*, *North East, Centre, South, Islands, Age97_06, Age87_96, Age77_86*[35], *Age69_76*; λ_i is the inverse Mill's ratio computed from the previous selection equation and is used to control for selection bias. Finally, υ_i is the error term with the usual statistical properties.

The direction of causality is another important issue in analysing the relationship between scientific productivity and co-authorship. While it has been shown that cooperation affects productivity, it may well be the case, as suggested by anecdotal evidence, that productivity might affect cooperation: i.e., very productive economists are seen as potentially better co-authors, thus generating a classic reverse causality problem. Also, the presence of endogeneity – i.e., one or more explanatory variables correlated to the true (but unobserved) error term – can generate biased and inconsistent OLS estimates of the coefficient under investigation. We deal with this problem by adopting an IV strategy. The variable that satisfies these conditions is the number of collective-volume articles (CVA) – i.e. chapters in edited books – authored by each economist. CVA are used as an instrument for the propensity for co-authorship because the most recent literature and university policy recommendations assume CVA are the effect of personal connections and relational attitudes which, while not comparable in quality with JA, may reflect alternative use of a researcher's time. If a researcher is contributing to a collected volume, this leaves less time to writing JAs. Thus CVA may measure the propensity to cooperate and interact with the wider scientific community irrespective of the impact in terms of the most known bibliometric indexes of scientific productivity.

Under the null hypothesis that the model is appropriately specified with all explanatory variables exogenous, both Hausman tests reject the null hypothesis of exogeneity of the co-authorship propensity variable, which suggests that IV method is the most appropriate method to estimate our model (see the p-value of the Hausman test in Table 6).

Finally, to account for the introduction as regressor of the inverse Mill's ratio, we compute standard errors using bootstrap methods (with 50 replications). Note that the inverse Mill's ratio variable is always negative[36] and often statistically significant. This suggests that correcting for sample selection is the right choice.

[35] Used as the reference and not included in the regression.

[36] When the coefficient on the inverse Mill's ratio is negative this means that there are unobserved variables increasing the probability of selection and the probability of a lower than average score on the dependent variable.

Table 6 reports the main results. The findings suggest that co-authorship is a significant determinant of scientific productivity. In fact, the coefficient of the variable measuring the propensity for co-authorship is positive and statistically significant. This means that economists that are more collaborative are also more productive in JA terms. The results underline another positive role of networking outside the MIUR community. Collaboration with foreign and Italian non-MIUR economists has a positive impact on scientific productivity. Belonging to an international network can be interpreted as "signalling" the intrinsic "quality" of the economist and his/her positive political and social attitudes to forging scientific relationships with foreign groups. The liaison with members of the editorial boards of international journals might also play an important role in these processes.

Table 6 – Determinants of scientific productivity: complete sample

Dependent variable: scientific productivity

	[1]		[2]	
	Coeff.	*t-values*	*Coeff.*	*t-values*
Propensity of coauthorship	5.686**	5.21	2.912**	3.40
Foreign	…	…	0.383**	2.25
Gender	0.404**	2.64	0.327**	3.79
Tenured	0.196	1.47	0.126	1.30
North West	-0.227	-1.40	-0.065	-0.56
North East	-0.373**	-2.05	-0.295**	-2.52
Centre	Ref.	Ref.	Ref.	Ref.
South	-0.092	-0.49	0.005	0.05
Islands	-0.156	-0.45	-0.086	-0.53
Age97_06	-0.326	-1.35	-0.052	-0.26
Age87_96	-0.310*	-1.55	-0.055	-0.46
Age77_86	Ref.	Ref.	Ref.	Ref.
Age69_76	0.475**	2.20	0.134	0.84
Inverse Mill's Ratio	-0,581	-1.07	-1.359**	-3.61
N. Obs.	1,317		1,015	
Hausman (p-value)	0.000		0.000	

Note: the regressions also include a constant term. Standard errors are bootstrapped (50 replications).
Legend: ** significant at 5%; * significant at 10%

V.2.2. The role of positional variables in scientific productivity: main component only

Finally, we test for the role of some positional variables characterizing the main component (MC), in explaining the scientific productivity of Italian economists, defined as before.

These variables, derived from the SNA exercise presented in Section IV[37], are: (i) the betweenness centrality index (*Betweennees*), which measures the strategic influence of an author with special reference to his "bridging role" in relation to different academic groups; (ii) closeness centrality index (*Closeness*), which measures the global centrality of a researcher, i.e. his/her capacity to reach all other researcher through the lowest number of co-authors; (iii)

[37] These values are calculated exclusively on the MC of the co-authorships network dichotomized according to a threshold value greater than zero.

clustering coefficient[38] (*CC*), which measures the proportion of researcher's co-authors who also co-author with one another[39]; (iv) a measure for the stability of the scientific cooperation (*Stability*) which ranges between 0, if all co-authors are different, and 1, if all co-authors are the same. We also introduce a set of covariates such as *Gender* and *Tenured*, and controls such as *North West, North East, Centre, South, Islands, Age97_06, Age87_96, Age77_86, Age69_76*.

We adopt the econometric approach described previously with the only difference that in this case the estimator is a simple OLS. This choice is justified by the fact that here endogeneity is less a problem because all network indexes refer to second order network features, which are not easily observable by the individual author when deciding on his/her co-authorship strategy.[40] The main findings of our analysis are presented in Table 7.

Table 7 – Determinants of scientific productivity (main component)
Dependent variable: scientific productivity

	[1]		[2]		[3]		[4]	
	Coeff.	*t-value*	*Coeff.*	*t-value*	*Coeff.*	*t-value*	*Coeff.*	*t-value*
CC	-0.183**	-3.15	-0.248**	-4.39	-0.095	-1.58	-0.130**	-2.12
Foreign	0.440**	3.46	0.449**	3.50	0.355**	3.15	0.368**	3.82
Stability	0.290**	8.85	0.303**	8.38	0.307**	7.85	0.310**	9.76
Closeness	…	…	2.881**	7.02	…	…	1.055**	2.90
Betweenness	…	…	…	…	4.198**	13.55	3.680**	9.17
Gender	0.277**	4.90	0.241**	5.00	0.242**	5.58	0.234**	4.84
Tenured	0.010	0.18	-0.017	-0.29	-0.044	-0.83	-0.047	-0.87
Geogr. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age dummnies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mill's Ratio	-0.453**	-2.09	-0.224	-1.00	-0.250	-1.05		
N. Obs.	661		661		661		661	
Adj. R^2	0.248		0.320		0.403		0.409	

Note: regressions also include a constant term. Standard errors are bootstrapped (50 replications).
Legend: ** significant at 5%; * significant at 10%.

The variables referring to researcher centrality (i.e., *Closeness* and *Betweenness*) are positive and statistically significant. Centrality in the co-authorship network boosts scientific productivity. However, if scientific productivity is the researcher's main goal, then acting as the "bridge" between different scientific sub-sectors and/or different "schools" and academic groups is more beneficial than being globally central. *Clustering* is negative and in most cases statistically significant, thus signalling the interests of "star" authors in hindering the

[38] This value is calculated for each individual based on the aggregated network of the Italian economists community for the whole period 1969-2006.
[39] In general SNA terms, if the neighbours of a single node are also neighbours of each other (i.e. if the clustering coefficient of that node is high), these neighborsneighbours are not reliant on the single node for their connection. Therefore, the intermediate node is completely needless for the connection between these two neighbors.
[40] This is one of the reasons for omitting the most common measure of centrality (degree centrality index) in our case equal to the number of previous coauthors, which is easily observed by the individual researcher. The other reasons are multicollinearity with other measures of centrality.

interactions among their co-authors in order to preserve a hierarchical and productive “hub and spoke” co-authorship structure.

Although there are plenty of theoretical reasons for and against frequent changes of co-authors (based on the trust-building vs intellectual novelty trade-off), our results shows that, at least for Italian economists, *Stability* pays. Keeping the same group of authors for successive JA, is the best strategy to improve scientific productivity.

VI. CONCLUSIONS

This paper presents a study of the collaborative behaviour (co-authorship) of a community of 1,620 economists working in Italian universities during the period 1969-2006, to examine whether this behaviour has a (relevant) effect on their scientific productivity. To achieve this we conducted a series of analyses, using an original database built by merging the CINECA-MIUR personnel dataset with the Econlit bibliographic dataset. SNA techniques allowed us to study the structure and evolution of co-authorship among Italian economists and to derive positional and relational data for each scientist, exploited in several econometric exercises to explain the variance in the scientific productivity (measured by number of journal articles published per year) of this community.

In terms of the structure of co-authorship, we found that collaborative behaviour has evolved over time and has become more frequent within this scientific community. The network of collaboration is complex (composed of many sub-networks) and is characterized by low density although since the mid 1980s the percentage of economists in the main component increased from about 20% to over 50% of the entire population. It is interesting also that the percentage of foreign and/or non-academic economists is higher in the main component than in the whole network, which is a sign of a highly cooperative attitude among these economists. There are several criteria that might explain the structure and composition of specific sub-networks (such as scientific discipline, geographical location, school of thought), but the overall structure of the main component is similar to small world structure, and the evolution of the network through time is guided by a non linear preferential attachment mechanism.

For the determinants of scientific productivity among Italian economists, the econometric analysis shows that ‘attributional’ (age, gender, academic position, tenure, scientific sub-discipline, geographical location), ‘relational’ (propensity to cooperate and stable cooperation patterns) and ‘positional’ (betweenness and closeness centrality indexes and clustering coefficient) variables matter. The econometric results show that the individual productivity of an Italian economist depends (among other factors) on his/her propensity to collaborate, his/her ‘international’ connections and the stability of his/her collaborative behaviour.

Finally, we found that the position of an individual economist in co-authorship networks affects his/her scientific productivity. Being ‘central’ increases the number of scientific publications per unit of time, but being a ‘bridge’ (i.e. connecting two almost separated parts of a network) is even more beneficial. It seems that maintaining the same set of co-authors over the years is the best strategy – at least for the community of Italian economists – to improve individual productivity.

References

Acedo F. J., Barroso C., Casanueva C., Galàn J. L., (2006), Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis, *Journal of Management Studies*, 43: 957-983.

Albert R., Barabasi A. L., (2002), Statistical Mechanics of Complex Networks, *Reviews of Modern Physics*, 74: 47-99.

Barnett, A.H., Ault, R.W., Kaserman, D.L., (1988), The Rising Incidence of Co-authorship in Economics: Further Evidence, *Review of Economics and Statistics* 70: 539–543

Borgatti S. P., Everett M. G., (1997), Network Analysis of 2-Mode Data, *Social Networks*, 18: 243-269.

Butler R.J.(2007) JRI, JF, and the Internet: Coauthors, New Authors, and Empirical Research, *Journal of Risk and Insurance*, 74: 713-37

Cainelli G., de Felice A., Lamonarca M., Zoboli R. (2006), The Publications of Italian Economists in ECONLIT. Quantitative Assessment and Implications for Research Evaluation, *Economia Politica Journal of Analytical and Institutional Economics*, XXIII: 385-423.

Cosme Costa Vieira P. (2008),. An economics journals' ranking that takes into account the number of pages and co-authors, *Applied Economics*, 40: 853-861

Dolado J. J, Garcia-Romero A., Ramarro G. (2003), Publishing Performance in Economics: Spanish Rankings (1990-1999), *Spanish Economic Review*: 85-100.

Doreian, P., Batagelj, V., & Ferligoj, A. (2004) Generalized blockmodeling of two-mode network data, *Social Networks*, 26: 29-53.

Durden, G. C. and T. J. Perri (1995). Coauthorship and Publication Efficiency. *Atlantic Economic Journal*, 23: 69-76.

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

Freeman L. C. (1979), Centrality in Social Networks. Conceptual Clarification, *Social Networks*, 1: 215-239.

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

Fisher, B. S., Cobane C.T., Vander Ven T. M.. Cullen T.M. (1998) How Many Authors Does it Take to Publish an Article? Trends and Patterns in Political Science. *Political Science and Politics* 31: 847–56.

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

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

Heckman J. J. (1979), Sample Selection as a Specification Error', *Econometrica*, 47:153-161.

Holder M., Langrehr, F., Schroeder D. (2000), Finance Journal Co-Authorship: How Do Co-Authors in Very Selected Journals Evaluate the Experience?', *Financial Practice and Education*, 10: 142-152.

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

Hudson, J. (1996), Trends in Multi-Authored Papers in Economics, *Journal of Economic Perspectives* 10: 153– 158

Kalaitzidakis P., Mamuneas T. P., Stengos T. (1999), European Economics: An Analysis Based on Publications in the Core Journals, *European Economic Review*, 43: 1150-1168.

Laband, D. N. *(*1987*):* A quantitative test of journal discrimination against women, *Easter Economic Journal*, 13: 149–54

Laband D., Piette M. J. (1995), Team Production in Economics: Division of Labour or Mentoring, *Labour Economics*, 2: 33-40.

Laband D., Tollison R. (2000), Intellectual Collaboration, *Journal of Political Economy*, 108: 632-661.

Laband D., Tollison R. (2006), Alphabetized Co-Authorship, *Applied Economics*, 38: 1649-1653.

Lee S., Bozeman B., (2005), The Impact of Research Collaboration on Scientific Productivity, *Social Studies of Science*, 35: 673-702.

Neary J.P., Mireless J.A., Tirole J. (2003), Evaluating Economics Research in Europe: an Introduction, *Journal of the European Economic Association*, 1: 1239-1249.

Maggioni M.A., (1995) Metodologie reticolari per l'analisi dei sistemi locali di produzione and innovazione, *Economia and Lavoro*, n. 1/2-95: 109-130

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

Maggioni M.A., Uberti E. (2010), Networks and geography in the economics of knowledge flows, *Quality & Quantity*, forthcoming

McDowell, J.M., Melvin, M. (1983), The Determinants of Co-Authorship: an Analysis of the Economics Literature, *Review of Economics and Statistics* 65: 155-160.

Medoff, M.H. (2003), Collaboration and the Quality of Economics Research, *Labour Economics* 10: 597–608.

Neary J.P., Mireless J.A., Tirole J. (2003), Evaluating Economics Research in Europe: An Introduction, *Journal of the European Economic Association*, 1: 1239-1249.

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

Newman M.E.J. (2003), The Structure and Function of Complex Networks, *SIAM Review*, 45: 167-256.

Piette M. J., Ross K. L. (1992), An Analysis of the Determinants of Co-Authorship in Economics, *Journal of Economic Education*, 23: 277-283.

Sauer R. (1988), Estimates of the Returns to Quality and Co-authorship in Economic Academia, *Journal of Political Economy*, 96: 855-866.

Sigelman L. (2009) Are Two (or Three or Four … or Nine) Heads Better than One? Collaboration, Multidisciplinarity, and Publishability, *Political Science & Politics*, 42: 507-512

Strogatz S. (2001), Exploring Complex Networks, *Nature*, 410: 268-276.

Sutter M., Kocher M. (2004), Patterns of Co-Authorship among Economists in the USA, *Applied Economics*, 36: 327-333.

Ter Wal, A.L.J., R.A. Boschma (2009), Applying social network analysis in economic geography: framing some key analytic issues, *Annals of Regional Science*, 43: 739-756

Uzzi B. Spiro J. (2005), Collaboration and creativity: The small world problem. *American Journal of Sociology*, 11: 447-504.

van der Leij M. J.,Goyal, S. (2006), Strong Ties in a Small World, *Tinbergen Institute Discussion Paper* N. 2006-008/1, Amsterdam, Tinbergen Institute.

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

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

Appendix

Table A.1: Academic Position, scientific sector and gender of Italian Economist as at December the 31st 2006

Scientific Secor (SECSP)	Gender	P	L	SL	AP	CAP	STP	FU	Total
POLITICAL ECONOMY (01)	Female	1	45	52	22	41	13	27	201
	Male	10	79	53	69	102	54	242	609
- Total 01		**11**	**124**	**105**	**91**	**143**	**67**	**269**	**810**
ECONOMIC POLICY (02)	Female	1	10	30	10	24	10	13	98
	Male	2	27	24	21	46	15	110	245
- Total 02		**3**	**37**	**54**	**31**	**70**	**25**	**123**	**343**
PUBLIC ECONOMICS (03)	Female	…	11	15	5	6	4	10	51
	Male	3	16	14	4	22	12	78	149
- Total 03		**3**	**27**	**29**	**9**	**28**	**16**	**88**	**200**
HISTORY OF ECONOMIC THOUGHT (04)	Female	…	…	5	…	1	3	2	11
	Male	1	3	4	3	10	2	13	36
- Total 04		**1**	**3**	**9**	**3**	**11**	**5**	**15**	**47**
ECONOMETRICS (05)	Female	…	3	3	3	2	1		12
	Male	…	5	5	3	6	7	23	49
- Total 05		**…**	**8**	**8**	**6**	**8**	**8**	**23**	**61**
APPLIED ECONOMICS (06)	Female	..	5	10	5	8	1	7	36
	Male	1	10	7	22	28	15	40	123
- Total 06		**1**	**15**	**17**	**27**	**36**	**16**	**47**	**159**
Total Economics	Female	2	74	115	45	82	32	59	409
	Male	17	140	107	122	214	105	506	1211
Total		**19**	**214**	**222**	**167**	**296**	**137**	**565**	**1,620**

Table A.2: Descriptive statistics of dependent varibles and regressors

	N. Obs.	Mean	Std. Dev.	Min	Max
Pro_ja	1,317	0.488	0.420	0.027	4.571
Pub_ja	1,620	0.812	0.390	0	1
Prop_coauth	1,317	0.472	0.349	0	1
Non-MIUR econ.	1,015	0.170	0.289	0	1
Stability	1,317	0.448	0.394	0	1
Clustering	1,317	0.144	0.271	0	1
Betweenness	1,034	0.066	0.217	0	1
Closeness	1,034	0.333	0.357	0	1
Gender	1,620	0.747	0.434	0	1
Tenured	1,620	0.680	0.466	0	1
North-West	1,620	0.251	0.434	0	1
North_East	1,620	0.208	0.406	0	1
South	1,620	0.230	0.421	0	1
Fac_economics	1,620	0.611	0.487	0	1
Economics	1,620	0.711	0.453	0	1
Public econ.	1,620	0.123	0.329	0	1
Econometrics	1,620	0.037	0.190	0	1