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

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

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

@: erika.uberti@unicatt.it
Outline

✓ Related Literature
  • Scientific production
  • Collaborative behavior and scientific production
  • Network Analysis and Network Topology

✓ The datasets:
  • Italian economists in 2006 according to MIUR database;
  • Journal Articles (JA) in 2006 indexed in Econlit.
  • Our dataset

✓ The models

✓ The empirical results

✓ Conclusions and further research
Recently, an increased specialization and the diffusion of scientific collaborations are becoming more and more diffused in the community of social scientists.

In the meantime the practice of evaluating scientific research and personnel through bibliometric indicators is increasingly used by departments, universities, government bodies and funding agencies.

These two phenomena are hardly independent; on the contrary we believe – in so comforted by an extensive stream of literature (see, among others: Sauer, 1988; Barnett et al., 1988; Piette and Ross, 1992; Laband and Piette, 1994; Hudson, 1996; Laband and Tollison 2000 and 2006) that they are strongly interdependent since the increased pressure to publish on academics has caused a rinsing propensity to co-authoring papers due to a series of demand-side and supply-side factors.

In economic terms a generic “scientific production function” is affected by different aspects:
- his/her attributive features (i.e. formation, gender, academic position, etc.);
- his/her relational features (i.e. co-authorships, scientific connections, etc.);
- his/her positional features (i.e. the structural position within the network of scientists of the same field);
Considering a generic “scientific production function” the role of co-authorship is very relevant because affects (demand and/or supply side factors):

the **quantity** of scientific production
- **Greater output and risk-spreading** (Barnett et al., 1988)

the **quality** of scientific production
- **Specialization** (McDowell and Melvin, 1983)
- **Technological complementarities** (Hudson, 1996)
- **Synergies from collaborative work** (Hudson, 1996)

But, especially in social sciences, the **number of co-authors** is very important because there exist

**Increasing/decreasing returns to the number of co-authors** in the same paper
**Increasing/decreasing returns to interactions** with the same co-author(s). (**Stability of interactions**)
The dataset of Italian Economists

We consider:

- a population of **1620 authors** composed by any person in the Cineca-MIUR database holding an **official academic position** in Italian Universities and belonging to one of 6 **economic scientific disciplinary groups** [Economics, SECS-P/01; Economic Policy, SECS-P/02; Public Finance, SECSP/ 03; History of Economic Thought, SECS-P/04; Econometrics, SECS-P/05; Applied Economics, SECS-P/06].

- **8679 journal articles** published between the January 1st 1969 and 31st December 2006 in the journals listed in the Econlit database;
  [Data on publications were downloaded between August 2007 and February 2008, and manually corrected for mistypes in names and double entries].
The database: some descriptive statistics

<table>
<thead>
<tr>
<th>Scientific Field</th>
<th>L</th>
<th>SL*</th>
<th>AP</th>
<th>TAP*</th>
<th>STP</th>
<th>FU*</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECONOMICS</td>
<td>14,0</td>
<td>15,0</td>
<td>10,6</td>
<td>18,5</td>
<td>8,0</td>
<td>34,0</td>
<td>1153</td>
</tr>
<tr>
<td>ECONOMETRICS</td>
<td>13,1</td>
<td>13,1</td>
<td>9,8</td>
<td>13,1</td>
<td>13,1</td>
<td>37,7</td>
<td>61</td>
</tr>
<tr>
<td>PUBLIC ECONOMICS</td>
<td>13,5</td>
<td>16,0</td>
<td>4,5</td>
<td>14,0</td>
<td>8,0</td>
<td>44,0</td>
<td>200</td>
</tr>
<tr>
<td>OTHERS</td>
<td>8,7</td>
<td>13,6</td>
<td>14,6</td>
<td>22,8</td>
<td>10,2</td>
<td>30,1</td>
<td>206</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>214</td>
<td>241</td>
<td>167</td>
<td>296</td>
<td>137</td>
<td>565</td>
<td>1620</td>
</tr>
</tbody>
</table>

* these academic positions are considered as **tenured** positions

*Source*: our calculations on MIUR database at 31 December 2006
Population pyramid of Italian economists

Source: our calculations on MIUR database at 31 December 2006
The dataset of the JA 1/2

M = 1620 people
E = 2972 people
Z = 1317 people, 8679 articles
O = 1655 people

M: MIUR population, “Italian” economists holding an official academic position
E: Econlit population, “Italian” economists (as defined in M) who are authors of publications indexed in Econlit and their co-authors (if any) with any affiliation
Z: Our Dataset “Italian” economists (as defined in M) who are authors of publications indexed in Econlit
N: Non publishing MIUR population, “Italian” economists (as defined in M) without entries in Econlit
O: publishing non MIUR population, co-authors of Z not holding a position in Italian Universities
The dataset of the JA 2/2

**Nodes**
- MODE 1 = 8679 journal articles ranked in Econlit from 1969 until 2006
- MODE 2 = 2972 economists writing papers

**Relational tie**: it appears when two scholars are writing an article together

**Bipartite (Two-mode) network**
- Authors and articles as nodes

**One-mode network**
- Authors as nodes, articles as links

Articles: 1, 2, 3, 4, 5… 8679th article

Authors: a, b, c, d, e, f … 2972nd economist

Authors: a, b, c, d, e, f … 2972nd economist
The results on the SNA of the economic network from 1969 until 2006 divided into 4 sub-periods
The evolution of the co-authorship behaviour
PERIOD 1: 1969-1976
274 articles

Nodes (N) 159
average path length 1.328
diameter 3
CC 0.458
centralization (degree) 0.0221
# diads 17
# triads 5
# subgroups with n>3 4
Isolated nodes 92
Main Component No
Average Degree 0.553
Min 0
MAX 4
Sd 0.774

Nodes (n) are economists, colors indicate n belonging to the same subgroup
PERIOD 2: 1977-1986
1195 articles

Nodes (N) 580
average path length 1.751
diameter 5
CC 0.606
centralization (degree) 0.0108
# diads 55
# triads 16
# subgroups with n>3 23
Isolated nodes 293
Main Component No

Average Degree 0.786
Min 0
MAX 7
Sd 1.147

Nodes (n) are economists, colors indicate n belonging to the same subgroup
**PERIOD 3: 1987-1996**

2309 articles

<table>
<thead>
<tr>
<th>Nodes (N)</th>
<th>1094</th>
</tr>
</thead>
<tbody>
<tr>
<td>average path length</td>
<td>8.180</td>
</tr>
<tr>
<td>diameter</td>
<td>21</td>
</tr>
<tr>
<td>CC</td>
<td>0.535</td>
</tr>
<tr>
<td>centralization (degree)</td>
<td>0.097</td>
</tr>
<tr>
<td># diads</td>
<td>88</td>
</tr>
<tr>
<td># triads</td>
<td>43</td>
</tr>
<tr>
<td># subgroups with n&gt;3</td>
<td>42</td>
</tr>
<tr>
<td>Isolated nodes</td>
<td>295</td>
</tr>
<tr>
<td>Main Component</td>
<td>Yes (214)</td>
</tr>
<tr>
<td>Average Degree</td>
<td>1.364</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
</tr>
<tr>
<td>MAX</td>
<td>12</td>
</tr>
<tr>
<td>Sd</td>
<td>1.498</td>
</tr>
</tbody>
</table>

Nodes (n) are economists, colors indicate n belonging to the same subgroup
PERIOD 4: 1997-2006
4901 articles

Nodes (N) 2424
average path length 8.282
diameter 20
CC 0.596
centralization (degree) 0.0098

# diads 114
# triads 46
# subgroups with n>3 56
Isolated nodes 263
Main Component Yes (1380)

Average Degree 2.328
Min 0
MAX 26
Sd 2.459

Nodes (n) are economists, colors indicate n belonging to the same subgroup
The whole network 1969-2006
2972 nodes writing 8679 articles (E):
1317 M, 1655 O

Nodes (n) are economists, colors indicate n belonging to the same subgroup
The whole network 1969-2006
2972 nodes writing 8679 articles (E):
1317 M, 1655 O

Nodes (N) 2972
average path length 8.290
Diameter 30
CC 0.569
Centralization (degree) 0.01029
# diads 113
# triads 43
# subgroups with n>3 44
Isolated nodes 283
Giant Component Yes (2061)

Average Degree 2.447
Min 0
Max 33
Sd 2.841

Nodes (n) are economists, colors indicate n belonging to the same subgroup
The topology of the economists’ network
Question:
Is the Italian economists network a small world network?
Is the Italian economists’ network a SWN network?

<table>
<thead>
<tr>
<th></th>
<th>real nets</th>
<th></th>
<th>random nets</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>2061</td>
<td>214</td>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>density</td>
<td>0,0015</td>
<td>0,0120</td>
<td>0,0022</td>
<td></td>
</tr>
<tr>
<td>average degree</td>
<td>3,026</td>
<td>2,561</td>
<td>3,041</td>
<td></td>
</tr>
<tr>
<td>average path lengt</td>
<td>8.296</td>
<td>8.533</td>
<td>8.295</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0.552</td>
<td>0.531</td>
<td>0.564</td>
<td></td>
</tr>
</tbody>
</table>

All networks display a SWN behaviour:

1) CC real >> CC random
2) APLreal ~ APLrandom
3) calculate Q index: \[
\frac{CC_{\text{real}}}{APL_{\text{random}}} \div \frac{CC_{\text{random}}}{APL_{\text{real}}}
\]
Existence of preferential attachment \((t_2 \rightarrow t_3)\) for \(k > 3\)

Marginal decreasing returns to co-authorships after 8

\[
R(k) = \frac{\text{proportion of new links to nodes of degree } k_i}{\text{proportion of nodes with degree } k_i}
\]

Under no p.a. \(R(k) = 1\)
The models for the econometric analysis

1\textsuperscript{st} model: Identification of the \textit{attributional} determinants of the probability to publish (i.e, to be an element of Z) [dprobit]

2\textsuperscript{nd} model: (only for Z members) identification of the \textit{relational} driving forces of scientific productivity [IV]

3\textsuperscript{rd} model: The impact of \textit{positional} variables (Co-authorship MC) on scientific productivity [Heckman procedure]
1st model

**Dependent variable:**
- Z is equal to 1 if the economist has published at least a JA in Econlit during 1969-2006, and 0 otherwise.

**Regressors:**
- set of *attributional* variables available for all the economists in the population (i.e. gender, tenure, scientific sub-sector, faculty of Economics, geographic location and lecturer academic position without tenure) that could affect this probability of publishing.
# 1st model: The results

Estimation method: dprobit  
Dep. var.: \( z = 1; 0 \)

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th></th>
<th>[2]</th>
<th></th>
<th>[3]</th>
<th></th>
<th>[4]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( dF/dx )</td>
<td>( t )-value</td>
<td>( dF/dx )</td>
<td>( t )-value</td>
<td>( dF/dx )</td>
<td>( t )-value</td>
<td>( dF/dx )</td>
<td>( t )-value</td>
</tr>
<tr>
<td>Gender</td>
<td>0.086**</td>
<td>3.80</td>
<td>0.086**</td>
<td>3.81</td>
<td>0.086**</td>
<td>3.80</td>
<td>0.075**</td>
<td>3.41</td>
</tr>
<tr>
<td>Tenured</td>
<td>-0.041**</td>
<td>-2.01</td>
<td>-0.038*</td>
<td>-1.88</td>
<td>-0.039*</td>
<td>-1.94</td>
<td>-0.163**</td>
<td>-6.10</td>
</tr>
<tr>
<td>Economics</td>
<td>...</td>
<td>...</td>
<td>0.064**</td>
<td>2.22</td>
<td>0.066**</td>
<td>2.27</td>
<td>0.080**</td>
<td>2.78</td>
</tr>
<tr>
<td>Econometrics</td>
<td>...</td>
<td>...</td>
<td>0.114**</td>
<td>2.36</td>
<td>0.112**</td>
<td>2.30</td>
<td>0.092*</td>
<td>1.81</td>
</tr>
<tr>
<td>Public Econ.</td>
<td>...</td>
<td>...</td>
<td>0.006</td>
<td>0.18</td>
<td>0.008</td>
<td>0.24</td>
<td>0.027</td>
<td>0.80</td>
</tr>
<tr>
<td>Others</td>
<td>...</td>
<td>...</td>
<td>Ref.</td>
<td></td>
<td>Ref.</td>
<td></td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Fac_Economics</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.038**</td>
<td>1.96</td>
<td>0.036*</td>
<td>1.90</td>
</tr>
<tr>
<td>North West</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.042*</td>
<td>1.76</td>
</tr>
<tr>
<td>North East</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.055**</td>
<td>2.13</td>
</tr>
<tr>
<td>Centre</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>-0.039</td>
<td>-1.33</td>
</tr>
<tr>
<td>Islands</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>-0.170**</td>
<td>-4.05</td>
</tr>
<tr>
<td>Lecturer</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>-0.345**</td>
<td>-7.14</td>
</tr>
<tr>
<td>N. Obs.</td>
<td>1.620</td>
<td></td>
<td>1.620</td>
<td></td>
<td>1.620</td>
<td></td>
<td>1.620</td>
<td></td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.011</td>
<td></td>
<td>0.018</td>
<td></td>
<td>0.020</td>
<td></td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Obs. P</td>
<td>0.812</td>
<td></td>
<td>0.812</td>
<td></td>
<td>0.812</td>
<td></td>
<td>0.812</td>
<td></td>
</tr>
<tr>
<td>Pred. P</td>
<td>0.815</td>
<td></td>
<td>0.817</td>
<td></td>
<td>0.817</td>
<td></td>
<td>0.834</td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors are robust to heteroskedasticity.

Legend: ** significant at 5%; ** significant at 10%
2nd model (1/2)

We explain the individual productivity (i.e. the number of JA weighted according to the “scientific age” of each individual) of Italian economists in terms of relational driving forces (i.e. propensity to co-authorship and to have international connections) and we control for attributional variables (i.e. gender, tenure, geography and age classes).
Dependent variable:
- **Scientific productivity**: log of number of JA publications weighted by the “scientific age” of each economist (estimated as the difference between the year of the first Econlit publication and 2006). We are missing the “biological age”.

Regressors:
- **Propensity to co-authorship**: for each author is calculated the proportion of collaboration on his/her papers. It ranges between 0 and 1: 0 = no collaboration, 1 = all papers are written in collaboration;
- **Foreign**: proportion of foreign co-authors;
- **Mills ratio**: to control for the sample selection bias
- **Dummy variables controlling for**: gender, tenure, age class
2nd model:
Some empirical issues (1/2)

- There could be a **sample selection bias**, since we are analysing exclusively those economists belonging to $Z$.
  
  We solve this problem calculating the inverse Mills ratio.

- **Endogeneity problem**: i.e. cooperation affects productivity, but productivity may affect cooperation (i.e. I may choose a co-author because he/she is very productive) generating a reverse causality problem.
  
  We solve this problem adopting IV strategy, instrumenting the propensity to cooperate using the number of collective volume articles (**CVA**) written by each author.
CVA as IV (2/2)

- We use collective volume articles (CVA) written by each author as instrument of propensity to co-authorship.

- In the literature, it is assumed that CVA are the effect of connections and are not quality comparable with JA and may reflect alternative use of a scientist’s time (i.e. if a scientist writes contribute for a CVA he/she has no time to write a JA). Thus CVA may measure the propensity to socialise irrespective of the impact on scientific productivity.

Tests confirm that both instruments are OK!
2nd model: The results
Estimation method: IV 2SLS
Dep. var.: log of JA by scientific age

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th>[2]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-value</td>
</tr>
<tr>
<td>Propensity of coauthorship</td>
<td>5.686 **</td>
<td>5.21</td>
</tr>
<tr>
<td>Foreign</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Gender</td>
<td>0.404 **</td>
<td>2.64</td>
</tr>
<tr>
<td>Tenured</td>
<td>0.196</td>
<td>1.47</td>
</tr>
<tr>
<td>North West</td>
<td>-0.227</td>
<td>-1.40</td>
</tr>
<tr>
<td>North East</td>
<td>-0.373**</td>
<td>-2.05</td>
</tr>
<tr>
<td>Centre</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>South</td>
<td>-0.092</td>
<td>-0.49</td>
</tr>
<tr>
<td>Islands</td>
<td>-0.156</td>
<td>-0.45</td>
</tr>
<tr>
<td>Age1_10</td>
<td>-0.326</td>
<td>-1.35</td>
</tr>
<tr>
<td>Age11_20</td>
<td>-0.310*</td>
<td>-1.55</td>
</tr>
<tr>
<td>Age31_40</td>
<td>0.475**</td>
<td>2.20</td>
</tr>
<tr>
<td>Mills Ratio</td>
<td>-0.581</td>
<td>-1.07</td>
</tr>
<tr>
<td>N. Obs.</td>
<td>1,317</td>
<td></td>
</tr>
<tr>
<td>Hausman (p-value)</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Note: regressions also include a constant term. Standard errors are bootstrapped (50 replications) in order to account for the generated regressor problem.

Legend: ** significant at 5%; * significant at 10%
3rd model

**Dependent variable:**
- **Scientific productivity**: defined as before

**Regressors:**
- we focus our empirical analysis on **positional** variables characterising the main component (MC)
  - clustering coefficient,
  - closeness
  - betweenness
  - instability of scientific cooperation (i.e. a value that ranges between 0 – i.e. all co-authors are the same – and 1 – i.e. all co-authors are different).

Here endogeneity is less of a problems because these are “unobservable” networks features.
Some elements of SNA (1/2)

Node d is the node with the highest degree value (direct connections)

Node h is the node with the highest betweenness value (bridging role)

Nodes g and f are the nodes with the highest closeness value (indipendency role)
Some elements of SNA (2/2)

Clustering coefficient (CC): is a measure of “neighborhood”, it could be synthesized as follows: The CC is the degree to which nodes in a graph tend to cluster together. It measures the cliquishness of networks of node $i$’s neighbors.

$$CC_i = \frac{\Lambda_i}{v_i}$$

where $\Lambda_i$ indicates the number of edges in the neighborhood of node $i$ and $v_i$ the total number of possible edges of node $i$. The index varies between 0 (no neighbor of any vertex is adjacent with any other neighbor of vertex $i$) and 1 (individually complete neighbors).
3rd model: The results
Estimation method: Heckman two-stages procedure
Dep. var.: log of JA by scientific age

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>-0.174**</td>
<td>-2.64</td>
<td>-0.224***</td>
<td>-3.73</td>
<td>-0.083</td>
<td>-1.30</td>
<td>-0.112*</td>
<td>-1.90</td>
</tr>
<tr>
<td>Foreign</td>
<td>0.460***</td>
<td>3.48</td>
<td>0.445***</td>
<td>3.85</td>
<td>0.362***</td>
<td>4.00</td>
<td>0.352**</td>
<td>3.27</td>
</tr>
<tr>
<td>Instability</td>
<td>-1.182***</td>
<td>-13.66</td>
<td>-1.190***</td>
<td>-11.34</td>
<td>-1.147***</td>
<td>-12.57</td>
<td>-1.192***</td>
<td>-13.33</td>
</tr>
<tr>
<td>Closeness</td>
<td></td>
<td></td>
<td>2.637***</td>
<td>5.79</td>
<td></td>
<td></td>
<td>1.103**</td>
<td>2.63</td>
</tr>
<tr>
<td>Betweenness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.878***</td>
<td>13.10</td>
<td>3.393***</td>
<td>8.39</td>
</tr>
<tr>
<td>Gender</td>
<td>0.314***</td>
<td>5.81</td>
<td>0.256***</td>
<td>4.34</td>
<td>0.256***</td>
<td>5.68</td>
<td>0.244***</td>
<td>4.65</td>
</tr>
<tr>
<td>Tenured</td>
<td>-0.050</td>
<td>-0.79</td>
<td>-0.047</td>
<td>-0.93</td>
<td>-0.075</td>
<td>-1.43</td>
<td>-0.084*</td>
<td>-1.80</td>
</tr>
<tr>
<td>Geogr. dummies</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Age dummies</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>N. Obs.</td>
<td>1,249</td>
<td></td>
<td>1,249</td>
<td></td>
<td>1,249</td>
<td></td>
<td>1,249</td>
<td></td>
</tr>
<tr>
<td>Censored Obs.</td>
<td>588</td>
<td></td>
<td>588</td>
<td></td>
<td>588</td>
<td></td>
<td>588</td>
<td></td>
</tr>
<tr>
<td>Uncens. Obs.</td>
<td>661</td>
<td></td>
<td>661</td>
<td></td>
<td>661</td>
<td></td>
<td>661</td>
<td></td>
</tr>
</tbody>
</table>

Note: Regression include a constant term. Standard errors are bootstrapped (50 replications) in order to account for the generated regressor problem.

Legend: *** significant at 1%; ** significant at 5%; * significant al 10%
Conclusions

1) The probability of publishing is influenced by gender, disciplinary groups, geographical area, faculty.

2) If one economist does publish, his/her productivity depends positively with his/her propensity to collaborate and on his/her international connections (intrinsic quality vs. editorial boards!?).

3) Position in the networks affects productivity. Being “central” is a plus and being a “bridge” is better than being globally central. Cliquishness is bad for science! (hint: star structure and exploitation of co-author gives good results?) Stability (and fidelity) pays! Keeping the same authors, at least for an Italian economist, is the best strategy in order to improve his/her productivity.
Further research

a. Structural stability test
b. Longitudinal analysis
c. Geography (e.g. distance between co-authors)