A social network analysis of research collaboration in the economics community

THOMAS KRICHEL[1]
NISA BAKKALBASI[2]

1 Long Island University College of Information and Computer Science
720 Northern Boulevard, Brookville, New York 11548 USA
2 Yale University Kline Science Library
PO Box 208111, New Haven, Connecticut 06520 USA

Abstract

RePEc (Research Papers in Economics) offers the RePEc Author Service (RAS). It allows registrants to claim authorship of the research papers that are described in RePEc archives. The data from this service forms a high-quality authorship database. We use this data to examine, as a practical example, how different network constructions affect the ranking of economists through authorship centrality. We use Spearman's rho test for evaluating the correlation between author centrality measures.

1. Introduction and motivation

There has been a vivid interest in scientific collaboration networks in recent years. Since Beaver [1] presented the first comprehensive theory of scientific collaboration, formally acknowledged by co-authorship of scientific papers, a growing number of scientists have been focusing in collaboration networks. In particular, co-authorship networks have been widely used to examine the patterns of collaborations within an academic community and determine the status and influence of individual researchers [2-7]. Social network analysis (SNA) is the principle tool used to examine patterns of collaboration in different scientific fields. Although there may not be a consensus on intellectual foundations of modern SNA, Freeman [8] traces back the history of the study of the patterned social interactions to 1800s. In the last decade or so, modern SNA produced many results concerning social influence, communication flows, and information sharing. In this paper we investigate the structure of research collaborations in economics, by applying social network analysis to the co-authorship network formed by RePEc Author Service (RAS).

RePEc, one of the earliest digital libraries in existence, has been conceived and developed to promote scholarly communication and enhance the dissemination of research findings in the field of economics [9]. According to its web site at http://repec.org, RePEc is a collaborative effort of hundreds of volunteers in 51 countries. As of this writing, RePEc describes over 362,000 items of interest such as working papers, journal articles, software components, and instructional datasets. All RePEc data is freely available online. University departments, institutions involved in economics research (e.g. central banks), publishers, and individuals contribute the contents and its associated metadata. RePEc is an interesting model of alliance and partnership among institutions, publishers, and researchers, each with a different stake in the scholarly communication process. Barrueco Cruz and Krichel [10], early pioneers of RePEc, provide a detailed discussion of their approach towards building the digital library and summarize the basic principles. Their vision for RePEc goes well beyond merely providing descriptions of and access to research papers. In fact, RePEc aims to create a relational database between the following four types of items:

Documents <=> Collections of documents (e.g., working papers series, journals, etc.)

Persons <=> Collection of persons (e.g., institutions involved in economics research)

The latter is important, as it serves as the basis for this paper. A single volunteer, Christian Zimmermann, maintains the registry of institutions at http://edirc.repec.org. However it would be impractical for one volunteer to register hundreds of authors on his own, in addition to maintaining the links from the author records to the documents they have written. As a result, this task has to be carried out by the authors themselves, using the aforementioned author service: RAS. Barrueco Cruz et al. [11] provide an early account of the service. Since 2003, RAS is based on the Academic Contribution Information System (ACIS), see http://acis.openlib.org, of which the Open Society Institute has sponsored the development.

When authors contact an ACIS-based service such as RAS for the first time, they provide basic information such as name, homepage, affiliation etc. Subsequently, they create a profile of spelling variations for their name. Periodically, ACIS scans the documents' author name fields to find name variations in an author's name variations profile. When a new document that matches a name variation is found, the RAS registrant is alerted via email. This email invites the author to claim the item. As the RePEc bibliographic database becomes more complete, an accurate academic contributions profile for an author is built. For illustration, here is an extract of the record for Christian Zimmermann:

```
Template-Type: ReDIF-Person 1.0
Name-First: Christian
Name-Last: Zimmermann
Name-Full: Christian Zimmermann
Workplace-Organization: RePEc:edi:deuctus
Email: christian.zimmermann@uconn.edu
Homepage: http://ideas.repec.org/zimm/
Author-Paper: repec:cre:crefwp:33
Author-Paper: repec:mtl:montde:2000-05
Author-Software: repec:dge:qmrbcd:99
Author-Software: repec:dge:qmrbcd:97
Author-Paper: repec:uct:uconnp:2005-01
Author-Article: repec:eee:jcecon:v:33:y:2005:i:1:p:88-106
Author-Article: repec:eee:jmacro:v:26:y:2004:i:4:p:637-659
Author-Paper: repec:sce:scecf5:372
Author-Paper: repec:red:sed005:561
Short-Id: pzi1
Handle: repec:per:1964-12-14:christian_zimmermann
Last-Login-Date: 2005-11-21 15:25:20 -0500
Registered-Date: 2004-02-29 17:36:09 -0600
```

Since not all authors are registered, the database is not yet complete. Bakkalbasi and Krichel [12] assess the completeness of RAS in detail. They report roughly one in three papers in RePEc is included in the RAS database and the ratio of the number of authorships in RAS to the total number of authorships in RePEc is about one in four. Despite the limited coverage, we believe that it makes more sense to analyze this dataset than use datasets that are not properly cleaned.

The main purpose of author registration is to build a dataset that can be used for the evaluation of academic research. One particular aspect of the evaluation, scientific collaboration, has been the motive for this paper. By using three most popular individual network measures, *degree centrality*, *closeness centrality*, and *betweenness centrality*, we want to determine the importance or prominence of an author in the discipline as a whole. We use three different approaches to build the co-authorship network: a traditional binary network and two weighted networks. In more detail, we make an inquiry into the rankings produced by different approaches to find out how closely they are related to each other. In next section we look at the overall RAS data and give overall summary statistics. In section 3, we discuss three different types of networks. In section 4, we analyze how close the rankings produced by different centrality measures. In section 5, we conclude.

2. Constructing collaboration networks

We extract data from the RAS to construct the collaboration network. A network can be presented as a graph, which consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations [13]. To build the co-authorship network model for this study we use authors as nodes and co-authorship of papers as edges.

In the RAS data, we find a total number of 13,049 registrants, of which 9,111 (i.e. 70%) have claimed at least one paper. It is difficult to explain why some individuals have registered for the service but not claimed any papers. These registrants are excluded from further analysis. Among the remaining 9,111 authors, there are 6,038 (i.e. 66%) co-authors, that is authors who have collaborated with another RAS registered author. All other authors, who do not meet this criterion, are eliminated from further analysis.[1]

A component is a connected subset of a graph in which there are paths between the nodes [13]. If two authors have written a joint paper, there is a path between them. If a third author has co-authored with any one of the first two authors, a chain of co-authorship path (henceforth "path") can be built connecting the first author with the third author and so on. If a path between two authors can be established, the two authors are said to belong to the same component of the network. Typically, the largest component of an observed network comprises more than 50% of all nodes. It is known as the giant component. The giant component of the RAS authorship network has 5,019 (i.e. 83%) authors. In the following, we only consider the authors that are in the giant component. Thus, our network has 5,019 nodes. We explore author centrality within this network.

One way to measure the centrality of an author is to look at how many immediate ties an author has. This is called the degree centrality of an author. Appendix A contains a table of the top twenty authors with the highest numbers of degree centrality scores, labeled as DEG. Note that since co-authorship is a symmetric relationship, we count co-authorship twice both as an incoming and outgoing edge. Thus, the numbers of edges reported in the table are even numbers and the degree centrality score divided by two gives the number of collaborators per author. Many authors have common degree values. The table in appendix A also furnishes the top twenty authors with the highest number of documents. This number yields another ranking of authors to reflect their prolificacy, labeled as NDO.

The previous measure, *degree centrality*, takes into account the immediate ties an author has rather than the ties to all others within a network. In other words, even though an author may be tied to a large number of others, if those others are disconnected from the larger network, then the author is only central within a "local" network. There are two other commonly used criteria that assess the global centrality of a node.

The first, *closeness centrality*, reflects the average length of the shortest paths leading from one node to all other nodes. The length of the shortest path is the sum of the edges on the path. The precise absolute value therefore depends on how the length of the edges is measured. But what counts for ranking is the relative value of the averages. Authors who have smaller average path lengths are considered to be the more central. Usually, in a large network, the closeness centrality scores of every

---

[1] Note that many of those authors have collaborated with others, but those others have not yet registered with RAS. Although we have not carried out a formal investigation, taking a glance at the RAS and the RePEc document data suggests that the fact an author registers does not increase the likelihood of his co-authors to register. Registration appears to be an individual decision made independently of co-authors.

author have a unique value. In terms of rankings over all authors, closeness centrality provides a more granular measure than the degree centrality, which, as we noted, yields many tied ranks.

The second, *betweenness centrality*, reflects the number of times a node appears on the shortest path between any two other nodes. Since there may be multiple shortest paths between two authors *A* and *B* we proceed as follows. Let $p(A,B)$ be the number of paths between two authors *A* and *B*. Let $n(A,B,C)$ of these paths pass through author *C*. Then the betweenness score of author *C* between authors *A* and *B* is:

$$\frac{n(A,B,C)}{p(A,B)}$$

The betweenness centrality score of an author is the sum of his betweenness counts between any two other authors[2]. All authors who only have one co-author, or, in other words who have a degree of two, are tied with zero scores at the bottom of betweenness ranking. In our study, we refer to these as "marginal" authors. There are 1,355 marginal authors in the data.

When we consider closeness and betweenness as centrality measures, the way we construct the edge lengths has an impact on the centrality scores of an author. The impact may be even more significant than the difference between closeness centrality scores and betweenness centrality scores. To assess this issue empirically, we construct three networks by using different edge lengths. These are the binary network, the weighted symmetric network and the random walk network. We discuss each in turn in Section 3.

3. *Collaboration network models*

We use three different approaches to model the co-authorship networks by calculating the strength of collaboration using three different methods.

3.1. *Binary network model*

By far the most common approach to assigning values to collaboration strength is to simply distinguish between relations being absent and being present. If two individuals have been co-authors on one or more papers, they receive a collaboration weight of one. Otherwise the collaboration weight is zero.

To calculate the shortest paths between authors in such a network we implement an algorithm proposed in Newman [14]. Our homegrown Perl script, available on request, calculates all shortest paths between each pair of nodes. Multiplicity of paths turns out to be numerically important. In the absence of multiplicity, each author has 5,018 shortest paths to others. In our dataset, the author with the smallest number of multiple paths has 12,116 paths while the author with the largest number of paths has a staggering number of 54,466 paths. But such a high number is rare. The median is 15,546. Only 24 authors have more than 30,000 paths.

Appendix B furnishes a list of top twenty authors with the highest numbers of betweenness centrality ranking, labeled as BIB, and closeness centrality ranking, labeled as BIC. Both measures are based on the binary network model. We note that the differences in successive betweenness values are

---

[2] To be consistent with the degree counts, we consider the paths between authors twice, once as a path from author *A* to author *B*, and again as a path from *B* to *A*. In a symmetric network the edge between *A* and *B*, if it exists, is of the same length as the edge between *B* and *A*. In such a network, if there is a path between *A* and *C*, all shortest paths between authors *A* and *C* are the same as the shortest paths between *B* and *A.* In an asymmetric network the edge between *A* and *B*, if it exists, is not necessarily of the same length as the edge between *B* and *A*. The latter may not even exist. In such an asymmetric network, if there is a path between *A* and *C*, shortest paths between *A* and *B* are likely to be different than shortest paths between *B* and *A*.

quite large to start with, but seem to be become progressively smaller. For closeness, the absolute values of differences between successive values are much smaller. Both lists seem to be well correlated at the top. Bakkalbasi and Krichel [12] discuss the differences between the two lists.

3.2. Weighted network models

In the binary network model, by dichotomizing the data, we lose valuable information relating to the strength of the collaborations. Common sense dictates that if two authors collaborate frequently, the tie between them is stronger than the tie between two authors who collaborate once or occasionally. In order to take the process further, we explore two weighted network models using two different edge weighting schemes.

3.2.1 Symmetric weight strength

First, we use a weighting scheme we find in Newman [14]. Suppose an author collaborates on a paper k that has n_k authors in total. Let δ_i^k be 1 if scientist i was a co-author of paper k and zero otherwise. Then the symmetric weight w_{ij} representing the strength of the collaboration (if any) between scientists i and j is:

$$w_{ij} = \sum_k \frac{\delta_i^k \delta_j^k}{n_k - 1},$$

Since for $k = 1$, this formula is not defined, we exclude from our sums all single authored papers.

For example, if author *A* has written one co-authored paper with author *B*, but on that paper author *C* also appears a co-author, then the co-authorship strength between *A* and *B* is only ½. If, in addition to this paper, the author *A* also has written another paper with author *C*, where only *A* and *C* are authors, then his collaboration strength with *C* is ½ from the first paper and 1 from the second paper. Thus the total collaboration strength of author *A* is ½ from the collaboration with author *B* and 1½ from the collaboration with author *C*. Thus, total co-authorship strength of *A* is the sum of all co-authored papers. This is in fact a general feature of this measure.

$$\sum_{j(\neq i)} w_{ij} = \sum_k \sum_{j(\neq i)} \frac{\delta_i^k \delta_j^k}{n_k - 1} = \sum_k \delta_i^k$$

To calculate the shortest paths, we adopt a Perl script found at http://www.sabren.net/code/perl/dijkstra. A single run takes around 10 hours under the same conditions as in the binary case. The algorithm is slower to find paths than the algorithm in the binary case. But the Dijkstra algorithm does not deal with multiplicity of paths. It finds only one shortest path.

Appendix C furnishes a list of top twenty authors with the highest numbers of betweenness centrality scores, labeled as SYB, and closeness centrality scores, labeled as SYC, based on a weighted network model. The rankings are very different. The figures for closeness appear to be much smaller than in the binary case. In the binary case, the sum of weights of an author is the sum of the collaborators. In the symmetric weighted case, the sum of weights is the number of papers that have been collaborated on. It appears that the latter sum is much greater than the former. This comes as no surprise given the tendency of prolific authors to register with RAS as reported by Bakkalbasi and Krichel [12]. Since there is only one shortest path found, the betweenness numbers are all integers. However, although most of them are even numbers, not all of them are. The algorithm may find a different path from *A* to *B* than from *B* to *A*.

3.2.2 Random walk strength

Here we adopt an approach that we find in Liu et al. [5]. The idea is that if an author has a lot of collaborators, the links with each individual collaborator should be reduced. One approach is to normalize the weights for the symmetric relationship. The new weights are

$$v_{ij} = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}}$$

This normalization ensures that the weights of an author's relationships sum to one. This weighting scheme has an intuitive interpretation in terms of a random walk on the collaboration graph. For this reason we call this weighting scheme the random walk scheme. For every paper written by author *A*, the probability to reach author *B* as the next node in the random walk is the number of papers that *A* wrote with *B*, divided by the number of papers that *A* has co-authored. Promiscuous authors who have written a lot of papers with a lot of co-authors will have weak ties to each of these authors. This implies that there are long outward edges from them. However, the edge from a minor author, who has only written a few papers but, with one major author, will have a strong tie to the major author, therefore have a short edge to this major author. Thus the network is asymmetric, but in a specific way. Prolific and promiscuous authors have strong incoming links (short edges) and weak outgoing links (long edges). This is a typical way of giving prolific and promiscuous authors higher prestige; see Wasserman and Faust [13]. The low penalty of getting to the high-prestige author node is balanced by the high penalty of leaving the author node. Thus we expect some authors to be central who have relatively fewer joint papers, but have written them with individuals who are in different subject areas. Such authors are serving as bridges among subject cliques that are otherwise far apart.

The calculations are done with the same Perl script that we use for the symmetrically weighted network. Similarly, calculations take around 10 hours and do not calculate multiple paths. But this should really not matter for the results because intuitively the amount of path multiplicity should be quite small in this network, even smaller than in the symmetrically weighted network. Appendix D furnishes a list of top twenty authors with the highest numbers of betweenness centrality scores, labeled as RWB, and closeness centrality scores, labeled as RWC, based on a random walk network model. Note that the closeness values are much higher. Normalization of weights has reduced each weight and has made the edges, which are the inverses of weights, longer.

4. Rank correlation of centrality metrics

The influence and prestige of the two economists, Joseph Stiglitz and Clive Granger, in RAS network is unquestionable. They are the recipients of the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel. However, based on which criteria do we rank the rest of the RAS registrants? Although we may not be able to verify which method performs best, we can compare the correlations between the metrics. We use Spearman's correlation coefficient to measure the strength of the association between centrality metrics obtained by using different edge weighting schemes. Table 1 furnishes a list of the Spearman's correlations for all possible pairs. A number of surprising facts emerge.

First, it seems the association between any pair of betweenness scores (e.g. BIB vs. SYB) is stronger than the association between any pair of betweenness and closeness scores (e.g. BIB vs. SYC). This observation is similar for the closeness scores with the exception of random walk closeness scores, which appears to be poorly correlated to all other scores.

Next, we take a look at the association between the numbers of papers an author has published (NDO) and the various centrality metrics. The moderately high correlation coefficient (i.e. 70%) between the NDO and the degree centrality rank (DGE) indicates productive authors have more collaborators than less productive authors. The association between the NDO rank and each of the betweenness centrality rank are moderately high whereas the association between the NDO rank and each of the closeness centrality ranks are surprisingly low. We observe that the lowest correlation coefficient (i.e. 19%) among all pairs is between NDO and random walk closeness rank (RWC).

In general, betweenness scores are strongly correlated with one another, whereas closeness scores appear to be moderately correlated with one another.

Table 1. Rank correlation of centrality metrics

	NDO	**DGE**	**BIB**	**BIC**	**SYB**	**SYC**	**RWB**	**RWC**
NDO	1	0.7	0.68	0.55	0.71	0.6	0.7	0.19
DGE	0.7	1	0.84	0.67	0.85	0.57	0.87	0.3
BIB	0.68	0.84	1	0.6	0.9	0.52	0.89	0.3
BIC	0.55	0.67	0.6	1	0.54	0.81	0.61	0.57
SYB	0.71	0.85	0.9	0.54	1	0.54	0.91	0.23
SYC	0.6	0.57	0.52	0.81	0.54	1	0.56	0.42
RWB	0.7	0.87	0.89	0.61	0.91	0.56	1	0.41
RWC	0.19	0.3	0.3	0.57	0.23	0.42	0.41	1

5. *Conclusions*

Co-authorship centrality rankings seem to be a promising way to generate incentives for registered authors to promote RAS to their unregistered co-authors. If each author understands that his position in the network is positively affected by convincing his co-authors to join the service and claim documents, RAS data will become complete. If centrality rankings are an indicator of prominence within the discipline, authors will have some incentive to request their co-authors to register with the service. Although we do not have a formal investigation, registration seems to be an isolated act that is not influenced by other authors. For example, when we take a look at Andrei Shleifer's record, we find that he has written more than 30 papers with Robert Vishny. However, Robert Vishny is not a registered author. If we present Shleifer with his centrality ranking, and he understands that his ranking will improve once his collaborator Vishny registers, he most likely encourages him to join the service.

What ranking should be present to authors? From a managerial point of view, we don't want to present all rankings as this would confuse the authors and send a message of relativism that we do not want to encourage. So we have to be selective. Simplicity of the proposed scheme appears to be one criterion. A simpler message gets through easier. But simplicity of the scheme should not be the only criterion. It should also be positively co-related with other quality criteria. No author will be motivated to rise up in a list that has a group of people at the top that do not appear to be accomplished. One way to ensure the inclusion of reputable authors is to also look at the total number of papers authors have written. The total number of papers is, within the measures that we have proposed here, probably the most immediate indicator of academic reputation. Thus if we want to pick one single network, and present both closeness and betweenness scores, it is the symmetric weighted network that we should pick. It dominates the two other models by showing a higher correlation to the NDO ranking. The only problem that we can see with this measure is that it is not the easiest to explain. Then again, we are dealing with economists here. They are used to sophisticated decision making processes and usually have good mathematic reasoning ability.

References

1. D. de B. Beaver and R. Rosen, Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship, Scientometrics, 1: 65-84, 1978.
2. M.E.J. Newman, Who is the best connected scientist? A study of scientific co authorship networks. Lecture Notes in Physics. 650: 337-3370, 2004.
3. M.E.J. Newman, Co-authorship networks and patterns of scientific collaboration. Proceedings Of The National Academy Of Sciences Of The United States Of America. 101 (Suppl. 1): 5200 – 5205, 2004.
4. A.L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, Evolution of the social network of scientific collaborations. arXiv:cond-mat/0104162v1, 2005.
5. X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, Co-authorship networks in the digital library research community, Information Processing & Management, 41 (6): 1462-1480, 2005.
6. S. Goyal, M. J. van der Leij. and J-L. Moraga, Economics: An emerging small world? CESifo Working Paper Series No. 1287; FEEM Working Paper No. 84.04; Tinbergen Institute Discussion Paper No. 04-001/1, 2004.
7. E. Otte, and R. Rousseau, Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science, Vol. 28, No. 6: 441-453, 2002.
8. L. C. Freeman, The Development of social network analysis: A study in the sociology of science, Empirical Press, Vancouver, BC, 2004.
9. E. Walshe, Creating an academic self-documentation system through digital library interoperability: the RePEc model. The New Review of Information Networking, 7: 43-58, 2001.
10. J. M. Barrueco Cruz and T. Krichel, Cataloging economics preprints: An introduction to the RePEc project. Journal of Internet Cataloging, 3(3): 227 -241, 2001.
11. J. M. Barrueco Cruz, M. J.R. Klink, and T. Krichel, Personal data in a large digital library. In Proceedings 4th European Conference on Research and Advanced Technology for Digital Libraries, Lisboa (Portugal), 2000.
12. N. Bakkalbasi and T. Krichel, Patterns of research collaboration in a large digital library, pre-print available at http://openlib.org/home/krichel/papers/elba.pdf, 2006.
13. S. Wasserman and K. Faust, Social network analysis: methods and applications. Cambridge University Press, 1994.
14. M. E. J Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Physical Review E. Vol. 64, 016132, 2001.

Appendix A. Top twenty authors with the highest numbers of degree centrality scores and numbers of papers.

Rank	Degree Centrality Scores (DEG)		Number of papers (NDO)	
1	Randall Wright	54	Barry Eichengreen	324
2	Clive Granger	52	Peter Phillips	322
3	Joseph Stiglitz	52	Joseph Stiglitz	318
4	Pierre Chiappori	46	M Pesaran	278
5	Philip Franses	44	Martin Shubik	278
6	Stephen Jenkins	44	Lars Svensson	270
7	Ronald MacDonald	44	Martin Feldstein	252
8	Gert Wagner	44	Jeffrey Frankel	234
9	Francis Diebold	42	Jean-Jacques Laffont	230
10	Costas Meghir	42	Stephen Turnovsky	226
11	Peter Phillips	42	Jean Tirole	225
12	Fabio Schiantarelli	42	Sebastian Edwards	223
13	Barry Eichengreen	40	James Heckman	219
14	Andrew Rose	40	James Poterba	217
15	Thomas Sargent	40	Andrei Shleifer	210
16	Friedrich Schneider	40	Bruno Frey	205
17	Olivier Blanchard	38	Maurice Obstfeld	202
18	Carlo Favero	38	Alan Krueger	200
19	Eric Ghysels	38	Daron Acemoglu	199
20	Francesco Giavazzi	38	Richard Freeman	199

Appendix B. Top twenty authors with the highest numbers of betweenness centrality scores and closeness centrality scores based on a binary network model.

Rank	Closeness Centrality Scores (BIC)		Betweenness Centrality Scores (BIB)	
1	Joseph Stiglitz	4.82742	Joseph Stiglitz	944006
2	Olivier Blanchard	4.90574	Fabio Schiantarelli	857455
3	Fabio Schiantarelli	4.96971	Juergen von Hagen	814814
4	Alison Booth	5.01455	Gert Wagner	628828
5	Juergen von Hagen	5.04344	Juan Dolado	607793
6	Costas Meghir	5.05899	Costas Meghir	596778
7	James Stock	5.06417	Klaus Zimmermann	578973
8	Barry Eichengreen	5.06736	Clive Granger	572371
9	Marcus Miller	5.07912	Friedrich Schneider	551241
10	Andrew Rose	5.09924	Mark Taylor	531349
11	William Brock	5.10283	Olivier Blanchard	529492
12	Michele Boldrin	5.11419	Alison Booth	516812
13	Michael Rothschild	5.12196	Pierre Chiappori	499438
14	Randall Wright	5.12216	Thierry Verdier	457124
15	Paul Beaudry	5.13113	John McMillan	431094
16	Juan Dolado	5.14169	Randall Wright	428264
17	Mark Taylor	5.14946	Harald Uhlig	421033
18	Pierre Chiappori	5.16182	Paul Beaudry	411765
19	Costas Azariadis	5.16481	Michele Boldrin	410979
20	Paul Masson	5.17079	Ronald MacDonald	408347

Appendix C. Top twenty authors with the highest numbers of betweenness centrality scores and closeness centrality scores based on symmetrical weight strength.

Rank	**Closeness centrality scores** (SYC)		**Betweenness centrality scores** (SYB)	
1	Lars Svensson	1.88131	Lars Svensson	3185434
2	Torsten Persson	1.88915	Daron Acemoglu	2518136
3	Guido Tabellini	1.89188	Thierry Verdier	2403658
4	Glenn Rudebusch	1.90883	Andrew Rose	2333892
5	Gérard Roland	1.91279	Francis Diebold	2202567
6	Francis Diebold	1.91501	Torsten Persson	2133208
7	Andrew Rose	1.91618	Alan Krueger	2047358
8	Thierry Verdier	1.91768	Andrei Shleifer	1937633
9	Daron Acemoglu	1.92965	Gérard Roland	1848082
10	Jeffrey Frankel	1.93197	Jeffrey Frankel	1794714
11	Andrei Shleifer	1.93646	George Mailath	1706441
12	Francesco Giavazzi	1.9427	Ernst Fehr	1687100
13	Michael Woodford	1.94383	Glenn Rudebusch	1586136
14	Tim Bollerslev	1.94549	Klaus Schmidt	1503542
15	Jorn-Steffen Pischke	1.95677	Jorn-Steffen Pischke	1455786
16	Robert Flood	1.95993	Guido Tabellini	1388466
17	Fabrizio Zilibotti	1.96115	Mark Taylor	1384477
18	Edward Glaeser	1.96118	Francesco Giavazzi	1340228
19	Simon Johnson	1.96194	Drew Fudenberg	1290778
20	Peter Christoffersen	1.96361	Joseph Stiglitz	1241014

Appendix D. Top twenty authors with the highest numbers of betweenness centrality scores and closeness centrality scores based on random walk strength.

Rank	Closeness centrality scores (RWC)		Betweenness centrality scores (RWB)	
1	Aloysius Siow	50.53682	Fabio Schiantarelli	622886
2	Argia Sbordone	50.95877	Andrei Shleifer	467193
3	Jose Tavares	50.99672	Daron Acemoglu	452742
4	Robert Gordon	51.17032	Joseph Stiglitz	449943
5	Maristella Botticini	51.22961	Cars Hommes	449620
6	John Matsusaka	51.31263	Pierre Siklos	430581
7	Guy Debelle	51.37366	Dilip Mookherjee	419434
8	Marc Rysman	51.49838	Alan Krueger	390444
9	Daniel Ackerberg	51.50198	Ronald MacDonald	388907
10	Richard Posner	51.53658	Ernst Fehr	383063
11	Xiaodong Zhu	51.6227	Geert Ridder	372212
12	Merton Miller	51.62681	Harald Uhlig	366237
13	Luis Cabral	51.68711	Gianni De Fraja	358873
14	Abigail Payne	51.79716	Timothy Bresnahan	356546
15	Richard Barnett	51.84694	Thierry Verdier	356166
16	Philippe Weil	52.08261	Debraj Ray	342839
17	Nicholas Economides	52.11043	David Card	341819
18	Eric Fisher	52.17904	Paul Masson	341202
19	Emin Dinlersoz	52.19531	Edward Glaeser	340802
20	Giuseppe Lopomo	52.21281	Jan van Ours	337870