Economics: an emerging small world*

Sanjeev Goyal[†]
University of Essex and Tinbergen Institute

Marco J. van der Leij[‡]
Tinbergen Institute and Erasmus University Rotterdam

José Luis Moraga-González[§]
University of Groningen, Erasmus University Rotterdam and CESifo

forthcoming in *Journal of Political Economy*

November 2, 2005

Abstract

We study the evolution of social distance among economists over the period from 1970 to 2000. While the number of economists has more than doubled, the distance between them, which was already small, has declined significantly. The key to understanding the short average distances is the observation that economics is spanned by a collection of inter-linked stars. A star is an economist who writes with many other economists, most of whom have few co-authors and generally do not write with each other.

*We thank the Editor, Robert Shimer, an anonymous referee and participants in various seminars for useful comments.

[†]E-mail: sgoyal@essex.ac.uk

[‡]E-mail:mvanderleij@few.eur.nl

[§]E-mail: j.l.moraga.gonzalez@rug.nl

1 Introduction

It is often argued that due to a series of technological and economic developments – such as the deregulation of airlines and telecommunications, the rise of facsimile technology and the internet– it is becoming cheaper for individuals to form and maintain more distant ties. This in turn, it is claimed, will reduce the 'distance' between people and will make the world 'smaller'.[1] We examine this argument by analyzing the evolution of social distance amongst economists who publish in journals, during the period from 1970 to 2000.

We split this period into three ten year intervals, 1970-1979, 1980-1989, and 1990-1999. Every publishing author is a node in a network, and two nodes are linked if they have published a paper or more together in the period under study. We thus have three co-authorship networks – one corresponding to each decade – and we examine whether these networks have become more integrated over time. In a network, two economists who co-author a paper are at a distance 1 from each other, while economists who do not write with each other but have a common co-author are at a distance 2 from each other, and so on. All economists who are either directly or indirectly linked with each other are said to belong to the same component and we shall refer to the largest group of inter-connected economists as the giant component. We will interpret a larger relative size of the giant component and a shorter average distance between economists in the giant component as evidence for the world becoming smaller.

[1]The popularity of terms such as 'globalization', the 'death of distance' and 'global village' is one indication of this widespread feeling; general references include Cairncross (2001). For formal studies of evolving social networks see Rosenblat and Mobius (2004) and Moody (2004).

2 Framework

Let $N = \{1, 2, ..., n\}$ be the set of authors in a network. For two authors, $i, j \in N$, we define $g_{i,j} \in \{0, 1\}$ as the academic relationship between them, with $g_{i,j} = 1$ signifying that the two authors have published one or more papers together and $g_{i,j} = 0$ otherwise. The collection of authors and the links between them yields a network of collaboration G.

Let $\mathcal{N}_i(G) = \{j \in N : g_{i,j} = 1\}$ be the set of authors with whom i collaborates in network G. The number of co-authors of a person i, $\eta_i(G) = |\mathcal{N}_i(G)|$, is referred to as the *degree* of individual i in network G. The *average degree* in a network G is $\eta(g) = \sum_{i \in N} \eta_i(G)/n$. We say that there is a path between authors i and j either if $g_{i,j} = 1$ or if there is a set of distinct intermediate co-authors $j_1, j_2, ..., j_n$, such that $g_{i,j_1} = g_{j_1,j_2} = ... = g_{j_n,j} = 1$. Two persons belong to the same component if and only if there exists a path between them. The components can be ordered in terms of their size, and we say that the network has a *giant component* if the largest component comprises a relatively large part of the population of economists and all other components are small (typically of order $\ln(n)$).

The distance between two authors i and j in network G, denoted $d(i, j; G)$, is the length of the shortest path between them. If there is no path between i and j in a network G then we set $d(i, j; G) = \infty$. For a connected network G (with a path between every pair of nodes), the average distance is given by

$$d(G) = \frac{\sum_{i \in N} \sum_{j \in N} d(i, j; G)}{n(n-1)} \tag{1}$$

The clustering coefficient of a network G is a measure of the overlap between the links of different authors. The level of clustering in the neighborhood of person i is given by

$$C_i(G) = \frac{\sum_{l\in N_i(G)} \sum_{k\in N_i(G)} g_{l,k}}{\eta_i(\eta_i - 1)} \tag{2}$$

for all $i \in N' \equiv \{i \in N : \eta_i \geq 2\}$. This ratio tells us the percentage of a person's co-authors who are co-authors of each other. The clustering coefficient of a network G is defined by the weighted average

$$C(G) = \frac{\sum_{i\in N'} \sum_{l\in N_i} \sum_{k\in N_i} g_{l,k}}{\sum_{j\in N'} \eta_j(\eta_j - 1)} = \sum_{i\in N'} \frac{\eta_i(\eta_i - 1)}{\sum_{j\in N'} \eta_j(\eta_j - 1)} C_i(G) \tag{3}$$

We say that a network G exhibits *small world* properties if it satisfies the following conditions: (1) The number of nodes is very large as compared to the average number of links, $n >> \eta(G)$. (2) The network is integrated; a giant component exists and covers a large share of the population. (3) The average distance between nodes in the giant component is small, $d(G)$ is of order $\ln(n)$. (4) Clustering is high, $C(G) >> \eta(G)/n$. This definition extends the one given by Watts (1999) by adding requirement 2.

3 Empirical Findings

We study the world of economists who published in journals included in EconLit. We cover all journal articles that appear in the 10 year windows 1970-1979, 1980-1989 and 1990-1999. The list of journal articles includes all papers in conference proceedings, as well as short papers and notes. We do not cover working papers and work published in books. In mapping the data onto the network, we distinguish different authors by their last name and the initials of all their first names.[2] About 1.6% of the articles have four

[2]We borrow this procedure from Newman (2001). This procedure is potentially subject to problems of underreporting (as when two distinct authors have common initials and surnames) as well as over-

or more authors and EconLit only reports the first author followed by the extension *et al.*; we therefore exclude these articles from our analysis.[3]

3.1. The small world hypothesis: We start by noting from Table 1 that the number of authors has grown from 33,770 in the 1970s to 81,217 in the 1990s. Thus the number of journal publishing economists is large and has grown substantially – more than doubling – over the period 1970 to 2000.

We now turn to the first statistic in the definition of a small world, the average number of co-authors. Table 1 shows that over a ten year period a typical economist has no more than 2 coauthors. Comparing the average degree of the networks with the total number of authors leads us to our first finding: *the average number of co-authors of an economist is very small relative to the total number of economists.*

We next discuss the existence and size of a giant component. Table 1 tells us that in the 1970s the largest component contained 5,253 authors, which constituted about 15.6% of the population. This largest component has expanded substantially over time and in the 1990s it contains 33,027 persons, which is roughly 40% of all economists. At the same time, there has been a sharp fall in the proportion of isolated individuals and in the size of the second largest component. These points are summarized in our second finding: *the giant component has grown substantially; it covered 15% of the nodes in the 1970s and over 40% of the nodes in the 1990s.*

---

reporting (as when an author appears with different initials in different articles). We have considered a number of alternative name extraction procedures and the main findings are robust; details of these procedures and the results can be obtained from the authors.

[3]Using other sources of information and the world wide web, we collected the missing names for a large subset of the journals. Including these articles did not alter our findings qualitatively.

We now turn to the distance between authors in the network. As is the norm, we use the average distance between persons in the giant component as a proxy for our measure of average distance in the network. Table 1 shows that this average distance was 12.86 in the 1970s, 11.07 in the 1980s, and 9.47 in the 1990s. This tells us that average distance has been very small throughout the period under study and moreover that it has declined, by approximately 25%, in spite of the tremendous growth in the giant component. We also note that this fall in average distance has been accompanied with a significant fall in the standard deviation in the distances between nodes from 4.03 in the 1970s to 2.23 in the 1990s.[4] These observations are summarized in our third finding: *the giant component has become significantly "smaller" in terms of distances.*

We next move to the level of overlap between co-authorship, which is measured by the clustering coefficient in the network. Table 1 shows that the clustering coefficient for the network as a whole was 0.193 in the 1970s, 0.182 in the 1980s and 0.157 in the 1990s. Could this network – and in particular these clustering levels – have emerged from a random process of generation of links? If the connections between authors were random, the probability of a relationship being formed would be approximately equal to the average number of co-authors divided by the total number of authors. Since link formation is independent, the clustering coefficient should be approximately equal to that number. For example, in the 1990s the actual clustering coefficient is 0.157, which is more

[4]If we consider distances between all pairs of authors in a giant component as an i.i.d. sample, we can use two-sample t-statistics to test the hypotheses of equal average distance in the 1970s and 1980s giant components and in the 1980s and 1990s giant components. The t-statistic is -1589.2 for the comparison between the 1970s and the 1980s and -4919.0 for the 1980s compared to the 1990s. In both cases the hypothesis of constant average distance is clearly rejected.

than 7,000 times the level predicted by the random process, 0.0000206. Since papers with three co-authors increase the clustering coefficient, we also computed it considering papers with two co-authors only. In the 1990s the clustering coefficient was around 0.015, still more than 700 times the level predicted by a model of random connections. These points put together yield our fourth finding: *the clustering coefficient for the networks is very high throughout the period under study.*

When we set these findings against the criteria for a network to display small world properties, we find that throughout the period 1970-2000 the collaboration networks satisfy properties (1), (3) and (4), i.e. the average degree of the networks under consideration is tiny relative to the number of nodes, distance within the giant component is small and clustering is high. As to criterion (2), we note that the coverage of the giant component was relatively modest in the 1970s but in the 1990s it covered over 40% of the nodes, i.e., a giant component has emerged. Thus in the 1990s the collaboration network satisfies all four criteria. Moreover, in spite of the growth in its size, the average distance within the giant component has declined significantly. This leads us to conclude that *economics is an emerging small world.*

3.2. Inter-linked stars: What is it about the number and arrangement of links in the network that generates these aggregate features? We start with the behavior of the average number of links. Table 1 tells us that there is almost a doubling in the average degree from 0.894 in the 1970s to 1.672 in the 1990s. This leads us to say that *the average number of co-authors is very small but it has been increasing consistently through the* *period 1970-2000.*[5]

We turn now to the inequality in the degree distribution. To get an appreciation of this inequality it is useful to compare the actual degree distribution with the degree distribution in a random network which has the same average degree. This degree distribution is binomial and thus approximately Poisson for large networks. We find that the variance in the actual degree distribution is much larger than the variance in the constructed degree distribution. For example, Table 1 tells us that in the 1990s the variance in the empirical distribution is 5.29, while the variance in the corresponding random network is only 1.67. We also find that the degree distribution is particularly skewed at very high degrees. Table 2 tells us that, in the 1990s, the 100 economists with the highest degree have (on average) 25 links, while the average degree in the population is 1.67. This difference in degree is (roughly) 10 times the standard deviation in the actual network. We summarize these findings by stating that *the distribution of co-authorship in the population of economists is very unequal.*

We now examine more closely the link pattern of the individuals in the network. Figure 1 shows that, for each of the decades, there is a clear negative relationship between clustering and degree. Let us look at the local network of the most connected individuals more closely. Table 2 tells us that, in the 1990s, the most connected author wrote 66 papers, had 54 co-authors and a clustering coefficient of 0.02. Thus the most connected individual collaborated extensively and most of his co-authors did not collaborate with

[5]As in the case of the average distances in the giant component, we can use two-sample t-statistics to test the hypothesis that the average number of collaborators is constant over time. The t-statistic is 32.1 for the comparison between the 1970s and the 1980s and 37.6 for the 1980s compared to the 1990s. The hypothesis of constant average degree is clearly rejected.

each other. Table 2 shows that the local network of the 100 most linked authors exhibits similar properties. These economists can thus be viewed as *'stars'* from the perspective of the network architecture. Figure 2 presents the local network of Joseph E. Stiglitz as an illustration. We summarize our observations as follows: *there are many 'stars' in the world of economics.*

<insert figures 1 and 2 here>

We next study the role of the stars in connecting different parts of the network. Here we follow the procedure of Albert *et al.* (2000) and compare the consequences of randomly deleting nodes as against deleting star nodes. We find that the removal of 5% of the authors at random leads to a marginal change in the giant component and clustering, while the deletion of the 5% most linked nodes leads to a complete breakdown of the giant component and a sharp increase in the clustering coefficient. For instance, in the 1990s the deletion of 5% of the nodes at random leads to a marginal fall in the size of the giant component from 40.7% to 38.9%, while the average distance within the giant component increases slightly from 9.47 to 9.68. By contrast, a removal of the 5% most connected nodes leads to a complete breakdown of the giant component and an increase in clustering from 0.157 to 0.344. This suggests that stars play the role of connectors and sharply reduce distance between different highly clustered parts of the world of economics. We therefore conclude that *the world of economists has been and still is spanned by a collection of inter-linked stars and that this is critical for understanding the short average distances.*

4 Concluding Remarks

In this paper we have found substantial evidence that the world of economists is becoming smaller. Further, we have identified stable and changing features of the structure of co-authorship in economics. The analysis allows us to make two general points. The *first* point is about a stable feature of the network: inter-linked stars span the network of collaboration and this explains the small average distance between economists. The *second* point is about an important change: there has been a significant increase in the average degree of the network.[6]

These results are very striking and lead us to ask questions about the process of network formation. In particular, we would like to understand the economic determinants of co-authorship better and how the inter-linked star architecture comes about. Our findings also raise questions about the impact of social interaction on scientific discovery and the diffusion of knowledge. We hope to explore these issues in future work.

References

Albert, Reka, Hawoong Jeong, and Albert-László Barabási. 2000. "Error and Attack Tolerance of Complex Networks." *Nature* 406 (July): 378–382.

Cairncross, Frances. 2001. *The Death of Distance: How the Communications Revolution*

[6]We have checked the robustness of these results by examining alternative data sets: 1) journals considered by the Tinbergen Institute (The Netherlands) to assess the research output of its fellows; 2) journals that were published and covered by EconLit throughout the period from 1970 to 1999; 3) a set of five core journals in economics (*AER, Econometrica, QJE, JPE* and *REStud*). Analysis of these data suggests that the qualitative features of our findings are robust. Details on the data, the list of journals and the results are provided in our working paper, Goyal, van der Leij, and Moraga-González (2004).

is Changing our Lives (2nd ed.). Boston: Harvard Business School Press.

Goyal, Sanjeev, Marco J. van der Leij, and José Luis Moraga-González. 2004. "Economics: An Emerging Small World?" TI Discussion Paper no. 04-001/1 (January), Tinbergen Institute, Amsterdam/Rotterdam, The Netherlands.

Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69 (April), 213–238.

Newman, Mark E. 2001. "The Structure of Scientific Collaboration Networks." *P.N.A.S.* 98 (January), 404–409.

Rosenblat, Tanya S., and Markus M. Mobius. 2004. "Getting Closer or Drifting Apart?" *Quarterly Journal of Economics* 119 (August), 971–1009.

Watts, Duncan J. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton Studies in Complexity. Princeton: Princeton University Press.

Table 1: Network statistics for the co-author networks.

decade	1970s	1980s	1990s
total authors	33770	48608	81217
average degree	.894	1.244	1.672
standard deviation of degree	1.358	1.765	2.303
size of giant component	5253	13808	33027
—as percentage	.156	.284	.407
second largest component	122	30	30
isolated authors	16735	19315	24578
—as percentage	.496	.397	.303
clustering coefficient	.193	.182	.157
average distance in giant component	12.86	11.07	9.47
standard deviation of distance in giant component	4.03	3.03	2.23

Table 2: Network statistics for most linked economists: 1990s.

Rank	Papers	% Co-authored	Degree	Distance 2	Clust.Coeff
1	66	.970	54	244	.022
2	58	.586	45	158	.019
3	67	1.000	41	172	.045
4	67	.940	41	57	.034
5	48	.938	34	169	.036
Average top 100	37.69	.849	25.31	99.40	.040
Average all	2.82	.409	1.67	3.57	.157

Note: economists are ordered by 1) degree and 2) number of nodes at distance 2. *Papers* is the number of papers published by economist i. *% Co-authored* is the fraction of papers published by i that are co-authored. *Degree* is the degree of i. *Distance 2* is the number of nodes at distance 2 from i. *Clust.Coeff* is the clustering coefficient of i. Average clustering coefficients are calculated as in (3).

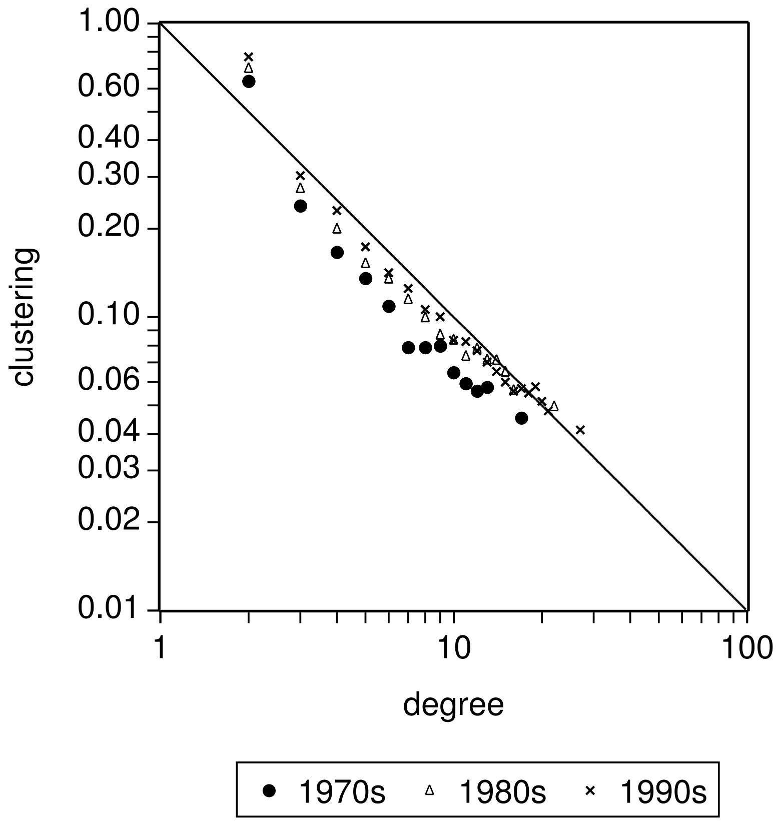

Figure 1: Average clustering and degree distribution

Note: Clustering for a given degree $k < \overline{k}$ ($\overline{k}$ is 14 in the 1970s, 17 in the 1980s and 22 in the 1990s) is the clustering coefficient of nodes with degree k. Observations for degree $k \geq \overline{k}$ are grouped together.

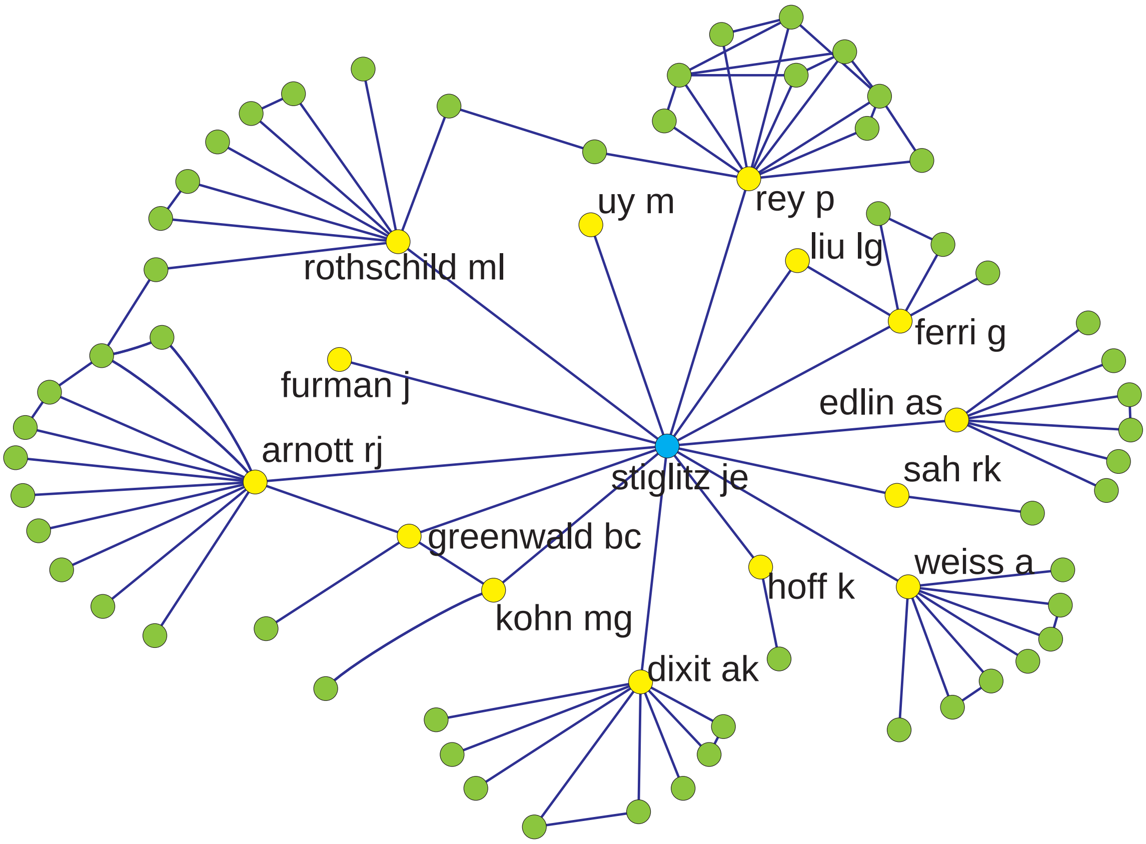

Figure 2: Local network of collaboration of Joseph E. Stiglitz in the 1990s.

Note: The figure shows all authors within distance 2 of J.E. Stiglitz as well as the links between them. Some economists might appear twice or are missing due to the use of different initials or misspellings in EconLit. The figure was created by software program *Pajek*.