Journal of Informetrics 2 (2008) 317–325

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# International collaboration in science and the formation of a core group

Loet Leydesdorff[a,*], Caroline S. Wagner[b]

[a] University of Amsterdam, Amsterdam School of Communications Research (ASCoR), Kloveniersburgwel 48, 1012 CX Amsterdam, The Netherlands
[b] SRI International, Arlington, VA 22209, USA

ARTICLE INFO

*Article history:*
Received 31 May 2008
Received in revised form 16 July 2008
Accepted 17 July 2008

*Keywords:*
Scientific collaboration
Social network analysis
Science policy
Social systems
Co-authorship

ABSTRACT

International collaboration as measured by co-authorship relations on refereed papers grew linearly from 1990 to 2005 in terms of the number of papers, but exponentially in terms of the number of international addresses. This confirms Persson et al.'s [Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometrics values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, *60*(3), 421–432] hypothesis of an inflation in international collaboration. Patterns in international collaboration in science can be considered as network effects, since there is no political institution mediating relationships at that level except for the initiatives of the European Commission. Science at the international level shares features with other complex adaptive systems whose order arises from the interactions of hundreds of agents pursuing self-interested strategies. During the period 2000–2005, the network of global collaborations appears to have reinforced the formation of a core group of fourteen most cooperative countries. This core group can be expected to use knowledge from the global network with great efficiency, since these countries have strong national systems. Countries at the periphery may be disadvantaged by the increased strength of the core.



## 1. Introduction

An increasing share of scientific papers is co-authored by scientists from two or more nations. During the 1990s, a rapid rise occurred in internationally co-authored papers indicating a rise in collaboration (Doré, Ojasoo, & Okubo, 1996; Georghiou, 1998; Glänzel, 2001). The increase was dramatic: international collaborations (as measured by internationally co-authored publications) doubled (Wagner & Leydesdorff, 2005a). The increase can be seen across all fields of science at more or less the same rate (Wagner, 2008). Collaboration continued to rise in the early 2000s. The number of internationally co-authored articles grew at a rate faster than traditional "nationally co-authored" articles (NSB, 2002). Moreover, internationally co-authored articles appear to be cited more often than nationally co-authored papers (Narin, 1991; Persson, Glänzel, & Danell, 2004).

We suggest that international collaboration in science can be considered as a communications network that is different from national systems and has its own internal dynamics (Gibbons et al., 1994; Price, 1963; Ziman, 1994; Katz & Hicks, 1998; Wagner, 2008; Wagner & Leydesdorff, 2005b). National systems have policies and institutions that mediate scientific communication, while at the global level the network exists primarily as a self-organizing system. The exception here is the European Union, where specific incentives exist to encourage formal international linkages among member countries.

* Corresponding author. Tel.: +31 20 525 6598; fax: +31 20 525 3681.
*E-mail addresses:* loet@leydesdorff.net, l.a.leydesdorff@uva.nl (L. Leydesdorff), Caroline.wagner@sri.com (C.S. Wagner).
*URLs:* http://www.leydesdorff.net (L. Leydesdorff), http://www.cswagner.net (C.S. Wagner).

1751-1577/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2008.07.003

Does the EU emerge as an "international actor" or are different patterns (e.g., elite structures including the USA) more dominant?

## 2. Data and methodology

Data were harvested from the CD-Rom version of the *Science Citation Index* for articles, reviews, letters, and notes for 1990, 2000, and 2005.[1] In his study entitled *Evaluative Bibliometrics*, Narin (1976) proposed counting only articles, reviews, and notes as indicators of scientific performance. Braun, Glänzel, and Schubert (1989) argued in favor of including letters as scientific output. However, the Institute for Scientific Information (ISI) no longer registered for the category of "notes" after 1995. We included 3090 internationally co-authored notes in the counts for 1990 because the data was already organized in this format during a previous project. Table 1 provides an overview of the data for the three years in question.

For example, of the 1,011,363 records contained in the *Science Citation Index 2005*, only the 734,750 articles, reviews, and letters were considered. Among these documents, 171,402 were internationally co-authored; this is 23.3% of the total in column (1).

Collaboration was indicated by a co-authorship event at the document level. The country counts were done using integer counting, which attributes a count of "1" to each occurrence of authorship from a country among the set.[2] This leads to an asymmetrical matrix of documents versus countries. The cosines are computed on the basis of this matrix (Leydesdorff, 1989; Leydesdorff & Vaughan, 2006).

Since the distributions are not expected to be normal, it has been suggested that the cosine instead of the Pearson correlation is the proper measure for normalization (Ahlgren, Jarneving, & Rousseau, 2003; Boyack, Klavans, & Börner, 2005; cf. Hamers et al., 1989). The cosine normalizes to a geometric mean (rather than an arithmetic mean) and the consequent vector space model (Salton & McGill, 1983) is useful for the visualization of latent structures in the set. Since the cosine runs from zero to one, a very small number of relations can be expected to generate a cosine larger than zero. We considered cosine >0.01 as a relevant threshold for discarding this incidental variation. Incidental variations may be caused by ongoing relations between supervisors and students or postdocs who have returned to their home countries.

The co-authorship events were additionally placed into a symmetrical matrix where country names appear on both axes, with the number of co-occurrence events appearing in the corresponding cell. Normalization, however, was based on the asymmetrical occurrence matrix because this matrix contains all information at the document level, including co-authorship relations among more than two countries. Leydesdorff and Vaughan (2006) showed that using the symmetrical co-occurrence matrix – which is based on multiplication of the asymmetrical occurrence matrix with its transposed – may lead to faulty conclusions about the underlying structure because information is lost. For example, correlations among co-occurrences can be spurious when based on multilateral co-occurrences at the document level (Waltman and van Eck, 2007; Leydesdorff, 2007a).

Both the co-occurrence and the normalized tables were used to conduct network analysis using UCINET and Pajek software (De Nooy, Mrvar, & Batagelj, 2005). The normalized data can reveal structures such as resemblances in patterns which in non-normalized data are overshadowed by the effect of stars in the network with a high degree of centrality (e.g., the USA; Leydesdorff, 2007b). The results are presented below.

## 3. Results

Fig. 1 visualizes the growth in international collaboration on the basis of the data provided in Table 1 above. The number of internationally co-authored publications has grown linearly ($r^2 > 0.99$). However, Fig. 1 shows an exponential growth in the number of addresses of internationally collaborating authors ($r^2 > 0.99$), suggesting that the growth of the network extends to many more places around the globe, with a corresponding growth in the possibility of knowledge diffusion. The average number of addresses in any one internationally co-authored publication has grown from an average of 2.86 in 1990 to 3.61 in 2005, and this trend is accelerating.

In other words, internationally co-authored publications are increasingly multi-nationally co-authored. Because the number of records increases only linearly, Persson et al. (2004) concluded that this trend indicates an inflation in international collaborations.

Table 2 shows the results of the social network analysis for the global network of collaborations for the 3 years studied. Social network analysis provides us with a large number of statistics (Newman, 2000). First, the number of nodes in Table 2 represents the number of countries with authors participating in global science: this number increased by 20 between 1990 and 2000, with half of this growth due to the break-up of the Soviet Union into individual states, many of which began to participate in science as separate political entities during the decade of the 1990s. Between 2000 and 2005, the number of countries in the data set increased only from 192 to 194.

[1] Because we were surprised by our findings we repeated the complete analysis with 2006 data, but the results are not essentially different from those of 2005. The discussion thus focuses on these 3 years.

[2] No effort was made to remove counts where the same author lists two different country addresses (Persson et al., 2004). We assume that authors listing two addresses have colleagues in both institutions and therefore fall within a broad definition of collaboration.