Scientometrics
DOI 10.1007/s11192-016-1878-5

# Evolution and structure of scientific co-publishing network in Korea between 1948–2011

**Jinseok Kim[1] · Liang Tao[2] · Seok-Hyoung Lee[3] · Jana Diesner[1]**

Received: 23 June 2015
© Akadémiai Kiadó, Budapest, Hungary 2016

**Abstract** This study investigates the evolution and structure of a national-scale co-publishing network in Korea from 1948 to 2011. We analyzed more than 700,000 papers published by approximately 415,000 authors for temporal changes in productivity and network properties with a yearly resolution. The resulting statistical properties were compared to findings from previous studies of coauthorship networks at the national and discipline levels. Our results show that both the numbers of publications and authors in Korea have grown exponentially in a 64 year time frame. Korean scholars have become more productive and collaborative. They now form a small-world-ish network where most authors can connect with one other within an average of 5.33 degrees of separation. The increasingly skewed distribution and concentration of both productivity and the number of collaborators per author indicate that a relatively small group of individuals have accumulated a large number of opportunities for co-publishing. This implies a potential vulnerability for the network and its wider context: the graph would disintegrate into a multitude of smaller components, where the largest one would contain <2 % of all authors, if approximately 15 % (57,724) of the most connected scholars left the network, e.g., due to retirement or promotion to higher-level administrative positions.

**Keywords** Bibliometrics · Coauthorship networks · Authority control · Network evolution · Small-world networks

✉ Jinseok Kim
jkim362@illinois.edu

[1] Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel St., Champaign, IL 61820, USA

[2] Department of Agricultural Engineering, University of Illinois at Urbana-Champaign, 1304 W. Pennsylvania, Urbana, IL 61801, USA

[3] Department of Overseas Information, Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-Gu, Daejeon 305-806, Korea

## Introduction

Collaboration has become a fundamental mode of scientific research. This is confirmed, for example, by the constant decline of the ratio of single-authored papers (Abt 2007). In order to explain the structural patterns and evolution of scholarly collaboration, researchers have applied social network analysis, among other techniques, to bibliometric data. From a network-analytic perspective, scientific collaboration is represented as authors (nodes) and links that are established between authors who have jointly authored at least one publication.

Analyzing the aggregate of those individual coauthorship relations in a specific discipline allows for investigating a variety of graph-theoretical and substantive questions about collaboration networks. For instance, scholars choose to collaborate with others who already have many collaborators; a phenomenon that is also referred to as the preferential attachment (Barabási et al. 2002). This mechanism has been shown to lead to the emergence of a scale-free network where the distribution of the node degree (i.e., the number of collaborators per author) follows a power law (Barabási et al. 2002; Newman 2001). Furthermore, prior work suggests that any random pair of scholars can reach each other via a small number of intermediaries in such a network; confirming the potential applicability of the small-world phenomenon to coauthorship networks (Newman 2000, 2004).

Scale-free as well as small-world properties have been reported for coauthorship networks from various fields: biomedicine (Newman 2001), computer science (Franceschet 2011), neuroscience (Barabási et al. 2002), physics (Liben-Nowell and Kleinberg 2007), and interdisciplinary collaboration (Börner et al. 2004). Other studies have focused on national-scale or international collaboration data, arriving at similar results: e.g., for the cases of Italy (Abramo et al. 2013), Japan (Yoshikane and Kageura 2004), Slovenia (Lužar et al. 2014; Perc 2010), and Turkey (Çavusoglu and Türker 2013; Gossart and Özman 2009).

In this paper, we describe and explain the evolution of the scientific collaboration network in Korea. Our goal is to understand the structure of an entire national coauthorship network with its change over a large time span (64 years), and to compare our findings to other studies of national-scale coauthorship networks. As scholars have suggested, international and domestic collaborations are affected by a variety of factors such as geographic distance and the language that scholars use (Schubert and Glanzel 2006). We argue that studying domestic collaboration networks is essential for developing theories about the evolution of international collaboration that takes cultural differences into account and allows for a more holistic understanding of coauthorship patterns across nations.

In working towards this goal, we construct a coauthorship network from more than 700,000 domestic publications in Korea. Here, domestic means papers published in Korea and includes studies by foreign scholars who reside either in Korea or overseas.[1] Scientific performance and citation patterns of Korean scholars have been previously investigated (e.g., Choi et al. 2011; Kim 2005; Park and Leydesdorff 2008). Our unique findings discussed in this paper are based on the structural analysis of collaboration patterns among domestic Korean scholars, which is an insufficiently researched area.

Another unique feature of this study is the underlying dataset, which covers not only the majority of scholarly publications (both from journals and conferences) published in Korea, but is also of high quality in terms of entity resolution. The author names were

[1] More than 97 % of the authors identified were found to be associated with universities, institutes, or companies in Korea. Although there might be scholars who reside overseas and may participate in papers published in Korea, we assume that that does not affect our study.

identified via a combination of highly accurate algorithmic solutions and manual disambiguation for uncertain cases (this work was completed by KISTI and is discussed in more detail in the next section). We expect this level of quality to allow us to map the collaboration patterns more correctly than if we relied on un-disambiguated data (Fegley and Torvik 2013; Kim and Diesner 2015; Kim et al. 2014). Similar to other collaboration network studies, our analysis is based on several assumptions. First, we consider collaboration networks as a medium for the diffusion of scientific information and knowledge. Second, we assume that coauthorship networks provide information about scientific collaboration among scholars (hereafter and in the context of this paper, we use "coauthors" as a synonym for "collaborators"). Those assumptions may limit the meaning and interpretation of our results.

## Dataset and measurements

The dataset for this study was obtained from the National Digital Science Library (NDSL), built by the Korea Institute of Science and Technology Information (KISTI). KISTI is a government-funded organization in charge of managing scientific publication information nationwide and abroad. Our instance of the dataset (KISTI data hereafter) includes publication information for 703,073 papers that were published between 1948 and 2011: 494,689 journal articles (71.56 %) and 208,384 conference papers (28.44 %).

One noticeable feature of the dataset is its quality. Asian names, including Korean ones, are known to be more difficult to disambiguate than Western names, mainly due to the strong prevalence of common surnames and given names (Strotmann and Zhao 2012; Torvik and Smalheiser 2009). KISTI took care of disambiguating author names as follows: first, author name instances were algorithmically clustered using features such as author name string, affiliation, coauthor, keywords of the publication, and publication venue. With this step, an accuracy of 93.44 % (F1) was achieved when tested against the ground-truth data of approximately 35,000 authors in nearly 12,000 papers. Second, the results of automatic clustering were passed to the next step of further manual refinement, where human experts checked and resolved erroneous cases leading to an accuracy of up to 98.00 %.

Using this disambiguation procedure, a total of 415,695 unique authors (335,937 from journal papers and 177,849 from conference papers) were identified in our dataset. The number of authors who appeared both in journal and conference papers was 98,091. For this project, we do not distinguish between journal and conference papers for analysis, because more than half of the authors of conference papers also appear as authors of journal papers, and because our purpose is to map the entire domestic collaboration structure.

If any two authors published a paper together, in the dataset, these authors are connected by a coauthor relationship. Like in previous studies (Barabási et al. 2002; Moody 2004; Newman 2001), we ignore the frequency and only consider the presence of collaboration relationships. The coauthorship network constructed through this process was analyzed by calculating several commonly used network metrics. More specifically, we identified the number of coauthors per paper, the number of collaborators per author, the clustering coefficient [a.k.a. transitivity (Newman 2001)], the proportionate ratio of the largest component, and the length of the shortest path (Brandes 2001). These measures were selected to make the findings of our study comparable to other relevant coauthorship network papers. Changes in these metrics were tracked over time.

Additionally, the inequality in (a) productivity and (b) the number of collaborators per authors was measured by using the Gini coefficient, following the example of prior coauthor network studies (Franceschet 2011; Martin et al. 2013; Yoshikane and Kageura 2004). In our paper, the Gini coefficient is calculated as follows (Glasser 1962):

$$G = \frac{1}{2\mu n^2}\sum_{i=1}^{n}(2i - n - 1)X_i \quad (n > 1) \tag{1}$$

Here $X_i$ is the publication frequency and number of collaborators of an author $X$ sorted from smallest to largest, $n$ is the total publication frequency and number of collaborators of all authors observed, and $\mu$ is the mean frequency. The value of $G$ can range from 0 (all authors have the same number of papers and coauthors) and 1 (only one author publishes all papers and monopolize collaborators).

## Results

We first turn our attention to temporal trends in publication volume (Fig. 1) and number of authors (Fig. 2). In both figures, solid lines represent all publications and authors, respectively, while dashed lines show the number of publications and authors excluding single-authored papers. For example, the total number of papers prior to 2011 is 703,073, while there were 530,994 (75.52 %) multi-authored papers.

Both graphs show an exponential growth over time. For example, when transformed into a log-linear graph, both solid lines can be fitted by an exponential function with an exponent of 0.18 ($R^2 = 0.97$) for publications and 0.17 ($R^2 = 0.97$) for authors. Possible explanations of this development are (1) Korea's economic growth, which began to accelerate around the mid-1980s, and (2) the subsequent increase in investments in R&D. The steep rise after the mid-1990s might be attributable to a strong focus on education: the number of universities rose from 242 in 1990 to 359 in 2009, and the number of university-

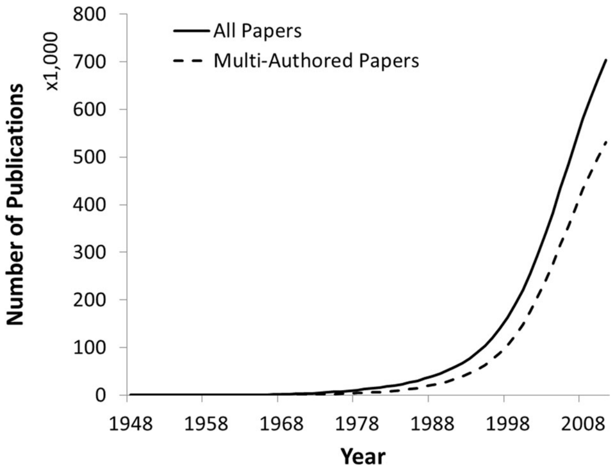


**Fig. 1** Cumulative number of publications per year. Each curve shows an exponential growth with exponents of 0.18 ($R^2 = 0.97$) for all papers (*solid line*) and 0.19 ($R^2 = 0.99$) for papers with two or more coauthors (*dashed line*)

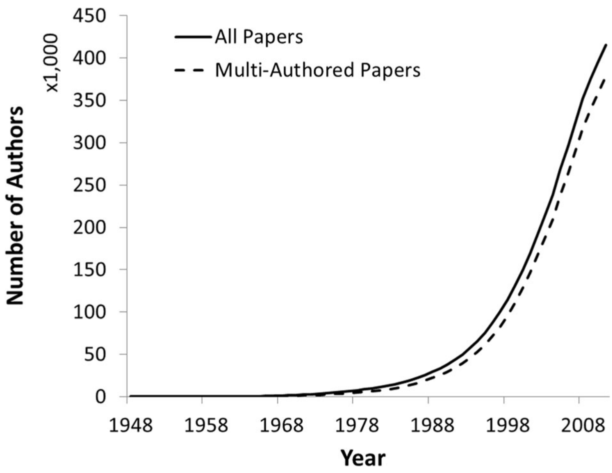


**Fig. 2** Cumulative number of unique author per year. Each curve shows an exponential growth with exponents of 0.17 ($R^2 = 0.97$) for all papers (*solid line*) and 0.17 ($R^2 = 0.98$) for papers with two or more coauthors (*dashed line*)

level teachers increased from 62,000 in 1990 to 230,000 in 2009 (Ministry of Education 2013; Ryoo 2011). A similar exponential growth in scientific productivity was reported for Turkey during the 1980–2010 period with exponents of 0.18 (publications) and 0.16 (authors) (Çavusoglu and Türker 2013). A few studies of academic fields reported similar exponential growth in publications and/or authors: computer science (Franceschet 2011)

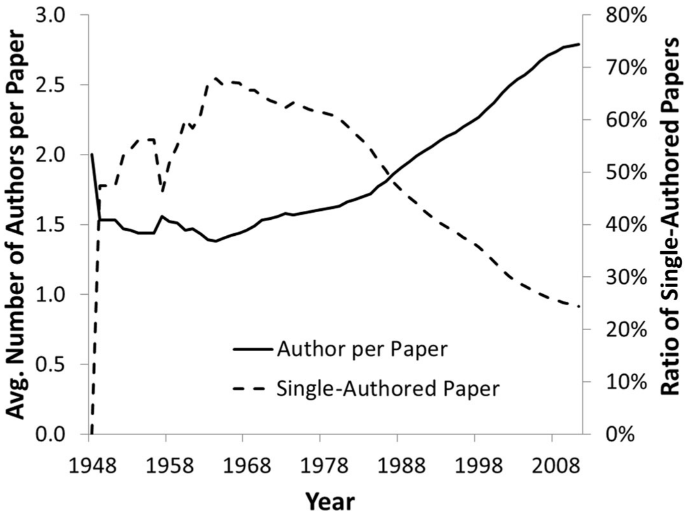


**Fig. 3** Average number of authors per paper (*left y*-axis; *solid line*) and ratio of single-authored papers (*right y*-axis; *dashed line*). Papers published in recent years tend to have more coauthors than older ones, which can be coupled with the observation that the ratio of papers written by a single author has continued to decrease since 1964

**Table 1** Statistical properties of comparative collaboration networks

| Country or field (data source) | Years | # of papers (# of authors) | Avg. # authors per paper | Avg. # papers per author | Avg. # coauthors per author | Ratio of largest component (%) | Clustering coefficient | Avg. shortest path length (diameter) |
|---|---|---|---|---|---|---|---|---|
| Korea (Korea Institute of Science and Technology Information) | 1948–2011 | 703,073 (415,695) | 2.79 | 4.73 | 7.38 | 87.50 | 0.19 | 6.33 (28) |
| Slovenia (Current Research Information System) | 1960–2010 | 76,194 (7380) | – | – | 10.7 | 98.7 | 0.20 | 4.6 (14) |
| Turkey (ISI Web of Science) | 1980–2010 | 237,409 (151,745) | 4.08 | 1.56 | 35.03 | – | 0.75 | 4.14 (–) |
| Biomedicine (MEDLINE) | 1995–1999 | 2,163,923 (1,520,251) | 3.75 | 6.4 | 18.1 | 92 | 0.066 | 4.6 (24) |
| Computer Science (DBLP) | 1936–2008 | 1,216,526 (731,333) | 2.56 | – | 6.63 | 85 | 0.24 | 6.41 (23) |
| Mathematics (Mathematical Reviews) | 1940–1999 | 1,598,000 (337,000) | 1.45 | 6.87 | 3.92 | 82 | 0.15 | 7.6 (27) |
| Physics (American Physical Society) | 1893–2009 | 458,799 (227,709) | 3.34 | 6.74 | 17.77 | 97.31 | 0.213 | 4.84 (15) |
| Sociology (Sociological Abstracts) | 1963–1999 | 281,090 (197,976) | 1.56 | – | – | 53 | 0.19 | 9.81 (–) |

Parameters obtained for Slovenia (Perc 2010), Turkey (Çavusoglu and Türker 2013), biomedicine (Newman 2001), computer science (Franceschet 2011), mathematics (Grossman 2002), and sociology (Moody 2004). For physics, we calculated parameters of the APS data (papers with 50 authors or fewer) disambiguated by the same algorithm as in Martin et al. (2013). The average number of authors per paper in sociology was calculated based on Table 1 in Moody (2004) only for the 1989–1999 period and physics (Martin et al. 2013). In contrast to that, other fields such as neuroscience and mathematics show a linear growth rate in author numbers (e.g., Barabási et al. 2002).

The increased tendency of scholars to collaborate is associated with the decrease of single-authored papers (Fig. 3); with the latter reaching a peak of 67.85 % in 1964, and then decreasing to 24.48 % in 2011. In line with this trend, the average number of authors per paper (Fig. 3 left *y*-axis) reached its minimum of 1.38 in 1964 and then increased to 2.79 in 2011. These trends towards more collaboration have been confirmed for science overall (Waltman 2012) and on the national level (Çavusoglu and Türker 2013). In the KISTI data, the average number of authors per paper (2.79) seems to be greater than or equal to the fields of mathematics, computer science, and sociology, but less than the values for biomedicine and physics, or the country of Turkey (see Table 1). These differences may be due to the fact that the social sciences and humanities, where single authorship or small-sized collaboration teams are dominant, constitute almost half of the domestic publications in Korea (Han et al. 2009).

The productivity of Korean scholars and the number of their collaborators have continuted to increase since 1948 (Fig. 4). The plateau seen in the 1950s may be attributable to the Korean War (1950–1953) and its aftermath. The average number of papers per author in 2011 was 4.73, which is greater than that for Turkey, but less than those for biomedicine, mathematics, and physics (for comparison, see Table 1). The average number of unique collaborators per author is 7.38, which is greater than those for computer science and mathematics, but less than those for Slovenia, Turkey, biomedicine, and physics (for comparison, see Table 1).

To gain a deeper understanding of the productivity of authors, a cumulative distribution of the number of papers per scholar is plotted in Fig. 5. Here, the values on the *x*-axis represent the number of publications. "Pr ($X \geq x$)" refers to the proportion of authors who have a value of *x* or above. For example, authors with one or two papers make up 47.7 % of all authors, and authors who have published 55 or more papers account for only 1 % of all authors. The plot shows that a small number of authors have large number of

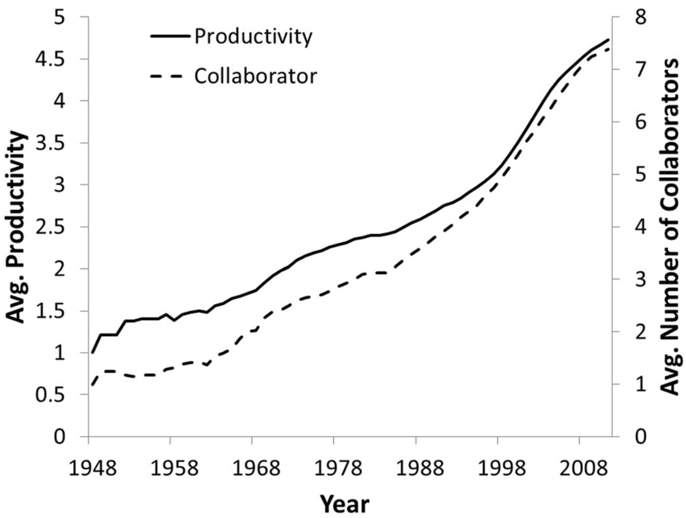


**Fig. 4** Average productivity (*left y*-axis; *solid line*) and average number of collaborator (*right y*-axis; *dashed line*). Both plots show a consistent increase from 1948 to 2011, indicating that Korean scholars have become more productive and collaborative over time

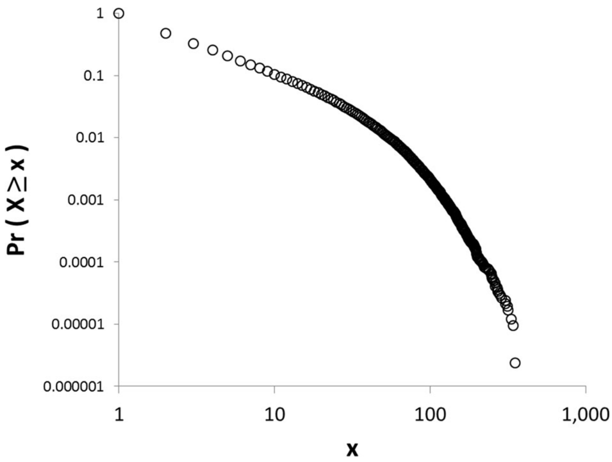


**Fig. 5** Cumulative log–log plot of productivity distribution as of 2011. The plot shows the proportion of authors (*y*-axis) who have published x or more papers (*x*-axis) against the total number of authors. The plot indicates that a small number of scholars have published many papers, while the majority of scholars have published a small number of papers

publications, while the majority of scholars have a small number of publications. Some studies have demonstrated that the slope of productivity plots form an almost straight-line; indicating that scientific productivity follows a power law; an effect also known as Lotka's law (Çavusoglu and Türker 2013; Martin et al. 2013; Newman 2001). In our data, however, we could not find any statistically meaningful power-law regime.[2]

Similar to the results for productivity, the cumulative distribution of the number of collaborators per author (circles in Fig. 6) suggests that a small number of authors have a large number of collaborators, while the majority have a small set of coauthors. Unlike in several previous studies, a power law fit, even with an exponential cutoff, does not approximate the KISTI data with the method of Clauset et al. (2009). To conduct a comparison of node degree distribution in KISTI versus a generic power law network, we generated a scale-free network, i.e., a network with a power law distribution of the number of coauthors per author, using a similar network size (378,296 nodes), average degree (7.38), and number of edges (1,395,357) as in KISTI; following the method used in Kim and Diesner (2015).

The resulting distribution of nodal degree in the synthetic network (crosses in Fig. 6) shows a nearly straight-line slope that approximates a scaling parameter of 2.43. The curve based on the KISTI data does not align with the one from the scale-free network. Rather, the KISTI degree distribution was best fitted by a probability distribution of Generalized Extreme Value (scale = 2.0781, shape = 0.8364, location = 2.4105, negative log likelihood = 1.0532e + 06). Thus, we can say that a scale-free network generated by the preferential attachment mechanism, i.e., the tendency of authors to collaborate with highly collaborative others (Barabási et al. 2002) is not a valid explanation for the formation of the collaboration network in Korea.

[2] In our study, a power law distribution is fitted to the observed degree distribution using the maximum-likelihood fitting method with the goodness-of-fit test based on the Kolmogorov–Smirnov statistic as described in Clauset et al. (2009).

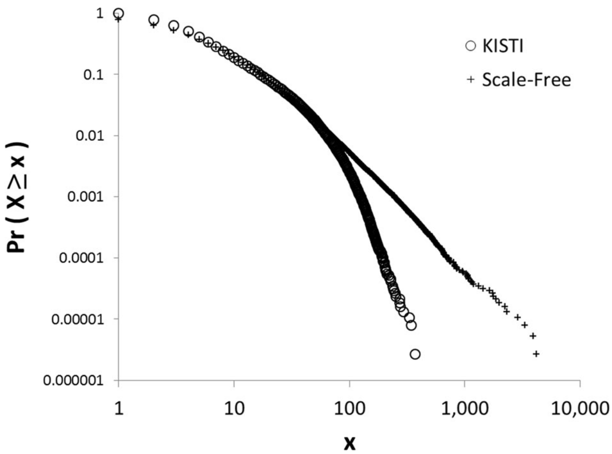


**Fig. 6** Cumulative log–log plots of collaborator distribution in KISTI data (*circles*) as of 2011 and of a synthetic scale-free network (*crosses*). The KISTI data show a negative curve, eventually deviating from the plot of scale-free network generated by the preferential attachment process where the number of nodes and average degree are similar to those of the KISTI network. This implies that the KISTI data are an outcome of a collaboration-seeking mechanism different from preferential attachment

As shown above, the distributions of both productivity and collaboration show a high skewness: a small number of authors have a large number of papers and collaborators, while others have small values. The inequality of these distributions can be measured via the Gini coefficient. For example, the Gini coefficient for collaboration as of 2011 is 0.59. The top 1 % of the most collaborative scholars attract 12.19 % of all collaborations, the top 5 % gather 34.10 %, and the top 10 % accumulate 48.25 %.

Plotting the temporal change in the Gini coefficient of productivity and the number of collaborators (Fig. 7) reveals that inequality has been increasing over time. This means that some scholars have managed to attract more opportunities for publication and collaboration than others have over time. This might be explainable by the Matthew's effect (Merton 1968). When compared to other studies, the Gini coefficient for productivity in 2011 is slightly lower than the coefficient found for physics (0.70) (Martin et al. 2013), while inequality in collaboration in Korea (in 2011) is higher than in the field of computer science (0.56) (Franceschet 2011).

Another structural characteristic of a collaboration network is the connectivity of the graph. Figure 8 shows the temporal change in the ratio of the largest component's size (dashed line). The Korean domestic collaboration network was fragmented until the early 1970s, but the connectivity has sharply increased so that the largest component entails 87.50 % of all network participants. This ratio is higher than those for computer science, mathematics, and sociology, but lower than those for Slovenia, biomedicine, and physics (see Table 1 for comparison).

The size of the largest component, when coupled with the concentration of collaboration as shown in Fig. 7, can provide additional insights into network cohesion. The highly concentrated Korean publication network is vulnerable to decomposition if some highly connected scholars would leave the network, e.g., due to retirement or taking high-level administrative positions. This vulnerability can be measured by the relative size of the largest component after removing a given fraction of nodes from the graph (Franceschet 2011). To assess the impact of node drop-out on graph connectivity, we identified the top

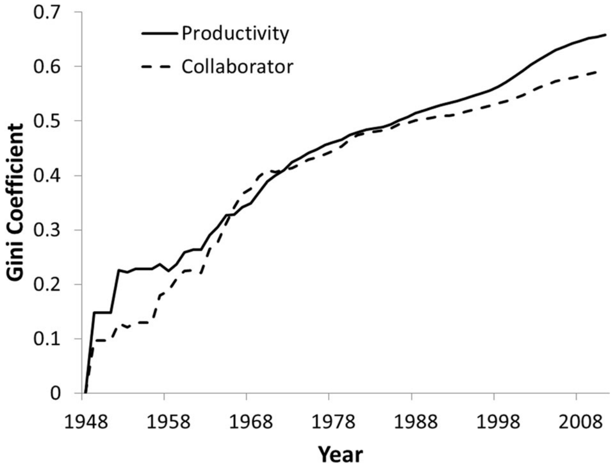


**Fig. 7** Gini coefficients for productivity (*solid line*) and number of collaborator (*dashed line*). The Gini coefficient measures the degree of inequality of distributions, and ranges from zero (=equality) to one (=monopoly). Both lines show an increase over time; indicating that publication and collaboration opportunities are not evenly distributed among scholars

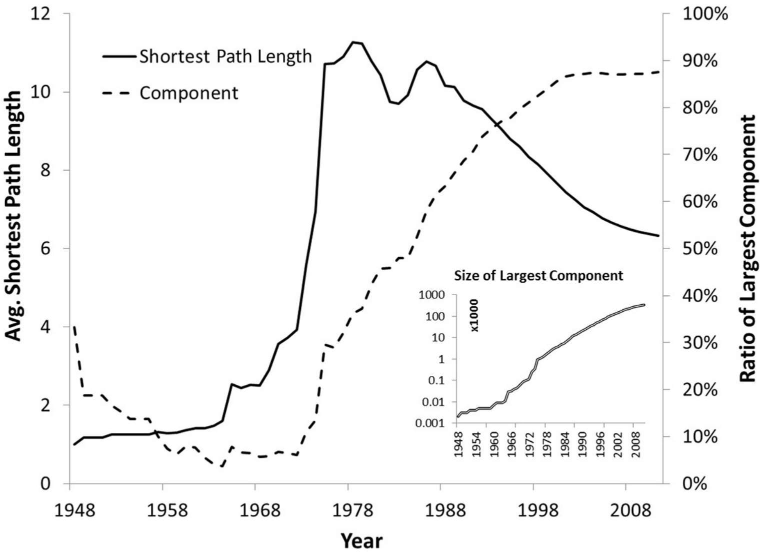


**Fig. 8** Average length of the shortest path (*left y*-axis; *solid line*), proportional ratio of the largest component (*right* y-axis; *dashed line*), and the size of largest component (*inset* figure; *y*-axis in log scale). The average shortest path length increased over time; reaching its peak at 11.27 in 1978 but decreased afterwards until 2011, where authors were separated by on average 5.33 intermediaries. The proportion of the largest group of authors who could reach each other increased from 3.7 % in 1964 to 87.50 % in 2011

1 to 15 % of authors (in terms of node degree), removed each set (1, 2, 3...15) from the network, and re-calculated the ratio of the largest component. The results from this procedure are shown in Fig. 9. Starting from 87.50 %, the largest component decreases to

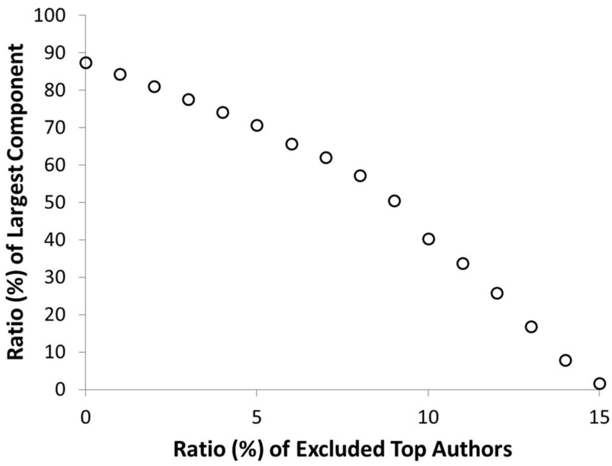


**Fig. 9** Effect of discarding highly collaborative scholars on network connectivity. The collaboration network is almost completely fragmented if the top 15 % of the most connected authors are excluded from the network

50 % when approximately 9 % of the most collaborative scholars (32,600) are disregarded, and to less than 2 % after 15 % get eliminated. This indicates that the connectivity of Korea's domestic collaboration network relies on a relatively small fraction of highly connected individuals. This finding can be compared to other studies. In computer science, the removal of the top 15 % of scholars scattered the connectivity of the network to almost zero (Franceschet 2011), while in physics, 20–30 % of scholars need to be disregarded to dissolve the network connectivity (Newman 2010).

The ratio of the largest component, i.e., 87.50 % in 2011, means that the majority of Korean authors can reach each other within a certain steps of coauthorship ties, with a maximum of 28 steps (i.e., diameter) among reachable pairs. The connection between scholars through coauthor ties can be measured in terms of the length of the shortest paths. The implications of this metric are rather graph-theoretical: a scholar from history, for example, might not aim to co-publish with someone from physics but knowing how long the distances are helps to understand how integrated a community is. In Fig. 8, the temporal change in the average length of the shortest path is plotted (solid line). It started at 2, reached its peak of 11.27 in 1978, and then kept decreasing until it reached its lowest value of 6.33 in 2011. This means that, as of 2011, the majority of pairs of Korean scholars can be reached within roughly six steps of a coauthoring relationship. This distance is longer than the ones for Slovenian and Turkish scholars, and for authors in biomedicine, physics, but shorter than in computer science, mathematics, and sociology.

We also calculated the clustering coefficient (also known as transitivity). The clustering coefficient expresses to what degree an author's collaborators publish with each other (Newman 2001). Figure 10 shows that the clustering coefficient in the KISTI data hit its peak of 0.47 in 1969, and then decreased to 0.19 in 2011. This value is similar to the ones for Slovenia and sociology, but it is higher than the ones for biomedicine and mathematics, but lower than the ones for Turkey, computer science, and physics.

The five degrees of separation among Korean scholars is an interesting finding because it aligns with the number of degrees of separation that is descriptive of small-world networks (Watts and Strogatz 1998). We tested whether the Korean collaboration network is a

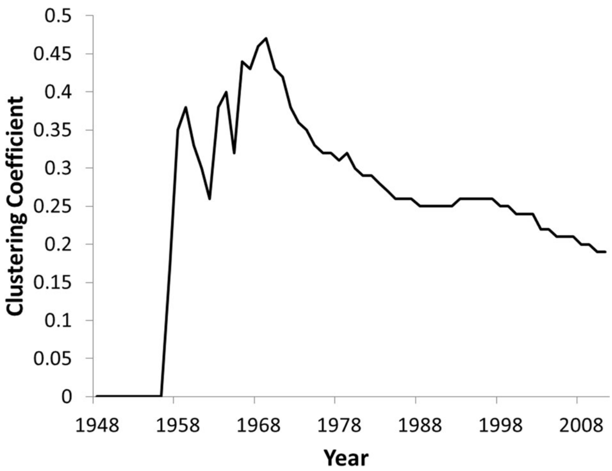


**Fig. 10** Temporal change of clustering coefficient. If authors A and B have worked with the same author C, C might introduce A to B (or B to A). The clustering coefficient measures how many of these potential transitive collaborations occur in the data. The plot shows that for Korean scholars, transitive collaboration kept decreasing after 1969

small-world by comparing the descriptive metrics of this type of network, namely the graph-level Watts–Strogatz clustering coefficient[3] and the shortest path length, to a synthetically generated random graph (Perc 2010; Watts and Strogatz 1998).[4] More specifically, we generated an undirected Erdős–Rényi random graph with the same node size (378,296) and average degree (7.38) as in the KISTI data, and calculated these network metrics. The resulting Watts–Strogatz clustering coefficient is 0.00002 (vs. 0.73 in KISTI), and the average length of the shortest path is 6.65 (vs. 6.33 in KISTI). This means that the Korean collaboration network has small-world properties as also confirmed for other previous national and field-level collaboration studies (Barabási et al. 2002; Çavusoglu and Türker 2013; Franceschet 2011; Perc 2010).

## Conclusion and discussion

In this paper, the domestic collaboration network of Korea for the period of 1948–2011 was described and compared to other national and field-level coauthorship networks. The studied community features an exponential growth in the number of publications and authors over the last 60 years, which may be due to economic development and improvements in higher education. Scholars in this system have become more productive and collaborative every year. In addition, their interconnectivity increased; leading to a giant component within the wider network. Furthermore, scholars can reach each other

[3] A graph-level Watts–Strogatz clustering coefficient is the average of the ego network densities of all authors.

[4] The coauthorship networks in our study were constructed by converting an author-by-paper matrix (i.e., two-mode or bipartite network) into an author-by-author matrix (i.e., one-mode or monopartite network). Several scholars have proposed to compare bipartite graphs (e.g. networks of interlocking directors/companies or inventors/patents) against bipartite random graphs. For more details, please refer to Robins and Alexander (2004) or Kogut and Belinky (2008).

within an average of six steps of a coauthoring relationship; confirming the small-world property for this community.[5]

Our findings imply that the given network structure enables efficient collaboration and knowledge sharing among its participants, which is a desirable feature of scholarly communication (Newman 2001; Perc 2010). We assume that the small-world configuration of this community is associated with the innovativeness and strong performance of the nation's eonomy and academic sector over the last decades. This observation substantiates prior findings on a relationship between a small-world network topology and creativity, innovation and effective knowledge diffusion (Chen and Guan 2010; Cowan and Jonard 2004; Uzzi and Spiro 2005).

These increased productivity and collaboration, however, are also associated with a concentration of productivity and collaborations applicable to a small number of scholars who consequently became increasingly influential throughout this process. This characteristic may represent a fragile network structure. In other words, the connectivity can break if a small fraction of highly collaborated scholars left the network. How will the inequality in productivity and collaboration as well as the potentially vulnerable network structure affect the future of domestic collaboration in Korea? This question requires further in-depth investigations and sociological insights in order to develop meaningful policy implications.[6]

Another valuable insight can be obtained from cross-national comparison (see Table 1). How different are countries in terms of scientific collaboration, and why are they different? The average Korean domestic scholar produces more papers (avg. 4.73) than scholars in Turkey (avg. 1.56). However, the Korean authors participate in a smaller-sized collaborations (avg. author size per paper 2.79) than Turkish scholars (4.08) and therefore might have a smaller number of coauthors (7.38 vs. Slovenia: 10.7 and Turkey: 35.03). This indicates that different norms and traditions may apply to each country. For example, in comparison to Korea, a large proportion of scholars in Turkey seems to produce knowledge in larger teams, e.g., laboratory-based collaborations that are common in medicine or the physical science.

With respect to the collaboration structure, average Korean scholars are connected via longer shortest paths (6.33 vs. Slovenia: 4.6 and Turkey: 4.14), and tend to collaborate less with others who share common coauthors (clustering coefficient 0.19 vs. Slovenia: 0.20 and Turkey: 0.75). Such structural differences may be due to different sizes of datasets and the different levels of interdisciplinarity in each country. For example, if collaboration among scholars from diverse fields increased in a country, scholars would be expected to be reachable through shorter paths than before. Of course, such inferences should be based on findings from a comparison of scholarly collaboration across different disciplines and possible explanations for any observed differences (e.g., Lužar et al. 2014 for Slovenia). As such, this study can be further developed by investigating interdisciplinarity of collaboration among Korean scholars. Findings from such an in-depth study would provide insights into the factors driving the evolution of productivity and connectivity of Korean scholars at a field level.

As we confined our interest to an analysis of domestic collaboration, our findings inevitably fail to capture the collaboration among Korean scholars in international settings.

[5] This interpretation needs to be taken with caution. Even though a path is relatively short, most potential within and across field collaborations will not be realized.

[6] For example, one topic for further investigations might be to analyze whether the increased productivity and connectivity are due to new cohorts of scientists who truly are more productive than more established scholars, or because scholars who have been in a field for many years have accumulated a multitude of publications.

For example, if two Korean scholars have only published together in an international journal but not in a domestic journal, their link is not considered in our study. Therefore, another open question is how the domestic scholarly relationships are associated with education abroad, international collaboration, and the international publication activity of Korean scholars.

**Acknowledgments** This work is supported by KISTI (Korea Institute of Science and Technology Information), grant P14033. The American Physical Society (APS, http://journals.aps.org/datasets) kindly provided the publication records of the Physical Review journals for our research. We would like to thank Brian Karrer (Facebook), Travis Martin (University of Michigan), Brian Ball (Dotomi Inc.) and Mark E. J. Newman (University of Michigan) for helping us to disambiguate author names in the APS dataset. We also thank the anonymous reviewers who helped us improve the quality of this paper, and Susan Lafferty (GSLIS, University of Illinois at Urbana-Champaign) for editing the manuscript.

## References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). The collaboration behaviors of scientists in Italy: A field level analysis. *Journal of Informetrics, 7*(2), 442–454. doi:10.1016/j.joi.2013.01.009.

Abt, H. A. (2007). The future of single-authored papers. *Scientometrics, 73*(3), 353–358. doi:10.1007/s11192-007-1822-9.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A—Statistical Mechanics and Its Applications, 311*(3–4), 590–614. doi:10.1016/s0378-4371(02)00736-7.

Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences of the United States of America, 101*(-suppl. 1), 5266–5273. doi:10.1073/pnas.0307625100.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177.

Çavuşoğlu, A., & Türker, İ. (2013). Scientific collaboration network of Turkey. *Chaos, Solitons and Fractals, 57*, 9–18.

Chen, Z. F., & Guan, J. C. (2010). The impact of small world on innovation: An empirical study of 16 countries. *Journal of Informetrics, 4*(1), 97–106. doi:10.1016/j.joi.2009.09.003.

Choi, S.-H., Kim, B.-K., Kang, M., You, B.-J., Lee, J., & Park, J.-W. (2011). A study of citing patterns of Korean scientists on Korean journals. *Journal of the Korean Society for Information Management, 28*(2), 97–115.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review, 51*(4), 661–703.

Cowan, R., & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control, 28*(8), 1557–1575. doi:10.1016/j.jedc.2003.04.002.

Fegley, B. D., & Torvik, V. I. (2013). Has large-scale named-entity network analysis been resting on a flawed assumption? *PLoS ONE, 8*(7), 1–16. doi:10.1371/journal.pone.0070299.

Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology, 62*(10), 1992–2012. doi:10.1002/asi.21614.

Glasser, G. J. (1962). Variance formulas for the mean difference and coefficient of concentration. *Journal of the American Statistical Association, 57*(299), 648–654. doi:10.2307/2282402.

Gossart, C., & Özman, M. (2009). Co-authorship networks in social sciences: The case of Turkey. *Scientometrics, 78*(2), 323–345. doi:10.1007/s11192-007-1963-x.

Grossman, J. W. (2002). Patterns of collaboration in mathematical research. *SIAM News, 35*(9), 8–9.

Han, S. H., Cho, S. R., Yang, J. M., & Ryu, D. A. (2009). *A study on academic research and development activities in Korea*. Seoul: National Research Foundation of Korea.

Kim, M.-J. (2005). Korean science and international collaboration, 1995–2000. *Scientometrics, 63*(2), 321–339.

Kim, J., & Diesner, J. (2015). The effect of data pre-processing on understanding the evolution of collaboration networks. *Journal of Informetrics, 9*(1), 226–236. doi:10.1016/j.joi.2015.01.002.

Kim, J., Kim, H., & Diesner, J. (2014). The impact of name ambiguity on properties of coauthorship networks. *Journal of Information Science Theory and Practice, 2*(2), 6–15. doi:10.1633/JISTaP.2014.2.2.1.

Kogut, B., & Belinky, M. (2008). Comparing small world statistics over time and across countries: An introduction to the special issue comparative and transnational corporate networks. *European Management Review, 5*(1), 1–10. doi:10.1057/emr.2008.6.

Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology, 58*(7), 1019–1031. doi:10.1002/asi.20591.

Lužar, B., Levnajić, Z., Povh, J., & Perc, M. (2014). Community structure and the evolution of interdisciplinarity in Slovenia's scientific collaboration network. *PLoS ONE, 9*(4), e94429. doi:10.1371/journal.pone.0094429.

Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation patterns in the physical review. *Physical Review E, 88*(1), 012814-1–012814-9. doi:10.1103/PhysRevE.88.012814.

Merton, R. K. (1968). Matthew effect in science. *Science, 159*(3810), 56. doi:10.1126/science.159.3810.56.

Ministry of Education. (2013). Statistical yearbook of education. Seoul, Korea: Ministry of Education.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213–238.

Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics, 101*(3–4), 819–841. doi:10.1023/A:1026485807148.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404–409. doi:10.1073/pnas.021544898.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America, 101*(suppl. 1), 5200–5205. doi:10.1073/pnas.0307545100.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford: Oxford University Press.

Park, H. W., & Leydesdorff, L. (2008). Korean journals in the Science Citation Index: What do they reveal about the intellectual structure of S&T in Korea? *Scientometrics, 75*(3), 439–462.

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics, 4*(4), 475–482.

Robins, G., & Alexander, M. (2004). Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Computational and Mathematical Organization Theory, 10*(1), 69–94. doi:10.1023/B:CMOT.0000032580.12184.c0.

Ryoo, J. W. (2011). The labor market for college professors in Korea. *Korean Journal of Labor Economics, 34*(2), 1–27.

Schubert, A., & Glanzel, W. (2006). Cross-national preference in co-authorship, references and citations. *Scientometrics, 69*(2), 409–428. doi:10.1007/s11192-006-0160-7.

Strotmann, A., & Zhao, D. (2012). Author name disambiguation: What difference does it make in author-based citation analysis? *Journal of the American Society for Information Science and Technology, 63*(9), 1820–1833. doi:10.1002/Asi.22695.

Torvik, V. I., & Smalheiser, N. R. (2009). Author name disambiguation in MEDLINE. *Acm Transactions on Knowledge Discovery from Data, 3*(3), 1–29. doi:10.1145/1552303.1552304.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world Problem1. *American Journal of Sociology, 111*(2), 447–504.

Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics, 6*(4), 700–711. doi:10.1016/j.joi.2012.07.008.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Yoshikane, F., & Kageura, K. (2004). Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics, 60*(3), 435–446. doi:10.1023/b:scie.0000034385.05897.46.