Scientometrics (2018) 116:1853–1865
https://doi.org/10.1007/s11192-018-2835-2

# Why do ecologists search for co-authorships? Patterns of co-authorship networks in ecology (1977–2016)

**Anderson Matos Medina**[1]

Received: 30 January 2018 / Published online: 9 July 2018


**Abstract**

Here, the pattern of co-authorship among ecologists was evaluated using a network approach that was built using four time periods (1977–1986, 1987–1996, 1997–2006, and 2007–2016). Furthermore, four potential explanations (geographic distance, word similarity, reputation asymmetry, and country development) for this pattern were evaluated. Distance and reputation asymmetry effects on collaboration have decreased in recent decades, whereas word similarity was a good predictor in recent decades. Of interest, country development was not a good predictor of co-authorship among ecologists.

**Keywords** Co-authorship networks · Research collaboration · Ecology · Homophily

## Introduction

Modern science is rooted on the shift from the exchange of letters between scientists to, what we now call, articles (Day 1989). This shift has made articles the building blocks of scientific knowledge; therefore, to understand how scientific knowledge is built, it is necessary to understand how scientists interact with their peers when writing or publishing their articles. Scientometric approaches help us understand the mechanisms behind scientific practices, including how social aspects affect peer-review (e.g., sexual bias) or how the characteristics of articles influence citation rates (Jamali and Nikzad 2011; Tregenza 2002).

A recent trend in science is the increasing number of publications that contain multiple authors (Wuchty et al. 2007), leading to the question of how knowledge flows in the scientific community. Three major approaches tackle this question (Neff and Corley 2009): Co-word analysis, co-citation analysis, and co-author analysis (Callon et al. 1983; de Solla Price 1965; Newman 2004). Co-citation analysis involves searching for groups of references that are commonly cited together, co-word analysis seeks similarities in the

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11192-018-2835-2) contains supplementary material, which is available to authorized users.

✉ Anderson Matos Medina
bojaum@gmail.com

[1] Programa de Pós-Graduação em Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil

frequency of words used in articles to define themes (cluster of words), and co-author analysis creates a network of interactions between authors and their co-authors.

In particular, I focused on analysing the co-authorship networks in the scientific field of ecology to detect the pattern of interactions among authors. Furthermore, I established four hypotheses to explain the pattern, and possibly temporal changes in these patterns, found on co-authorship networks and, therefore, how authors collaborate with their peers. The first hypothesis was based on the rationale that coauthors seek partners with similar lines of thought, belonging to the same field of research. A direct consequence of this, it can be assumed that co-authorship is more frequent between authors with higher word similarity. Words are a crucial step to determine who reads an article and accesses knowledge. For instance, a recent study showed an increase in jargon terms in scientific texts (Plaven-Sigray et al. 2017), which are associated to specialization of subfields. Therefore, authors that use different words are less likely to form co-authorships. The second hypothesis was that the frequency of author interactions is facilitated when authors work within the same geographical area, or is more intense between authors from closer geographical locations (Pan et al. 2012). If this hypothesis is true, I expected to find an increase in the frequency of interactions at shorter geographical distances. The third hypothesis was that authors aim to collaborate with peers that have more experience than them (Merton 1968), leading to preferential attachment (Barabási and Albert 1999) in which authors with more prestige attract authors with less prestige. If this hypothesis is true, I expected co-authorship to be more frequent between authors that have a greater asymmetry in publications. The fourth hypothesis was based on the centre-periphery hypothesis, whereby researchers from well developed countries will attract researchers from less developed countries. Two explanations for this attraction are the diffusion of knowledge produced by more developed countries towards less developed countries and the higher investment in science by the former (Wagner and Leydesdorff 2005). Thus, co-authorships will be more frequent between researchers from countries with higher asymmetry in developmental status.

## Methods

### Inclusion criteria

I selected the ten most influential ecology journals based on their impact factor in journal citation reports for 2015 (Thomson Reuters 2016; Table 1). Afterwards, I retrieved information from the articles of these journals on the SCOPUS database, searching for the journal names from 1943 to 2016. This search resulted in 21,440 articles that, after removing duplicates, erratum, and editorials, resulted in 20,392 unique articles.

I separated the articles into four time periods (1977–1986, 1987–1996, 1997–2006, and 2007–2016) to analyse temporal changes in co-authorship networks. For each time period, I extracted the list of authors ($n_{1977–1986} = 523$; $n_{1987–1996} = 3128$; $n_{1997–2006} = 10,308$; $n_{2007–2016} = 31,827$) and, if an author had only one publication in the time period, they were excluded from further analysis (remaining authors: $n_{1977–1986} = 112$; $n_{1987–1996} = 818$; $n_{1997–2006} = 3207$; $n_{2007–2016} = 10,224$); however, the publication was counted for other authors if they published more than one article. I abstained from synonymising authors with the names of multiple authors to avoid creating connections between researchers that are, in reality, disconnected. This synonymy error had more negative effects on the network pattern than leaving the authors with multiple names because it can creates connections between groups of unconnected authors (Borrett et al.

**Table 1** Top ten journals of the ecology area and their respective impact factor on the journal citation report 2016

| Rank | Journal | IF15 |
|---|---|---|
| 1 | Trends in Ecology and Evolution | 16.7 |
| 2 | Ecology Letters | 10.7 |
| 3 | Annual Review of Ecology Evolution and Systematics | 9.4 |
| 4 | ISME Journal | 9.3 |
| 5 | Frontiers in Ecology and the Environment | 8.5 |
| 6 | Global Change Biology | 8.4 |
| 7 | Ecological Monographs | 8.0 |
| 8 | Methods in Ecology and Evolution | 6.3 |
| 9 | Journal of Ecology | 6.2 |
| 10 | Bulletin of the American Museum of Natural History | 6.0 |

2014). For example, if *Virgulino Ferreira da Silva* had two author names (*da Silva*, *VF* and *Silva*, *VF*) they were left as two different authors. These two author names will have a minor impact on the network because they probably interact with a similar group of authors associated with the real author (*Virgulino Ferreira da Silva*). However, one setback is that they increase the number of authors in the network. When checking this potential issue, I found that the authors had possible multiple names in the dataset in less than 20% of cases for all time periods ($n_{1977–1986} = 3.95\%$; $n_{1987–1996} = 11.05\%$; $n_{1997–2006} = 15.42\%$; $n_{2007–2016} = 17.27\%$).

## Co-authorship network

I created a weighted co-authorship matrix in which both lines and columns were authors and an author was considered to have interacted with another author if they published at least one paper together. Based on this weighted co-authorship matrix, I created four undirected graph networks, one for each time period, using the package *igraph* (Csardi and Nepusz 2006). An undirected graph is a representation of the co-authorship matrix that displays the authors as vertices and the number of shared publications as edges, with the edges having no direction. When plotting these graphs, I used the Fruchterman–Reingold algorithm, which generates aesthetic graphs with vertices that are evenly distributed, while maintaining edges of similar lengths (Fruchterman and Reingold 1991).

Research communities are defined here as sets of authors that interact more among themselves than with other authors from the co-authorship networks. These research communities might be caused by geographic distance, similar research interests, or student/tutor relationships. I used the fast greedy modularity optimisation algorithm to detect the communities of researchers based on the co-authorship network structure, because it is a fast algorithm for large and sparse networks (Clauset et al. 2004). This algorithm starts with each vertex being an exclusive community, and searches for a combination of vertices that cause the higher increase on modularity until finishing the number of vertices available (Clauset et al. 2004). The modularity is a measure of how a subset of the graph (set of vertices) is more linked among the subset through edges than is linked with other subsets of the network.

#### Predictors of interactions among authors

The similarity of lines of thought between authors was measured using the similarity of words in the articles titles of author A is comparison with those of author B. This approach is based on the assumption that the words used in a text convey ideas; therefore, a set of words that co-occurs in a text should represent a line of thought. The same assumption underlies the use of co-word analysis (Callon et al. 1983). To measure word similarity, a list containing the titles of all publications was compiled and punctuations, numbers, and stop words were removed from the title using package *tm* (Feinerer and Hornik 2017). Furthermore, the 200 most frequent words were synonymised by removing plural suffixes manually (e.g.: The term *interactions* was synonymised to *interaction*). After cleaning the text for each time period, Sørensen similarity was calculated between the words used in all articles of author A (including articles not published with co-authors) and author B (including articles not published with co-authors).

Before measuring the geographic distance between co-authors, I extracted a list of affiliations from authors included in the networks for each time period, and removed any duplicates present (number of affiliation sites: $n_{1977–1986} = 168$; $n_{1987–1996} = 1491$; $n_{1997–2006} = 6296$; $n_{2007–2016} = 22{,}484$). I then searched for the geographic coordinate using the package *ggmap* (Kahle and Wickham 2013). If the search failed to retrieve a geographic coordinate, the country was used as a surrogate in the search (number of cases: $n_{1977–1986} = 3$; $n_{1987–1996} = 8$; $n_{1997–2006} = 24$; $n_{2007–2016} = 165$). Afterwards, I calculated the mean geographic coordinates for authors with multiple affiliations during the time period. For example, if author J was affiliated in Sweden (17′64°E 59′85°N) in one article and Canada in another article (76′48°W 44′22°N), their geographic coordinate was placed as 29′42°W and 52′04°N. Note that the averaging process could lead to possible bias to distance measures if authors with multiple affiliations also increase with time. In some cases, the author did not have an affiliation available or the search was not able to retrieve a geographic coordinate; such affiliations were included in subsequent analyses as missing data ($n_{1977–1986} = 14$; $n_{1987–1996} = 50$; $n_{1997–2006} = 1$; $n_{2007–2016} = 20$). Then the distance in kilometres between the two geographic coordinates of the author was measured using the *raster* package (Hijmans 2016).

I measured the capacity of an author to attract other researchers based on their reputation using citation asymmetry as a surrogate. Citation asymmetry is the absolute difference between the mean numbers of citations from articles published in a single time period of author C (including articles not published with co-authors) and the mean number of citations from articles published in a single time period of author D (also including articles not published with co-authors). This statistic measures differences in the reputation of authors C and D. For instance, a higher citation asymmetry indicates a higher capacity of attracting other co-authors.

I measured the developmental status of countries using the human development index, which encompasses information on healthy lifestyles, access to knowledge, and standards of living (United Nations Development Programme 2016). I measured the absolute difference in human development between Author A and Author B (hereafter HDI asymmetry) as a proxy for differences between the developmental status of the country of a given author. If an author was affiliated with more than one country in a time period, the most frequent country was used to obtain the human development index and measure HDI asymmetry. Note that some authors did not have information on country affiliation or were from a country absent from the human development index ($n_{1977–1986} = 8$; $n_{1987–1996} = 14$; $n_{1997–2006} = 11$; $n_{2007–2016} = 94$).

## Statistical analysis

Data residuals presented a high degree of heterogeneity in both linear models and generalized linear models with Poisson distribution, and the large number of observations prevented the use of generalized least squares. Thus, I used Spearman rank correlation for each explanatory variable (words similarity, citation asymmetry, geographic distance, and HDI asymmetry) for the whole dataset (overall pattern, number of correlations = 4) and for each time period (temporal trends, number of correlations = 16). I also obtained Spearman rank correlations between the explanatory variables to test whether there were confounding effects generated by correlating them. However, only two explanatory variables were highly correlated; thus, this effect was not considered (Table 2). All analyses were performed in R environment (R Development Core Team 2017).

**Table 2** Spearman rank correlation between explanatory variables used to predict coauthorship frequency

| Time period | Predictor | | | | |
|---|---|---|---|---|---|
| | Predictor | Word | Citation | Distance | HDI |
| All | Word | | | | |
| | Citation | 0.1910 | | | |
| | Distance | 0.1042 | 0.0044 | | |
| | HDI | − 0.0348 | − 0.0437 | 0.4090 | |
| 1977–1986 | Word | | | | |
| | Citation | − 0.5421 | | | |
| | Distance | − 0.5570 | 0.2271 | | |
| | HDI | − 0.2898 | 0.0429 | 0.8389 | |
| 1987–1996 | Word | | | | |
| | Citation | 0.3862 | | | |
| | Distance | 0.1774 | 0.0319 | | |
| | HDI | 0.1550 | 0.0502 | 0.4563 | |
| 1997–2006 | Word | | | | |
| | Citation | 0.2841 | | | |
| | Distance | − 0.0747 | − 0.0227 | | |
| | HDI | − 0.2307 | − 0.0560 | 0.4766 | |
| 2007–2016 | Word | | | | |
| | Citation | 0.2120 | | | |
| | Distance | 0.1450 | 0.0317 | | |
| | HDI | 0.0047 | − 0.0312 | 0.3943 | |

Word corresponds to words similarity. citation corresponds to citation asymmetry. distance corresponds to geographic distance and HDI corresponds to HDI asymmetry

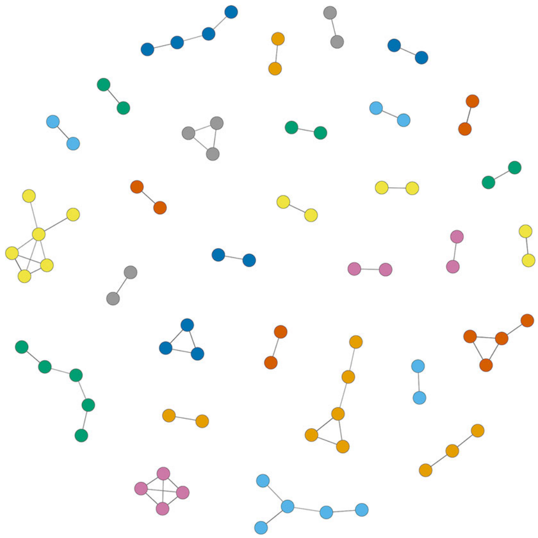

**Fig. 1** Coauthorship network based on articles from the top ten journals of ecology published during the period of 1977–1986. Each circle is an author and different colors represent different research communities; lines linking two authors denote a coauthorship and disconnected authors are omitted. (Color figure online)

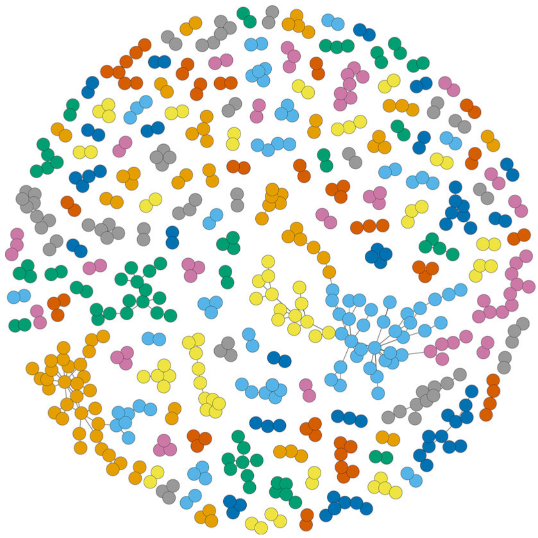

**Fig. 2** Coauthorship network based on articles from the top ten journals of ecology published during the period of 1987–1996. Each circle is an author and different colors represent different research communities; lines linking two authors denote a coauthorship and disconnected authors are omitted. (Color figure online)

## Results

A pattern found on the co-authorship networks (Figs. 1, 2, 3, 4) was an increasing number of authors over time ($n_{1977–1986} = 112$; $n_{1987–1996} = 818$; $n_{1997–2006} = 3207$; $n_{2007–2016} = 10{,}224$; $\mathrm{Mean}_{1977–1986} = 1.85$; $\mathrm{Mean}_{1987–1996} = 1.86$; $\mathrm{Mean}_{1997–2006} = 3.09$; $\mathrm{Mean}_{2007–2016} = 4.98$). There was also a decreased in the mean frequency of co-authorship over time ($n_{1977–1986} = 2.03$; $n_{1987–1996} = 1.59$; $n_{1997–2006} = 1.46$; $n_{2007–2016} = 1.30$). In addition, there was an increase in the number of research communities over time ($n_{1977–1986} = 60$; $n_{1987–1996} = 409$; $n_{1997–2006} = 419$; $n_{2007–2016} = 279$) and mean number of authors in the research communities over time ($n_{1977–1986} = 1.87$; $n_{1987–1996} = 2.00$;

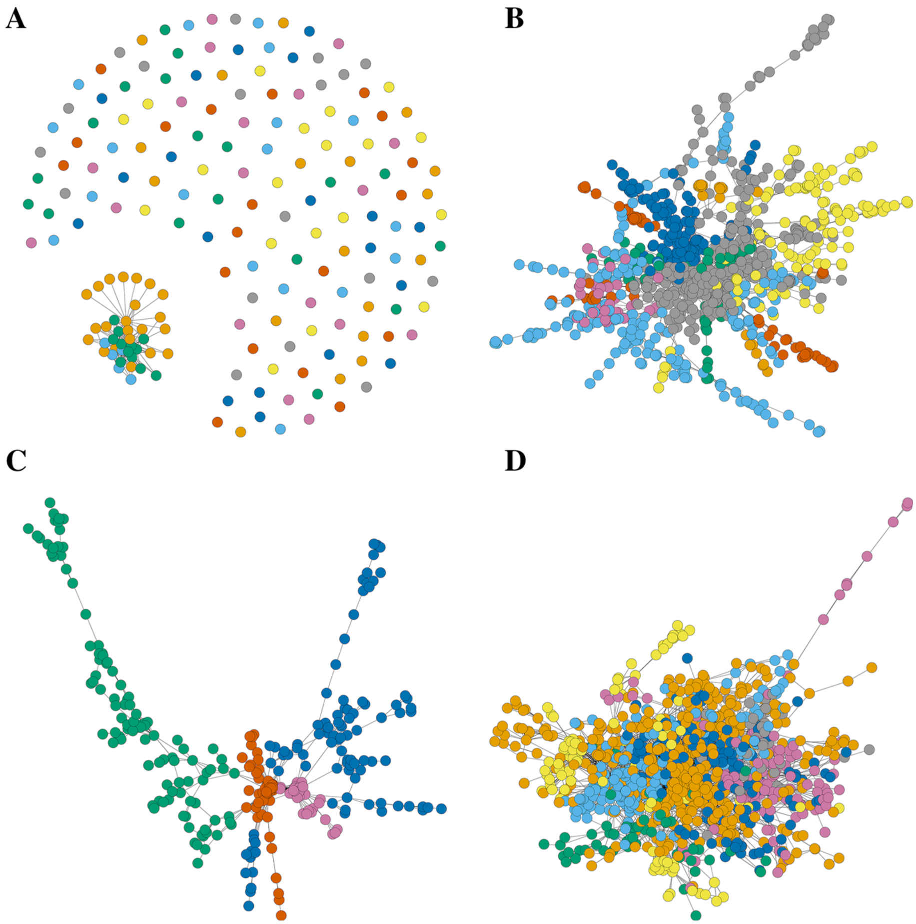


**Fig. 3** Coauthorship network based on articles from the top ten journals of ecology published during the period of 1997–2006. Given the high number of authors, I built a network of researchers communities which have each vertex represents a group of co-authors (**a**), and the researchers communities with more authors is presented in more detail (**b**, **c**, **d**). **b**, **c** and **d**: Each circle is an author and different colors represent different research communities; lines linking two authors denote a coauthorship and disconnected authors are omitted. (Color figure online)

$n_{1997–2006}$ = 7.65; $n_{2007–2016}$ = 36.65). Finally, there was a decrease in modularity over time ($n_{1977–1986}$ = 0.94; $n_{1987–1996}$ = 0.97; $n_{1997–2006}$ = 0.83; $n_{2007–2016}$ = 0.70).

Overall, an increase in word similarity was correlated with an increase in the frequency of co-authorship between authors (rho = 0.4448, $p < 0.05$; Fig. 5). However, this effect was negative during the first decade ($\text{rho}_{1977–1986}$ = − 0.4448, $p < 0.01$), becoming positive in the subsequent decades ($\text{rho}_{1987–1996}$ = 0.5956, $p < 0.01$; $\text{rho}_{1997–2006}$ = 0.5913, $p < 0.01$; $\text{rho}_{2007–2016}$ = 0.4154, $p < 0.05$).

Distance played a small negative effect on the frequency of co-authorship (rho = − 0.1728, $p < 0.01$; Fig. 6); however, this effect was strongest in the early decades, growing weaker in the latter decades ($\text{rho}_{1977–1986}$ = −0.5014, $p < 0.01$; $\text{rho}_{1987–1996}$ = − 0.3358, $p < 0.01$; $\text{rho}_{1997–2006}$ = − 0.0749, $p < 0.01$; $\text{rho}_{2007–2016}$ = − 0.1785, $p < 0.01$; Fig. 6).

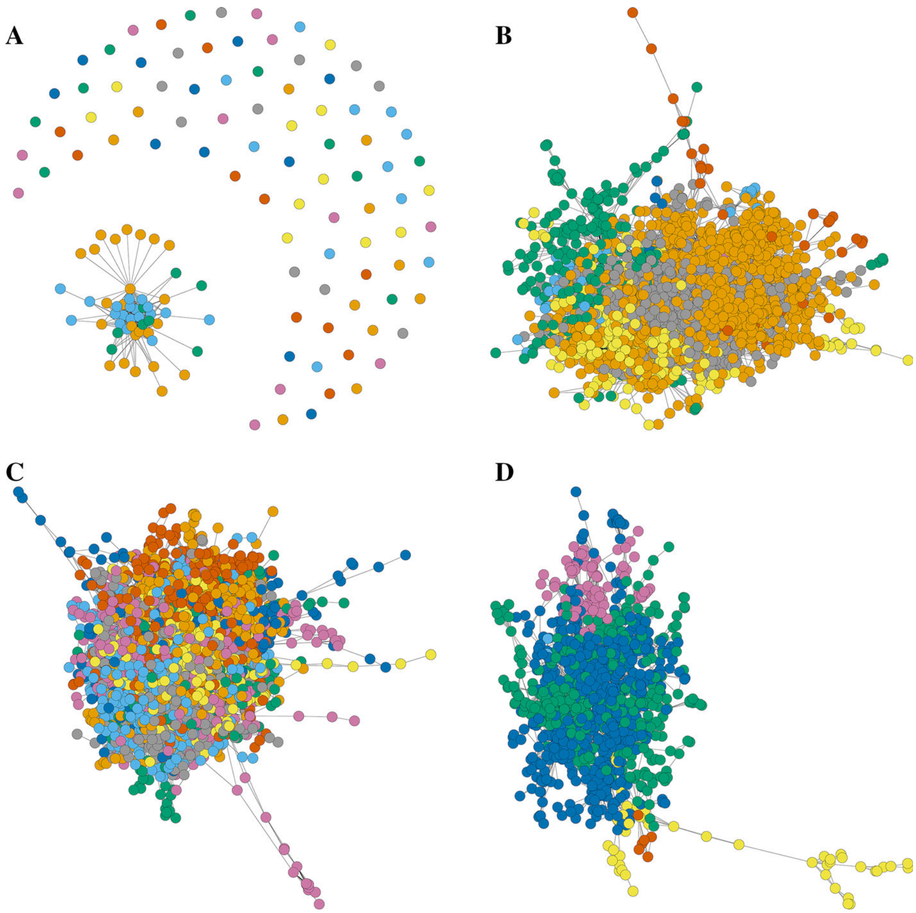


**Fig. 4** Coauthorship network based on articles from the top ten journals of ecology published during the period of 2007–2016. Given the high number of authors, I built a network of researchers communities which have each vertex represents a group of co-authors (**a**), and the researchers communities with more authors is presented in more detail (**b**, **c**, **d**). **b**, **c** and **d**: Each circle is an author and different colors represent different research communities; lines linking two authors denote a coauthorship and disconnected authors are omitted. (Color figure online)

Contrary to my expectation, citation asymmetry was negatively correlated with the frequency of co-authorship (rho = $-$ 0.2170, $p < 0.05$; Fig. 7), demonstrating that authors seek coauthors with similar reputations. This trend was stronger during the early decades, and became weaker during the latter decades ($rho_{1977-1986} = -0.4552$, $p < 0.01$; $rho_{1987-1996} = -0.4771$, $p < 0.01$; $rho_{1997-2006} = -0.3421$, $p < 0.01$; $rho_{2007-2016} = -0.2024$, $p < 0.05$; Fig. 7).

The overall effect of HDI asymmetry on authorship frequency was negative (rho = $-$ 0.1306, $p < 0.05$; Fig. 8). Furthermore, this relationship between HDI and authorship oscillated from small ($rho_{1987-1996} = -0.2272$, $p < 0.01$; $rho_{2007-2016} = -0.1413$, $p < 0.05$) to no effect ($rho_{1977-1986} = -0.1310$, $p = 0.336$; $rho_{1997-2006} = 0.0021$, $p = 0.823$).

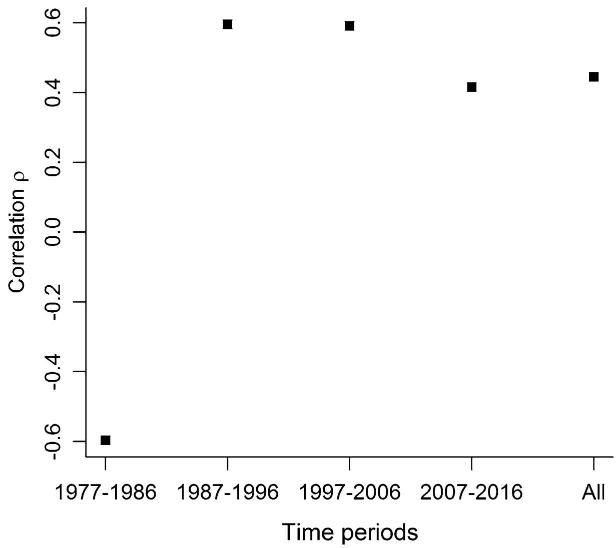


**Fig. 5** Values of Spearman correlation coefficient coauthorship frequency and word similarity on each time period and full dataset. Black squares are correlations with $p < 0.05$ while gray squares are correlations with $p > 0.05$

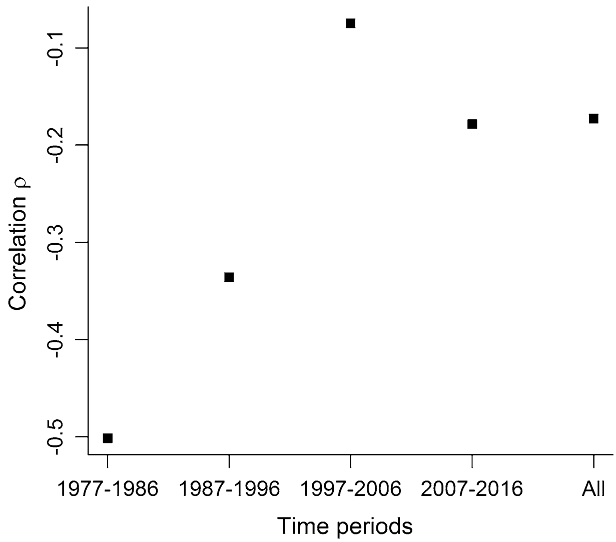


**Fig. 6** Values of Spearman correlation coefficient between coauthorship frequency and geographic distance on each time period and full dataset. Black squares are correlations with $p < 0.05$ while gray squares are correlations with $p > 0.05$

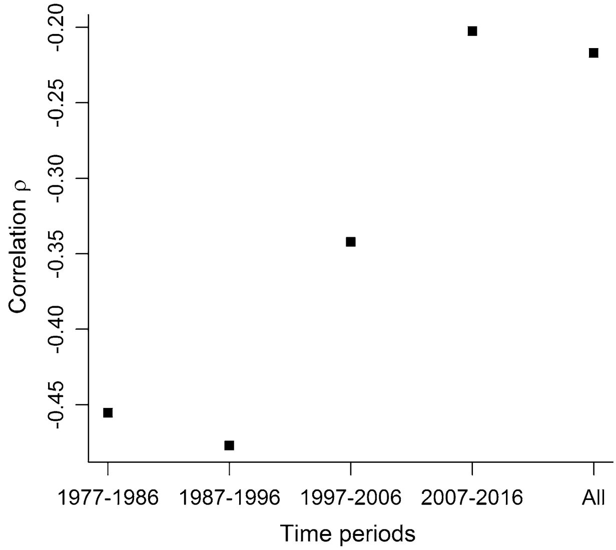


**Fig. 7** Values of Spearman correlation coefficient between coauthorship frequency and citation asymmetry on each time period and full dataset. Black squares are correlations with $p < 0.05$ while gray squares are correlations with $p > 0.05$

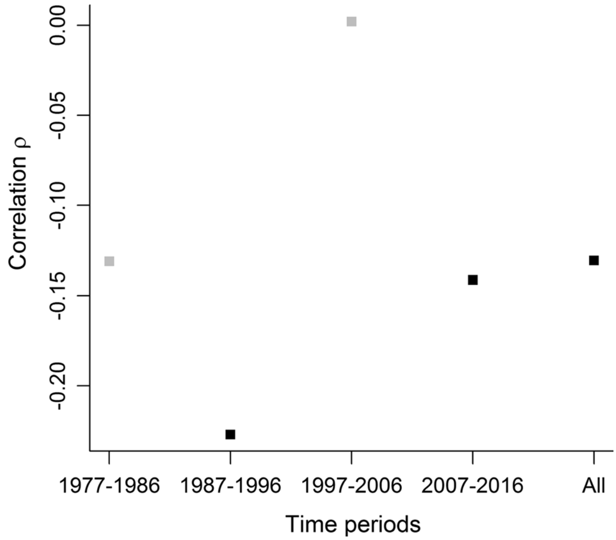


**Fig. 8** Squares are Spearman correlation coefficient between coauthorship frequency and HDI asymmetry on each time period and full dataset. Black squares are correlations with $p < 0.05$ while gray squares are correlations with $p > 0.05$

## Discussion

An increase in the number of authors in networks is a direct consequence of the increased number of publications. For example, during the early decades, there were only three ecological journals meeting my criteria, whereas there were 20 journals in the last decade. Furthermore, a general trend in science and ecology in particular, is the increased number of authors per article (Fig. S1; Logan 2016; Wuchty et al. 2007). This trend of science produced by larger teams also helps to explain why there was a decrease in the mean frequency of interactions. For instance, the increased number of new authors might increase the chance of occasional publications between authors. Furthermore, the number of research communities increased until the third decade (1997–2006) and decreased in the last decade (2007–2016). A possible explanation for this pattern is that new research centres were established in developing countries during the first three decades (1977–1986, 1987–1996, and 1997–2006). This suggestion is supported by a study showing that countries in Latin America increased their scientific output during 1990–2000 (Holmgren and Schnitzer 2004). It is possible that the number of research communities declined during 2007–2016 because these communities were isolated from developed countries during the early decades and became more connected during the last decade.

My hypothesis that authors sought partners with similar interests was corroborated in the global dataset, and over the latter last three decades (1987–1996, 1997–2006, and 2007–2016). Thus, authors are indeed carrying out science with similar interests. Interestingly, the opposite was true for the first decade (1977–1986), revealing that authors were performing science with partners from other research schools. This phenomenon might be explained by the fact that the capacity of authors to interact with and select partners was limited to nearby researchers that studied different topics. This effect of distance on the selection of peers by authors was supported by the strong effect of distance during the first decade.

As mentioned above, distance had an overall negative effect on the frequency of co-authorship. Thus, this result confirmed that, at least at large scales, distance was a good predictor of social interactions (Liben-Nowell et al. 2005), with scientific collaborations following a gravity law. In other words, authors are attracted to geographically closer co-authors (Pan et al. 2012). Albeit, distance effects have been minimised with lower costs of transports and new ways of communicating (e.g., video conferences). This phenomenon was demonstrated by the decreasing effect to distance from strong during the first decades to weak during the latter decades. A possible hypothesis for the decreasing importance of distance is the need for data to test large-scales patterns in ecology, so called macroecology (Brown and Maurer 1989), might stimulate long-distance partnerships with authors scattered around the world to sample and test large scale patterns and hypotheses.

I found no evidence that authors interact more often with authors of more repute than themselves. Actually, I found evidence in favour of homophily, or *bird of a feather* hypothesis, which predicts that people seek interactions with other people with similar characteristics (Freeman 1996). Similarly, there was no consistent effect on how an ecologist seeks co-authorship based on the level of development of the country, and no evidence to support the centre-periphery hypothesis. Furthermore, the negative effects over two decades support the homophily hypothesis, as authors interacted more often with peers from countries with similar levels of development. Of note, even if developing countries increased their publications in relation to developed countries, authors from developing countries failed to publish, or published less often, in ecological journals with higher impact factor than their counterparts from developed countries (Holmgren and Schnitzer 2004). Therefore, the number of authors from developing countries was limited, and less likely to interact with authors from developed countries.

Here, I demonstrated that co-authorships among ecologists are explained by the homophily hypothesis, whether they have similar research interests, reputations, or country developmental status. Furthermore, I believe that distance might limit the ability of authors to choose partners using the stated criteria. Similar patterns might be found in other fields of biology that are characterised by multiple coauthors (Newman 2004) or distance effects on collaboration (Ponds et al. 2007). However, caution is needed when extrapolating my results to other scientific fields, which might have different drivers of co-authorship or different degrees of importance in the drivers assessed here. For example, economists are more prone to form co-authorships with more experienced researchers and when there are opportunities for field specialisation (McDowell and Melvin 1983). In comparison, social sciences co-authorships are more frequent between authors with similar methodological and theoretical approaches (Moody 2004).

**Acknowledgements** I received a scholarship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). I am grateful for all comments made on the manuscript during the peer-review process.

## References

Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509.

Borrett, S. R., Moody, J., & Edelmann, A. (2014). The rise of network ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling, 293,* 111–127. https://doi.org/10.1016/j.ecolmodel.2014.02.019.

Brown, J. H., & Maurer, B. A. (1989). Macroecology: The division of food and space among species on continents. *Science, 243*(4895), 1145–1150. https://doi.org/10.1177/0309133308094425.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*(2), 191–235. https://doi.org/10.1177/053901883022002003.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), 66111. https://doi.org/10.1103/PhysRevE.70.066111.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Day, R. (1989). The origins of the scientific paper: The IMRaD format. *American Medical Writers Association Journal, 4*(2), 16–18.

de Solla Price, D. J. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515. https://doi.org/10.1126/science.149.3683.510.

Feinerer, I., & Hornik, K. (2017). tm: Text mining package. R package version 0.7-1. https://CRAN.R-project.org/package=tm. Accessed 30 May 2017.

Freeman, L. C. (1996). Some antecedents of social network analysis. *Connections, 19*(1), 39–42.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102.

Hijmans, R. J. (2016). raster: Geographic data analysis and modeling. R package version 2.5-8. https://CRAN.R-project.org/package=raster. Accessed 30 May 2017.

Holmgren, M., & Schnitzer, S. A. (2004). Science on the rise in developing countries. *PLoS Biology, 2*(1), 10–13. https://doi.org/10.1371/journal.pbio.0020001.

Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics, 88*(2), 653–661. https://doi.org/10.1007/s11192-011-0412-z.

Kahle, D., & Wickham, H. (2013) ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences, 102*(33), 11623–11628. https://doi.org/10.1073/pnas.0503018102.

Logan, J. M. (2016). Historical changes in co-author numbers in ecology. *Frontiers in Ecology and the Environment, 14*(6), 297–299. https://doi.org/10.1002/fee.1301.

McDowell, J. M., & Melvin, M. (1983). The determinants of co-authorship: An analysis of the economics literature. *The Review of Economics and Statistics, 65*(1), 155–160.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56.

Moody, J. (2004) The structure of a social science collaboration network: Disciplinary cohesio from 1963 to 1999. *American Journal of Sociology* 69(2):213–238. http://www.jstor.org/stable/3593085.

Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics, 80*(3), 657–682. https://doi.org/10.1007/s11192-008-2099-3.

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

Pan, R. K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: Uncovering the role of geography in science. *Scientific Reports, 2,* 902. https://doi.org/10.1038/srep00902.

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*. https://doi.org/10.7554/eLife.27725.

Ponds, R., van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science, 86*(3), 423–443. https://doi.org/10.1111/j.1435-5957.2007.00126.x.

R Development Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 30 May 2017.

Thomson Reuters. (2016). 2015 Journal Citation Reports® Science Edition. http://jcr.incites.thomsonreuters.com/. Accessed 30 May 2017.

Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology and Evolution, 17*(8), 349–350. https://doi.org/10.1016/S0169-5347(02)02545-4.

United Nations Development Programme. (2016). Human development report 2016: Human development for everyone. Washington, DC: Communications Development Incorporated.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608–1618. https://doi.org/10.1016/j.respol.2005.08.002.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099.