Scientometrics (2021) 126:201–224
https://doi.org/10.1007/s11192-020-03746-5

The coauthorship networks of the most productive European researchers

Marian-Gabriel Hâncean[1] · Matjaž Perc[2,3,4] · Jürgen Lerner[5]

Received: 4 February 2020 / Published online: 29 October 2020

Abstract

The world of science possesses an inherent inequality in the distribution of research output and impact. Only a small minority of researchers is responsible for the majority of the knowledge production. However, little is known about the factors that might explain the prestige and the working habits of these researchers. In this paper, we therefore examine the coauthorship networks of the most productive European researchers over a 12-year time window, between the years 2007 and 2018. Explicitly, we look at the impact that these collaborative structures have upon the citations of the researchers. We show that highly productive researchers gain benefits in terms of citations by increasing their research output, by embedding themselves in large geographically dispersed coauthorship networks, as well as by publishing with highly cited collaborators. These results substantiate a prestige effect (the best tend to collaborate with the best) that governs the behaviour of the most productive researchers. Our study thus contributes to the currently coalescing literature on profiling the European research elite, and we hope it will be informative to policy-makers in their efforts of driving Europe towards an integrated research area.

Keywords Highly productive researchers · Coauthorship networks · Citations · Research productivity

✉ Marian-Gabriel Hâncean
gabriel.hancean@sas.unibuc.ro

1 Department of Sociology, University of Bucharest, 90-92 Panduri St., 050663 Bucharest, Romania

2 Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia

3 Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404, Taiwan

4 Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria

5 Department of Computer and Information Science, University of Konstanz, Universitaetsstrasse 10, 78464 Constance, Germany

Introduction

It has been shown that the individual-level citation distributions are affected not only by the research productivity, but also by the collaborative structures embedding the researchers (Uddin et al. 2019). These structures can be fruitfully approached, explored and statistically modelled by employing *social network analysis*, i.e., a social science research methodology focused on collecting data about the relationships among social entities (Wasserman and Faust 1994). Basically, two researchers are said to be connected or to share a coauthorship tie if they sign together at least one publication. These dyadic ties are the building blocks of coauthorship networks (Moody 2004).

Our study seeks to address the effects of coauthorship networks on the research performance at an individual level. Previous literature addressing this topic is notably diverse in terms of the employed methods, the approached subjects or the availing statistical models and techniques. Network structure measurements (Abbasi et al. 2012b; González-Alcaide et al. 2020) as well as network compositional attributes (Badar et al. 2015; Ronda-Pupo and Pham 2018) have been shown to explain high research performance, preferential attachment or rich get richer effect (Perc 2014). Cross-sectional (Abbasi et al. 2011a) or longitudinal (Ductor 2015) data have been collected to reflect one or more research fields, academic journals, countries or regions (Hou et al. 2008; Scarazzati and Wang 2019; Sun and Rahwan 2017; Wang et al. 2015). Statistical models and techniques, such as correlation and cluster analysis (Abbasi et al. 2012b; Medina 2018), factor analysis (Martín-Alcázar et al. 2019), node-level regression models (Hâncean et al. 2014; Tu 2019), as well as machine learning algorithms (Sarigöl et al. 2014) have been put to work to explain or predict scientific success out of network characteristics. Studies looking at the relationship between coauthorship networks and research performance have typically adopted two research designs. On one hand, *the ego-centric network design* that takes stock of a focal researcher (or *an ego*), her collection of co-authors (or *the alters*) and the special configuration of coauthorship ties connecting them (Abbasi et al. 2012a; Wang 2016). On the other hand, *the socio-centric network design* that pays attention to how coauthorship ties configure among the researchers embedded in bounded social units (departments, universities, domestic fields of research, countries etc.) (Abbasi et al. 2010, 2018; Guan et al. 2015a, b, 2016; Hâncean et al. 2014; Perc 2010).

In the literature, few studies have been addressing the particular case of top performers, such as Nobel laureates (Gallotti and De Domenico 2019; Merton 1968), network stars (Andrikopoulos et al. 2020), top Italian scientists (Abramo et al. 2019), academic elite working in China (Yin and Zhi 2017), top researchers in Polish higher education system (Kwiek 2018). Evidence profiling successful researchers is still coalescing, despite the growing interest, among scholars, policy-makers and university top-management (Kwiek 2016). Publication productivity and citations were proved to have a skewed distribution (a small minority accounts for the largest share of the output) (Perc 2014). In this context, the characteristics of the top performers are intrinsically crucial for disentangling excellence in research.

The aim of this study is to extend current knowledge of the most productive European Union (EU) researchers ($n=4588$ that account for 0.03% of Scopus indexed researchers). We examine and analyse their coauthorship networks, both structurally and compositionally. Also, we assess the impact of these collaborative structures on the individual research success (citations). We provide evidence that might be useful for the currently emerging efforts of profiling most productive researchers. In this regard, we advance a novel longitudinal-wise research focused on a balanced panel data of focal researchers embedded in ego-centric networks. This could possibly support designing research policies and productivity strategies on multiple layers (faculty, institutional, country and EU level). Also, we contribute to the area of collaboration network studies, through employing an ego-centric network design wherein EU top researchers are regarded as focal nodes.

The remainder of this paper is organized as follows. First, we introduce the hypotheses referring to the effect of coauthorship networks on individual citations and draw the corresponding literature. Second, we present the key elements of the study design, the data sources and measurement, the variables and the statistical methods. Third, we highlight the main findings and end up by discussing our data, their implication and limits.

Background

The network perspective in studying scientific coauthorship

There is a considerable amount of literature arguing that research activity and productivity are inherently collaborative (Cummings and Kiesler 2005; Katz and Martin 1997). Some of the work conducted on this venue has grasped a *collaboration network perspective* (Newman 2004). According to this perspective, it has been found that not only the attributes of the researchers but also their patterns of relationships are relevant for advancing explanatory models (Badar et al. 2013, 2014, 2016; Rotolo and Messeni Petruzzelli 2013). Essentially, studies using this network perspective can be divided into two classes (Brass et al. 2004). One class investigates the antecedents (how networks emerge) while the other examines the consequences of the coauthorship networks (their impact). Put it differently, one considers selection processes while the other, networks as conduits for influence.

In terms of the antecedents, various social selection effects have been proposed: the tendency to co-author with similar others (homophily), with more prestigious researchers (preferential attachment, Matthew effect), or the tendency of one's collaborators to write together (transitivity). For instance, it has been provided evidence on research performance homophily in sociology (Hâncean and Perc 2016), ethnic homophily (Freeman and Huang 2015), sex homophily in economics (Boschini and Sjögren 2007), Matthew effect in the case of coauthorship ties to Nobel laureates (Merton 1968) or more prestigious others (Borjas and Doran 2015; McCarty et al. 2013; Yin and Zhi 2017), transitivity in computer science (Zhang et al. 2018), clustering and preferential attachment in physics, biology (Newman 2001), mathematics and neuro-science (Barabási et al. 2002; Jeong et al. 2003), or preferential attachment in Slovenia's scientific collaboration network (Perc 2010). Additionally, other studies have examined the preference to establish ties with coauthors in the physical proximity (same country, same city, same institution) (Hoekman et al. 2010). These findings have suggested that coauthorship structures are not random but exhibit *network autocorrelation* as a specific feature (Friemel 2015). This class of research has been critical for understanding how researchers select their collaboration ties and cluster together. Selection mechanisms have been brought forth to explain why authors create ties with similar others (*positive selection*) and drop ties with different others (*negative selection*).

In parallel, a great deal of attention has been paid to the consequences that the coauthorship network properties have on research productivity and performance. Networks have been addressed both structurally and compositionally (i.e., patterning and content).

Irrespective of the research design (either ego-centric or socio-centric), one line of work has looked at the impact of structural network features such as: size (number of co-authors) (Abbasi et al. 2010; Biscaro and Giupponi 2014), density and centrality measures (Abbasi et al. 2011b, 2018; Li et al. 2013; Liao 2011), normalized betweenness and closeness (Abbasi et al. 2011a), or unweighted betweenness (Abbasi et al. 2012a, b). Another line has focused on the compositional traits of the networks such as: the geographical diversity of the co-authors (Abbasi and Jaafari 2013; Gazni et al. 2012; Gazni and Didegah 2011; Sugimoto et al. 2017; Wang et al. 2015), the strength of ties (Ding 2011; Petersen 2015; Wang 2016) or collaboration frequency (Abbasi 2013). All these studies have assumed the idea that networks act as channels for influence and diffusion. And, consequently, specific network configurations and compositions have been found to hold a significant effect on research prominence and success.

Research hypotheses

Research productivity and citations are the result of complex dynamics. These are affected by both network formation and by resulting configurations. Based on the above discussion, we state that research performance is the effect of how co-authors are selected and, at the same time, of how coauthorship networks configure. Selection and influence processes are difficult to disentangle unless a time framework is provided. Accordingly, in this paper, adopting a 12-year time window, we control for selection effects and examine the impact of structural and compositional characteristics. Our work looks at the consequences of coauthorship networks and seeks to account for the citations distributions of the 4588 EU most prominent researchers.

In terms of structural network effects, we test for the degree centrality (the number of collaborators or the size of the ego-network) as well as for the strength of ties (the repeated collaboration). There is mixed evidence concerning the impact of these variables on the citation counts. On a dataset file of 5585 papers published between 1985 and 2013, in the area of climate change and disaster risk, it was argued that the degree centrality has a positive impact on citations counts (Biscaro and Giupponi 2014). Also, publication records in the field of information science and library science, between 2000 and 2009, indicated that researchers with a higher degree centrality perform better (Abbasi et al. 2012a) or have a higher quality output (Abbasi et al. 2018). Similar results were reported for the field of chemistry, while controlling for the quality of the publications (Bornmann et al. 2012). However, an analysis on longitudinal data (1980–2002, 1192 articles) in the field of strategic management highlighted that the number of co-authors does not have any effect on citations counts, in the presence of brokerage (a network characteristic indicating authors connecting otherwise disconnected researchers) (Collet et al. 2014). Another study on articles in the vascular and endovascular literature revealed that the number of co-authors was not among the factors predicting increased citations (Antoniou et al. 2015). Despite these inconclusive evidences, we hypothesize a positive impact of the number of collaborators (Hypothesis 1: *Degree centrality positively affects the citations of most prominent EU researchers*). In setting this hypothesis, we build on a previously reported longitudinal study (1960–2000) showing the positive effect of publishing in coauthorship in the context of 2.1 million patents and 19.9 million papers published in various fields (sciences and engineering, social sciences, arts and humanities) (Wuchty et al. 2007).

Few studies have examined in detail whether the strength of the collaboration ties affects the quality of research output. Results are inconclusive. Some authors reported a positive impact that average tie strength has on researchers' performance (Abbasi et al. 2011a). On the other hand, an inverted U-shaped relationship was found between network average tie strength and citation impact, on a panel of 1042 American scientists in several disciplines (biology, chemistry, computer science, earth and atmospheric sciences, electrical engineering, and physics), with papers published between 1980 and 2009 (Wang 2016). We aim to clarify these findings by testing a positive relationship between the frequency of co-authoring with the same collaborators and the citation counts (Hypothesis 2: *Repeated collaboration positively affects the citations of most prominent EU researchers*).

Moreover, we extend the knowledge on the geographical dispersion of the collaborators embedded in the networks of the most prominent EU researchers. Several studies have already demonstrated that an increase in the number of countries or of collaborations across geographical boundaries leads to an increase in the number of citations (Abramo et al. 2019; Gazni et al. 2012; Larivière et al. 2015; Puuska et al. 2014). However, a bias to collaboration with physically proximate partners was found in Europe (between 2000 and 2007) (Hoekman et al. 2010), in Brazil (Sidone et al. 2017), Turkey (Gossart and Özman 2009) or Finland (Puuska et al. 2014). Spatial patterns of collaboration were also found among researchers in agricultural sciences, humanities, health sciences or ecology (Parreira et al. 2017). Therefore, our third hypothesis tests for the positive impact of geographical dispersion on citation counts (Hypothesis 3: *Geographical dispersion of collaborators positively affects the citations of most prominent EU researchers*). This predictor takes stock of the number of unique different countries wherein collaborators are based.

The impact of these structural (the individual-level degree centrality and the strength of collaboration ties) and compositional (the geographical dispersion of co-authors) network predictors is assessed while controlling for selection process (network antecedents) effects. Specifically, collaborators' performance (co-authors' citations) and proximity homophily (same country, same city, same institution). Additionally, the number of papers (author's productivity) is also controlled for, due to its positive impact upon the citation counts and the number of co-authors (Bornmann and Daniel 2007).

Methods

Data and study design

Our paper aims to profile the most productive European Union (EU) researchers, irrespective of their research field. Accordingly, we cut-off the first 5000 researchers solely based on paper productivity. Subsequently, to build a balanced panel, we keep 4588 of them. The rationale of profiling this specific sample of researchers is to increase the understanding of the processes underpinning their out-of-the-ordinary research productivity as well as the citations of their work. We explicitly stress that our study is not about the most productive researchers in specific research fields. Given our general objective (aim), we do not perform any stratified or quota sampling and we are not interested in looking at the most productive researchers in various specific fields. The sample of 4588 researchers is followed longitudinally for 12 years (2007–2018) resulting in a total of 55,056 observations. The individuals included in the panel are the EU-based researchers with the highest research output in the analysed time window.

In our study, we use the term *ego* to refer to any of the panel members and the term *alters* to designate an ego's co-authors (this terminology is specific to social network studies). For both egos and alters, data on personal characteristics (institutional affiliation, country, and city), on research productivity (papers and citations) and on coauthorship relationships are retrieved from Scopus (the world's largest database of peer-reviewed research literature; http://www.scopus.com). This information is used to test the research hypotheses.

For each of the panel members, the papers are indexed based on the publication year. We include in the analysis papers yearly published between 2007 and 2018. Subsequently, papers are assigned citation counts, i.e. the number of citations received since the publication year. The citation counts are dependent on the publication date. The older the publication, the more likely to receive more citations. In effect, for a specific ego, papers published in 2007 are assigned the citations received from the year of publication and until 2018. In a similar fashion, for example, papers published in 2011 are assigned the citations received since the year of publication and until 2018. A similar procedure is used to compute the citations of the alters' publications.

We employ an ego-centric network study design. We build and analyse coauthorship ego-networks wherein the ego (the focal node) is a panel researcher while the alters are her co-authors. Coauthorship ties connect ego to alters as well as alters to alters, providing that they co-published at least one paper. In our collection of coauthorship ego-networks, we also mark whether two nodes repeatedly co-authored (the strength of the tie or *repeated collaboration*). For each of the panel researchers, we create 12 ego-networks (Fig. 1). Each ego-network reflects one of the 12 years (2007–2018). These networks capture the variables used in the analysis. Some of these variables measure network properties: number of *co-authors* (network

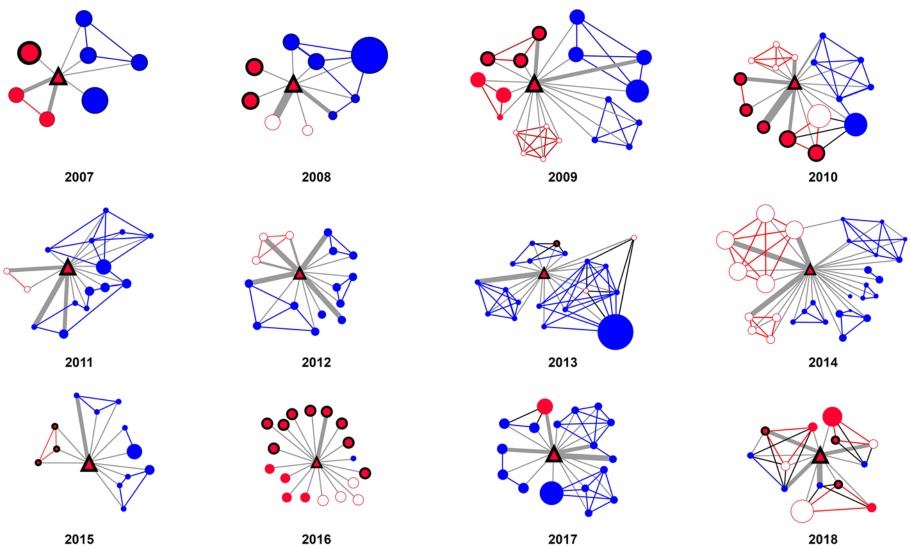

Fig. 1 The coauthorship ego-networks of a researcher in each year, over 12 years (2007–2018). *Note*. Two nodes (researchers) share a tie if they co-authored at least one paper. As the tie thickness increases, the number of co-authored papers increases (the strength of the tie). The size of the nodes marks the number of citations. The triangle indicates the ego (the researcher) and the circled-nodes designate her alters (co-authors). The colours illustrate geographical location. Nodes in blue are alters based in a different country than the ego's. Red variants mark: *same institution* as the one of the ego's (black-bordered red nodes), *same city* (full red bodes) and *same country* (red-bordered empty nodes). (Color figure online)

size) and *repeated collaborations* (the strength of the ties or how many times two researchers co-published). Other variables measure compositional properties: the number of *papers*, the number of *co-authors*, the geographical *dispersion* (the number of unique countries wherein alters are based), the institutional affiliation data (the number of alters from the *same country*, *same institution* and *same city*), and alters' research productivity (papers and citations).

Statistical methods and variables

In our study design, the dependent variable is the count of the citations received by the egos' articles published in a specific year. Additionally, our data come from a time-series cross-sectional balanced panel of individual researchers. Following the practice in the field (Hausman et al. 1984) and due to handling over-dispersion in our data (Hilbe 2011), we employ negative binomial regression models (Shoukri 2018). Specifically, we use the negative binomial distribution as it is implemented in a generalized linear model frame by the *glm.nb ()* function in the MASS R-package (Venables and Ripley 2010). In the models, we keep the citations counts (the dependent variable) as raw integer numbers. We also include the publication year as a factor variable to control for the fact that papers' citations are dependent on the publication date. This is an alternative equivalent solution to normalizing by computing the average number of citations over all papers published in any given year and divide the actual counts by that average.

Our statistical models include three independent variables: the number of co-authors (the ego degree centrality or ego-network size), repeated collaboration (the strength of coauthorship ties) and geographical dispersion (the number of unique countries wherein co-authors are based). Before being introduced into the models, *the number of co-authors* is divided by the number of authors' papers while the *number of repeated collaborations* is divided by the number of co-authors. Additionally, we use several control variables in the full version of our statistical models. Firstly, to control for the homophily effect, we enter the number of co-authors from *same country*, *same institution* and *same city* (before being introduced into the full models, these variables are divided by the number of co-authors). Secondly, to control for prestige effect, we enter *the co-authors' citations* (divided by the number of co-authors' publications). The *year* (as a factor variable), the ego's *country* and ego's *number of papers* are introduced in all the models (simple and full models) to control for global effects. A lagged variable for yearly citation counts with a shift of one year (lag 1 autocorrelation) is also entered to correct for serial correlation, i.e., the number of citations received by a researcher in the previous year ($t-1$).

The negative binomial regression models are fitted both on the whole sample panel of 4588 researchers, as well as on two sub-samples of 1155 individuals each (the upper and the lower 25% of observations on paper productivity). This multiple application of the models allows us to examine the predictors' behaviour. The results of the statistical fit are expected to indicate, for example, whether the upper 25% of the observations (the most productive 25% of the most productive) exhibit different patterns of coauthorship compared to the whole sample and to the lower 25% subsample.

Results

Descriptive statistics

The largest shares of panel researchers are based in Germany (24%), United Kingdom (14%), Netherlands (12%), Italy (11%), France (9%), and Spain (6%). These six countries account for more than 76% of all EU most productive researchers (Table 5 in the "Appendix"). Additionally, ten countries host approximately 90% of the 4588 individuals (among these, we do not find any East European country). More than a third (35%) of the co-authors embedded in the co-publishing networks of the most productive researchers are based in countries of the European Union. Six countries account for approximately 25% of the total number of co-authors: Germany, United Kingdom, Italy, France, Netherlands and Spain (Table 6 in "Appendix").

Table 1 reports descriptive statistics for the dependent variable, as well as for all the other variables. For illustration, we also present in Table 1 the number of co-authors based in different countries. On average, we found that citations exhibit a decreasing pattern. We note that the older the publication, the more likely to receive more citations. In addition, the citation window comes into play for the last year in our timeline (2018). In this respect, we comment that articles published in the first months of the year probably attract more citations compared to articles published in the last months (Levitt and Thelwall 2011). Particularly, papers published in 2007 received on average 721.7 citations ($SD=856.6$) while papers, in 2018, received on average 11.5 citations ($SD=18.6$). A similar trend is displayed by the co-authors' citations divided by their articles. The ratios constantly decreased from 2007 ($M=53.1$, $SD=23.0$) to 2018 ($M=1.0$, $SD=0.5$). It is noteworthy that the number of egos' papers illustrate a drop at the end of the time window. This is most probably due to the lower coverage of SCOPUS. However, we do not exclude other factors (e.g., career age, funding opportunities, etc.). We decided to keep the entire time window as long as we do not know the precise causes. The number of co-authors (divided by the number of ego's papers) and the geographical dispersion exhibit positive trends. Specifically, the number of co-authors increased from an average of 4.2 ($SD=3.1$) in 2007 to an average of 7.9 ($SD=6.9$) in 2018. The number of unique countries wherein co-authors are based followed a positive slope: from an average of 7.1 ($SD=5.2$) in 2007 to an average of 9.7 ($SD=7.6$) in 2018. At the same time, the repeated collaborations are rather decreasing (this development is probably affected by the positive trend in the number of co-authors): $M=1.8$ ($SD=1.1$) in 2007 and $M=1.5$ ($SD=1.3$) in 2018. The time pattern of the various groupings of co-authors (i.e., from the same country, same city and same institution) is rather constant; with small variations that are observable if scores were reported with three digits after the decimal point. Except for 2007 ($M=0.5$, $SD=0.2$), the average of same country co-authors is 0.4 ($SD=0.2$). For the case of the same city and same institution co-authors, the average is rather constant ($M=0.2$, $SD=0.2$ and $M=0.1$, $SD=0.1$, respectively). The number of co-authors from different countries (divided by the total number of co-authors) is higher compared to the number of domestic partners. Furthermore, from 2007 until 2018, this number slightly increased ($M=0.5$, $SD=0.2$ and $M=0.6$, $SD=0.2$, respectively).

Tables 2, 3 and 4 present the results of the negative binomial regression models for predicting citation counts for the panel data ($n=4588$), as well as for the lower 25% ($n=1155$) and upper 25% ($n=1155$) of the sample, respectively. In each table, two statistical models are displayed: a simple model that do not control for *co-authors' citations* (the prestige

Table 1 Descriptive statistics (number of observations per each year: 4588). *Source*: Authors' own calculations

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Citations												
Mean	721.7	723.3	768.7	743.4	668.9	630.2	515.7	418.3	311.0	180.7	71.4	11.5
SD	856.6	863.8	1269.3	896.2	837.4	766.1	551.5	522.4	425.8	236.7	83.2	18.6
Min	0.0	0.0	0.0	0.0	0.0	0.0	4.0	2.0	0.0	0.0	0.0	0.0
Median	469.0	480.5	503.0	490.0	443.0	418.0	358.0	277.0	194.0	117.0	46.0	6.0
Max	11,158.0	14,164.0	40,562.0	17,945.0	25,587.0	9450.0	11,970.0	15,642.0	10,478.0	4043.0	1047.0	631.0
Papers												
Mean	16.0	16.6	18.1	18.8	19.9	20.4	21.4	21.3	20.9	20.2	19.1	15.9
SD	10.6	10.4	10.7	10.8	11.1	11.1	11.6	11.8	11.8	12.3	12.4	11.1
Min	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Median	14.0	15.0	16.0	17.0	17.0	18.0	19.0	19.0	18.0	18.0	17.0	14.0
Max	139.0	112.0	150.0	142.0	142.0	136.0	138.0	132.0	177.0	214.0	241.0	220.0
Co-authors (divided by ego's papers)												
Mean	4.2	4.5	4.6	4.8	4.9	5.3	5.1	5.5	5.8	6.2	6.6	7.9
SD	3.1	3.9	3.4	3.8	3.4	4.3	3.4	3.8	3.9	4.1	4.9	6.9
Min	0.3	0.3	0.3	0.4	0.3	0.4	0.4	0.4	0.5	0.4	0.5	0.5
Median	3.5	3.6	3.8	3.9	4.1	4.3	4.3	4.6	4.9	5.3	5.5	6.1
Max	42.8	99.0	55.0	80.0	43.7	86.0	38.0	48.0	89.0	58.0	81.0	93.0
Repeated collaboration (divided by ego's number of co-authors)												
Mean	1.8	1.8	1.9	1.8	1.9	1.8	1.9	1.8	1.8	1.7	1.7	1.5
SD	1.1	1.6	1.6	1.2	1.7	1.2	2.3	1.7	1.9	1.5	1.6	1.3
Min	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Median	1.6	1.5	1.6	1.6	1.5	1.5	1.6	1.5	1.5	1.5	1.4	1.3
Max	18.1	36.7	30.2	21.1	30.3	21.3	51.7	45.6	40.7	36.4	41.6	32.7
Dispersion												
Mean	7.1	7.7	8.5	9.3	9.6	10.0	10.1	10.4	10.6	10.5	10.2	9.7
SD	5.2	5.5	5.9	6.3	6.1	6.2	6.2	6.6	6.8	7.1	7.1	7.6

Table 1 (continued)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Median	6.0	7.0	7.0	8.0	8.0	9.0	9.0	9.0	9.0	9.0	9.0	8.0
Max	35.0	39.0	44.0	45.0	46.0	39.0	43.0	61.0	46.0	52.0	54.0	61.0
Co-authors' citations (divided by the number of co-authors' papers)												
Mean	53.1	49.6	49.8	45.2	39.4	36.1	28.7	23.5	18.1	10.9	4.9	1.0
SD	23.0	20.0	28.2	18.5	17.2	16.7	11.6	10.0	8.5	4.5	1.9	0.5
Min	5.4	3.7	3.4	3.1	3.9	2.9	2.9	2.9	2.0	1.5	0.4	0.0
Median	52.8	50.5	48.7	45.8	39.1	35.2	28.7	22.9	17.3	10.4	4.7	0.9
Max	168.3	184.4	779.5	179.3	234.7	201.9	217.0	135.1	103.9	60.5	17.8	8.2
Different country co-authors (divided by the number of co-authors)												
Mean	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
SD	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Median	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Max	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Same country co-authors (divided by the number of co-authors)												
Mean	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
SD	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Median	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Max	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Same city co-authors (divided by the number of co-authors)												
Mean	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
SD	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1 (continued)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Median	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Max	1.0	1.0	1.0	0.9	1.0	1.0	0.9	1.0	1.0	1.0	1.0	1.0
Same institution co-authors (divided by the number of co-authors)												
Mean	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SD	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Median	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Max	1.0	1.0	1.0	0.9	1.0	1.0	0.8	1.0	1.0	1.0	0.9	1.0

Table 2 Negative binomial regression models predicting citation counts for full panel data ($n=4588$). *Source*: Authors' own calculations

Independent variables[b,c]	Model 1[a]					Model 2[a]				
	Est.	S.E.	z	p value	95% CI [LL, UL][d]	Est.	S.E.	z	p value	95% CI [LL, UL][d]
Co-authors	0.004	0.001	4.746	0.000	[0.003, 0.006]	−0.009	0.001	−10.204	0.000	[−0.010, −0.007]
Repeated collaboration	−0.001	0.002	−0.336	0.737	[−0.005, 0.003]	0.005	0.002	2.667	0.008	[0.001, 0.008]
Dispersion	0.047	0.001	71.744	0.000	[0.046, 0.048]	0.035	0.001	53.783	0.000	[0.034, 0.036]
Co-authors' citations						0.026	0.000	126.171	0.000	[0.026, 0.027]
Same country						0.061	0.019	3.165	0.002	[0.023, 0.098]
Same city						−0.095	0.027	−3.577	0.000	[−0.148, −0.043]
Same institution						−0.147	0.031	−4.732	0.000	[−0.208, −0.086]
Papers	0.022	0.000	61.154	0.000	[0.021, 0.022]	0.028	0.000	87.084	0.000	[0.027, 0.029]
Citations (lagged, 1 year)	0.000	0.000	90.264	0.000	[0.000, 0.000]	0.000	0.000	46.387	0.000	[0.000, 0.000]
(Intercept)	3.858	0.222	17.408	0.000	[3.424, 4.293]	2.983	0.199	14.996	0.000	[2.593, 3.373]
AIC					644,867.836					632,894.246
BIC					645,238.658					633,300.384
Pseudo-R^2 (McFadden)					0.166					0.216

[a]Observations: 50,468 (4588 missing observations deleted as an effect of entering the citations lagged variable (1-year lag autocorrelation))

[b]*Country* (factor) and *Year* (factor) variables were omitted from the table

[c]All variables have variance inflation factor scores (VIF) less than three

[d]95% Confidence intervals with lower and upper limits

Table 3 Negative binomial regression model predicting citation counts for the lower 25% of the panel data ($n = 1155$). *Source*: Authors' own calculations

Independent variables[b,c]	Model 1[a]					Model 2[a]				
	Est.	S.E.	z	p value	95% CI [LL, UL][d]	Est.	S.E.	z	p value	95% CI [LL, UL][d]
Co-authors	0.012	0.002	6.464	0.000	[0.008, 0.016]	−0.003	0.002	−1.517	0.129	[−0.006, 0.001]
Repeated collaboration	−0.038	0.009	−4.136	0.000	[−0.056, −0.020]	−0.028	0.008	−3.441	0.001	[−0.044, −0.012]
Dispersion	0.043	0.002	27.925	0.000	[0.040, 0.046]	0.033	0.002	21.516	0.000	[0.030, 0.036]
Co-authors' citations						0.026	0.000	63.128	0.000	[0.026, 0.027]
Same country						0.102	0.037	2.735	0.006	[0.029, 0.175]
Same city						−0.118	0.052	−2.265	0.024	[−0.221, −0.016]
Same institution						−0.148	0.064	−2.304	0.021	[−0.273, −0.022]
Papers	0.063	0.001	43.173	0.000	[0.060, 0.066]	0.065	0.001	49.189	0.000	[0.062, 0.067]
Citations (lagged, 1 year)	0.000	0.000	42.968	0.000	[0.000, 0.001]	0.000	0.000	22.570	0.000	[0.000, 0.000]
(Intercept)	3.403	0.227	15.002	0.000	[2.959, 3.848]	2.536	0.205	12.389	0.000	[2.135, 2.937]
AIC					155,104.709					152,118.297
BIC					155,402.699					152,446.086
Pseudo-R^2 (McFadden)					0.152					0.200

[a]Observations: 12,705 (1155 missing observations deleted as an effect of entering the citations lagged variable (1-year lag autocorrelation))

[b]*Country* (factor) and *Year* (factor) variables were omitted from the table

[c]All variables have variance inflation factor scores (VIF) less than three

[d]95% Confidence intervals with lower and upper limits

Table 4 Negative binomial regression model predicting citation counts for the upper 25% of the panel data ($n = 1155$). *Source*: Authors' own calculations

Independent variables[b, c]	Model 1[a]					Model 2[a]				
	Est.	S.E.	z	p value	95% CI [LL, UL][d]	Est.	S.E.	z	p value	95% CI [LL, UL][d]
Co-authors	−0.005	0.002	−3.524	0.000	[−0.008, −0.002]	−0.017	0.001	−11.664	0.000	[−0.020, −0.014]
Repeated collaboration	0.005	0.002	2.081	0.037	[0.000, 0.009]	0.010	0.002	4.946	0.000	[0.006, 0.014]
Dispersion	0.046	0.001	44.995	0.000	[0.044, 0.049]	0.035	0.001	33.092	0.000	[0.033, 0.037]
Co-authors' citations						0.025	0.000	66.436	0.000	[0.025, 0.026]
Same country						0.030	0.039	0.766	0.444	[−0.046, 0.106]
Same city						0.007	0.055	0.134	0.893	[−0.101, 0.116]
Same institution						−0.265	0.060	−4.422	0.000	[−0.383, −0.148]
Papers	0.011	0.000	24.071	0.000	[0.010, 0.012]	0.016	0.000	37.299	0.000	[0.015, 0.017]
Citations (lagged, 1 year)	0.000	0.000	55.421	0.000	[0.000, 0.000]	0.000	0.000	30.494	0.000	[0.000, 0.000]
(Intercept)	4.740	0.201	23.574	0.000	[4.346, 5.134]	3.969	0.181	21.981	0.000	[3.615, 4.323]
AIC					171,647.998					168,697.64
BIC					171,938.539					169,017.98
Pseudo-R^2 (McFadden)					0.188					0.239

[a]Observations: 12,705 (1155 missing observations deleted as an effect of entering the citations lagged variable (1-year lag autocorrelation))

[b]*Country* (factor) and *Year* (factor) variables were omitted from the table

[c]All variables have variance inflation factor scores (VIF) less than three

[d]95% Confidence intervals with lower and upper limits

effect) and co-authors from *the same country*, *city* and *institution* (proximity homophily effect), and a full model (including all the variables). Also, in the tables, the estimates (*Est.*) are provided together with their corresponding 95% confidence interval (*95% CI*), standard errors (*SE*) and the exact *p value*.

In Table 2, as predicted, the full model shows that *repeated collaboration* (*Est.* = 0.005, *SE* = 0.002, *p* = 0.008, 95% CI [0.001, 0.008]) and *dispersion* (*Est.* = 0.035, *SE* = 0.001, *p* = 0.000, 95% CI [0.034, 0.036] show a positive statistically significant impact on the citation counts. Contrary to our expectations, the size of the ego-network (*the number of co-authors*) displays a negative statistically significant effect (*Est.* = −0.009, *SE* = 0.001, *p* = 0.000, 95% CI [−0.010, −0.007]). Referring to the control variables, we note that research productivity (*papers*) positively impacts upon citations (*Est.* = 0.028, *SE* = 0.000, *p* = 0.000, 95% CI [0.027, 0.029]). Across the ego-networks of the panel data researchers, the prestige of the alters (*co-authors' citations*) is shown to be statistically significant (*Est.* = 0.026, *SE* = 0.000, *p* = 0.000, 95% CI [0.026, 0.027]). Results concerning the geographical proximity homophily are mixed. The number of co-authors from the *same country* has a statistically significant effect on citations (*Est.* = 0.061, *SE* = 0.019, *p* = 0.002, 95% CI [0.023, 0.098]). However, co-authors from *the same city* (*Est.* = −0.095, *SE* = 0.027, *p* = 0.000, 95% CI [−0.148, −0.043]) and from *the same institution* (*Est.* = −0.147, *SE* = 0.031, *p* = 0.000, 95% CI [−0.208, −0.086]) entail a negative effect on the dependent variable.

From the balanced panel data (*n* = 4588), we extracted two sub-samples. Specifically, we cut-off the lower and the upper 25%, based on authors' paper productivity (number of papers). In this section, Tables 3 and 4 present the results of negative binomial regression models predicting citation counts for the lower 25% observations (*n* = 1155) and for the upper 25% observations (*n* = 1155), respectively. In Table 3, the full model indicates that the number of *co-authors* do not have any impact on the citation counts (*Est.* = −0.003, *SE* = 0.002, *p* = 0.129, 95% CI [−0.006, 0.001]). Additionally, *repeated collaboration* displays a statistically significant negative impact (*Est.* = −0.028, *SE* = 0.008, *p* = 0.001, 95% CI [−0.044, −0.012]) while the geographical *dispersion* exhibits a positive effect (*Est.* = 0.033, *SE* = 0.002, *p* = 0.000, 95% CI [0.030, 0.036]). *Papers* (*Est.* = 0.065, *SE* = 0.001, *p* = 0.000, 95% CI [0.062, 0.067]) and *co-authors' citations* (*Est.* = 0.026, *SE* = 0.000, *p* = 0.000, 95% CI [0.026, 0.027]) hold a statistically significant impact on the citation counts of the less productive 25% of the full sample. In terms of spatial homophily, we observe similar results to the full panel data. Specifically, writing with somebody from *the same country* positively affects the dependent variable (*Est.* = 0.102, *SE* = 0.037, *p* = 0.006, 95% CI [0.029, 0.175]), whereas co-publishing with researchers from *the same city* (*Est.* = −0.118, *SE* = 0.052, *p* = 0.024, 95% CI [−0.221, −0.016]) or based in *the same institution* (*Est.* = −0.148, *SE* = 0.064, *p* = 0.021, 95% CI [−0.273, −0.022]) display negative effects. Table 4 reports the negative binomial regression results predicting citation counts for the sub-sample of the most 25% productive researchers of the entire panel. In the full model, *repeated collaboration* (*Est.* = 0.010, *SE* = 0.002, *p* = 0.000, 95% CI [0.006, 0.014]) and *dispersion* (*Est.* = 0.035, *SE* = 0.001, *p* = 0.000, 95% CI [0.033, 0.037]) are revealed to have a statistically significant positive effect on the citation counts. At the same time, contrary to our expectations, the size of the ego-network (the number of *co-authors*) (*Est.* = −0.017, *SE* = 0.001, *p* = 0.000, 95% CI [−0.020, −0.014]) has a negative impact. Among the control variables, paper productivity (*Est.* = 0.016, *SE* = 0.000, *p* = 0.000, 95% CI [0.015, 0.017]) and the prestige of co-authors (*co-authors' citations*) (*Est.* = 0.025, *SE* = 0.000, *p* = 0.000, 95% CI [0.025, 0.026]) hold positive impact on the dependent variable. Co-publishing with researchers based in *the same institution* has a negative effect

(*Est.* = −0.265, *SE* = 0.060, *p* = 0.000, 95% CI [−0.383, −0.148]), while collaborations with researchers from *the same country* (*Est.* = 0.030, *SE* = 0.039, *p* = 0.444, 95% CI [−0.046, 0.106]) and *the same city* (*Est.* = 0.007, *SE* = 0.055, *p* = 0.893, 95% CI [−0.101, 0.116]) are not statistically significant.

For all the negative binomial regression models, the reported scores for variance inflation factor (VIF) are under a threshold of three. This indicates a lack of multi-collinearity between a specific independent variable and the other independent variables in the model (O'brien 2007). The VIF scores are computed using the *summ ()* function available in the *jtools* R package (Long 2019).

Discussion

In this study, we profile the most productive EU based researchers in terms of published papers. A balanced panel of 4588 individuals is longitudinally observed between 2007 and 2018. Data on research productivity and coauthorship networks are collected. We start this section by summarizing the key results with reference to our study objectives. We find that six EU countries (Germany, United Kingdom, Netherlands, Italy, France and Spain) account for 76% of the EU most productive researchers and for 25% of their co-authors. Remarkably, only a third of the authors' collaborators are based in EU countries. This feature unveils the global dimension of these coauthorship networks. The high concentration of the most productive researchers in a limited number of European countries corroborates with other country-level bibliometric measurements of research output. For instance, Germany, United Kingdom and France are among the top ten countries that internationally dominate natural-science research (Nature 2019). Furthermore, United Kingdom, Germany, France, Italy, Spain, and Netherlands are, in this order, the first EU countries in the world rank of indexed research output (both for 2018 and for the 1996–2018 interval); according to *Scimago Journal & Country Rank* (www.scimagojr.com).

Our research is unsuccessful in supporting the positive impact of the ego degree centrality (*Hypothesis 1*). The size of the ego-network is a positive statistically significant predictor only if the prestige of the collaborators is not included as a control variable in the model. This result is similar to other previous findings in the literature. For instance, it was claimed that *g*-index (as a measure of research performance) is positively associated with ego's normalized centrality (Abbasi et al. 2011a), with ego's degree centrality (Abbasi et al. 2012a), and with node degree centrality (Bordons et al. 2015). Other studies underlined that co-authored publications achieve above-average visibility and impact (Abramo and D'Angelo 2015) or individual performance (Ductor 2015). However, we find that, in the presence of the co-authors' citations effect, there is a negative association between the number of alters and the citation counts (the outcome variable). In our full negative binomial regression models, the co-authors' prestige is indicated to have a positive statistically significant contribution to the citation counts. This may substantiate the existence of a selection effect in co-publishing: the tendency to collaborate with prestigious alters. Taken together, it is the prestige (co-authors' citations) and not the number of the collaborators that positively predicts an increase in the citations received by the researchers in the panel. However, this result should be interpreted with caution. In a previous study, it was reported that the partners' degree-centrality can affect ego's performance and productivity (Abbasi et al. 2018). Consequently, it is unclear whether the attributes (citations, in our study) or the structural position of the co-authors positively affect ego's citation counts. Future work is needed to refine this result by controlling for both co-authors' citations and structural characteristics. Moreover, a special attention should be given to the so-called *H-index paradox* according to which the average *H*-index of co-authors is usually higher than ego's *H*-index (Benevenuto et al. 2016).

The evidence from our study gives support to a positive association between repeated collaboration and citations counts (*Hypothesis 2*). Our data illustrate that researchers gain benefits in terms of citations as a result of developing strong ties with their collaborators. This relationship is revealed after controlling for the spatial proximity homophily (same county, same city and same institution co-authors). This result lends support to previous evidence showing that the average ties strength of a researcher has a positive impact on her research performance measured as *g*-index (Abbasi et al. 2011a; Bordons et al. 2015) or by citation counts (Petersen 2015). However, a study on the academic productivity of Nobel laureates' teams stressed that publications produced earlier in a sequence of repeated collaborations with a co-author are cited more compared to publications that come later (Chan et al. 2016).

Our findings also confirm that geographical dispersion (number of countries) of the collaborators positively impacts upon the citation counts (*Hypothesis 3*). According to our data, the geographical dispersion of the authors' networks and the number of international collaborators continuously increased, one year after another. Interestingly, for the same time window (2007–2018), the number of domestic co-authors (from the same country) followed a decreasing slope. Our findings provide support to previous studies that argued top researchers to be inclined to rather engage in international collaborations (Abramo et al. 2019). Also, we confirm previous results showing that international collaboration across six specialities is positively related to impact (Wagner et al. 2017). Additionally, we extend previous findings indicating the preference of scholars from Southern European countries (Fernández et al. 2018) or of researchers in general (Larivière et al. 2015) to develop collaborations over long spatial distances.

Taken all together, our negative binomial regression models highlight that most productive researchers gain benefits in terms of citations if they deploy repeated coauthorship and constantly expand the geographical coverage of their collaboration networks. Also, our results show the positive effects of some of the control variables: paper productivity, co-publication with prestigious collaborators and with alters from the same country (*spatial proximity homophily*). The results observed on the full sample of individual researchers are confirmed by the results observed on the sub-samples. There are nevertheless two exceptions. For the researchers in the first quartile (the lower 25% of the full sample), repeated collaborations have a negative estimate while the number of co-authors is not statistically significant.

These findings, however, need to be treated with care, given that our study may have several limitations. The first limitation refers to measuring research impact (prestige) and productivity by employing proxy variables (i.e., citations and papers, respectively). The second limitation is the focus on ego-networks and, especially, on first-order neighbourhoods (direct coauthorship ties). Therefore, our study does not address the structural aspects related to global networks (whole-networks) and indirect ties. The third limitation lies in the fact that the current study is not specifically designed to take stock of the research fields. We are aware of the fact that different patterns of coauthorship, citations etc. may exist in different fields (Glänzel 2001) or in the course of a scientist's career (Glänzel 2014). However, in accordance with the aim of our paper, we built a very general sample that included the most productive EU based researchers within a specific time window (2007–2018), i.e., the individuals with the largest number of published articles in a 12-year time interval. The examination of the most productive researchers stacked in various specific fields or across individual-level working environments exceeds the scope of this paper. For this reason, we did not perform stratified or quota sampling, or other statistical data selection procedures employing as sampling criteria: research fields, career stages, academic positions etc. Consequently, our findings should be interpreted with great deal of attention under an extremely general framework. It should be clearly noted that our findings might not be extended to the most productive researchers working in a specific research field, environment, or occupying specific academic positions. It may be very helpful for future research to use different research designs and data to validate our results. Furthermore, controlling for research fields could be an excellent step towards increasing the generalisability of this study's findings.

Despite these limitations, our paper however makes several contributions. First, we believe that our results may improve and increase current knowledge on the EU research elite (Kwiek 2016, 2018). We hope our research could support the efforts of the decision-makers to integrate the EU research (Chessa et al. 2013) and to increase inclusion under the institutional framework of the European Research Area. On another level, we expect our findings to inform policy-makers and other interested parts about how network factors affect the impact of research output. As our data show, the coauthorship networks of the most productive EU researchers have geographically expanded and increased in volume. Despite this development, collaboration remains under a geographical bias (the tendency to co-publish with alters from the same country was found statistically significant). For instance, 76% of the research elite are based in a core of six EU countries and only 35% of co-authors are based in EU institutions. Additionally, collaboration is generally driven by the quest of co-writing with prestigious others (in our statistical models, co-authors' citations account for research impact). The highly productive minority of EU researchers seem to be rather densely-knitted by repeated collaborations. At the same time, it is embedded in coauthorship networks that are inherently domestic and globally oriented. Second, our paper contributes to the studies of collaboration networks. It highlights the importance of combining structural and compositional network properties in accounting for citation counts. Also, it makes use of a longitudinal bibliometric balanced panel data (Hâncean et al. 2020) to study the research productivity of individual scientists.

In summary, our work provides insights into the relationship between coauthorship ego-networks and research impact. And, also, it adheres to the currently coalescing efforts of profiling the EU research elite.

Acknowledgements Funding was provided by Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetarii, Dezvoltarii si Inovarii (Grant No. PN-III-P1-1.1-TE-2016-0362), The Slovenian Research Agency (Grant Nos. J1-2457, J1-9112, and P1-0403) and Deutsche Forschungsgemeinschaft (Grant No. LE 2237/2-1).We thank Ms. Laura Trandafir for her help in formatting Fig. 1.

Appendix

See Tables 5 and 6.

Table 5 European Union-based most productive researchers (egos) by country. *Source:* Authors' own calculations

	Country	Frequency	Percent (%)	Cumulative (%)
1	Germany	1111	24.22	24.22
2	United Kingdom	662	14.43	38.64
3	Netherlands	530	11.55	50.20
4	Italy	522	11.38	61.57
5	France	406	8.85	70.42
6	Spain	273	5.95	76.37
7	Belgium	251	5.47	81.84
8	Sweden	146	3.18	85.03
9	Denmark	123	2.68	87.71
10	Finland	123	2.68	90.39
11	Poland	103	2.24	92.63
12	Austria	96	2.09	94.73
13	Portugal	54	1.18	95.90
14	Czech Republic	49	1.07	96.97
15	Greece	49	1.07	98.04
16	Ireland	26	0.57	98.61
17	Slovenia	21	0.46	99.06
18	Hungary	19	0.41	99.48
19	Romania	6	0.13	99.61
20	Cyprus	5	0.11	99.72
21	Luxembourg	4	0.09	99.80
22	Lithuania	3	0.07	99.87
23	Slovakia	3	0.07	99.93
24	Bulgaria	1	0.02	99.96
25	Croatia	1	0.02	99.98
26	Estonia	1	0.02	100.00
	Total	4588	100.00	100.00

The 4588 most productive researchers account for 0.03% of all the Scopus accounts

Table 6 The European Union-based coauthors (alters) by country. *Source:* Authors' own calculations

	Country	Number of coauthors	Share (%)	Cumulative percentage (%)
1	Germany	85,473	7.32	7.32
2	United Kingdom	57,576	4.93	12.25
3	Italy	49,138	4.21	16.45
4	France	42,749	3.66	20.11
5	Netherlands	33,400	2.86	22.97
6	Spain	26,045	2.23	25.20
7	Belgium	16,762	1.43	26.64
8	Sweden	13,825	1.18	27.82
9	Denmark	11,087	0.95	28.77
10	Poland	10,916	0.93	29.70
11	Finland	10,354	0.89	30.59
12	Austria	10,212	0.87	31.46
13	Greece	8903	0.76	32.23
14	Portugal	7709	0.66	32.89
15	Czech Republic	7662	0.66	33.54
16	Hungary	4264	0.37	33.91
17	Ireland	3996	0.34	34.25
18	Slovenia	2108	0.18	34.43
19	Slovakia	1309	0.11	34.54
20	Croatia	1176	0.10	34.64
21	Lithuania	1093	0.09	34.74
22	Bulgaria	934	0.08	34.82
23	Estonia	921	0.08	34.90
24	Romania	704	0.06	34.96
25	Cyprus	469	0.04	35.00
26	Iceland	398	0.03	35.03
27	Latvia	383	0.03	35.06
28	Malta	153	0.01	35.08
29	Other countries	758,365	64.92	100.00
	Total	1,168,084	100.00	100.00

References

Abbasi, A. (2013). *h*-Type hybrid centrality measures for weighted networks. *Scientometrics, 96*(2), 633–640. https://doi.org/10.1007/s11192-013-0959-y.

Abbasi, A., Altmann, J., & Hossain, L. (2011a). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics, 5*(4), 594–607. https://doi.org/10.1016/j.joi.2011.05.007.

Abbasi, A., Altmann, J., & Hwang, J. (2010). Evaluating scholars based on their academic collaboration activities: Two indices, the RC-index and the CC-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics, 83*(1), 1–13. https://doi.org/10.1007/s11192-009-0139-2.

Abbasi, A., Chung, K. S. K., & Hossain, L. (2012a). Egocentric analysis of co-authorship network structure, position and performance. *Information Processing and Management, 48*(4), 671–679. https://doi.org/10.1016/j.ipm.2011.09.001.

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012b). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics, 6*(3), 403–412. https://doi.org/10.1016/j.joi.2012.01.002.

Abbasi, A., Hossain, L., Uddin, S., & Rasmussen, K. J. R. (2011b). Evolutionary dynamics of scientific collaboration networks: Multi-levels and cross-time analysis. *Scientometrics, 89*(2), 687–710. https://doi.org/10.1007/s11192-011-0463-1.

Abbasi, A., & Jaafari, A. (2013). Research impact and scholars' geographical diversity. *Journal of Informetrics, 7*(3), 683–692. https://doi.org/10.1016/j.joi.2013.04.004.

Abbasi, A., Jalili, M., & Sadeghi-Niaraki, A. (2018). Influence of network-based structural and power diversity on research performance. *Scientometrics, 117*(1), 579–590. https://doi.org/10.1007/s11192-018-2879-3.

Abramo, G., & D'Angelo, C. A. (2015). The relationship between the number of authors of a publication, its citations and the impact factor of the publishing journal: Evidence from Italy. *Journal of Informetrics, 9*(4), 746–761. https://doi.org/10.1016/j.joi.2015.07.003.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2019). The collaboration behavior of top scientists. *Scientometrics, 118*(1), 215–232. https://doi.org/10.1007/s11192-018-2970-9.

Andrikopoulos, A., Bekiaris, M., & Kostaris, K. (2020). Stars in a small world: Social networks in auditing research. *Scientometrics, 122*(1), 625–643. https://doi.org/10.1007/s11192-019-03272-z.

Antoniou, G. A., Antoniou, S. A., Georgakarakos, E. I., Sfyroeras, G. S., & Georgiadis, G. S. (2015). Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature. *Annals of Vascular Surgery, 29*(2), 286–292. https://doi.org/10.1016/j.avsg.2014.09.017.

Badar, K., Frantz, T. L., & Jabeen, M. (2016). Research performance and degree centrality in co-authorship networks: The moderating role of homophily. *Aslib Journal of Information Management, 68*(6), 756–771. https://doi.org/10.1108/AJIM-07-2016-0103.

Badar, K., Hite, J. M., & Ashraf, N. (2015). Knowledge network centrality, formal rank and research performance: Evidence for curvilinear and interaction effects. *Scientometrics, 105*(3), 1553–1576. https://doi.org/10.1007/s11192-015-1652-0.

Badar, K., Hite, J. M., & Badir, Y. F. (2013). Examining the relationship of co-authorship network centrality and gender on academic research performance: the case of chemistry researchers in Pakistan. *Scientometrics, 94*(2), 755–775. https://doi.org/10.1007/s11192-012-0764-z.

Badar, K., Hite, J. M., & Badir, Y. F. (2014). The moderating roles of academic age and institutional sector on the relationship between co-authorship network centrality and academic research performance. *Aslib Journal of Information Management, 66*(1), 38–53. https://doi.org/10.1108/AJIM-05-2013-0040.

Barabási, A. L., Jeong, H., Néda, 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. https://doi.org/10.1016/S0378-4371(02)00736-7.

Benevenuto, F., Laender, A. H. F., & Alves, B. L. (2016). The *H*-index paradox: Your coauthors have a higher *H*-index than you do. *Scientometrics, 106*(1), 469–474. https://doi.org/10.1007/s11192-015-1776-2.

Biscaro, C., & Giupponi, C. (2014). Co-authorship and bibliographic coupling network effects on citations. *PLoS ONE, 9*(6), e99502. https://doi.org/10.1371/journal.pone.0099502.

Bordons, M., Aparicio, J., González-Albo, B., & Díaz-Faes, A. A. (2015). The relationship between the research performance of scientists and their position in co-authorship networks in three fields. *Journal of Informetrics, 9*(1), 135–144. https://doi.org/10.1016/j.joi.2014.12.001.

Borjas, G. J., & Doran, K. B. (2015). Which peers matter? The relative impacts of collaborators, colleagues, and competitors. *Review of Economics and Statistics, 97*(5), 1104–1117. https://doi.org/10.1162/REST_a_00472.

Bornmann, L., & Daniel, H.-D. (2007). Multiple publication on a single research study: Does it pay? The influence of number of research articles on total citation counts in biomedicine. *Journal of the American Society for Information Science and Technology, 58*(8), 1100–1107. https://doi.org/10.1002/asi.20531.

Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics, 6*(1), 11–18. https://doi.org/10.1016/j.joi.2011.08.004.

Boschini, A., & Sjögren, A. (2007). Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics, 25*(2), 325–365. https://doi.org/10.1086/510764.

Brass, D. J., Galaskiewicz, J., Greve, H. R., & Tsai, W. (2004). Taking stock of networks and organizations: A multilevel perspective. *Academy of Management Journal, 47*(6), 795–817. https://doi.org/10.5465/20159624.

Chan, H. F., Önder, A. S., & Torgler, B. (2016). The first cut is the deepest: Repeated interactions of coauthorship and academic productivity in Nobel laureate teams. *Scientometrics, 106*(2), 509–524. https://doi.org/10.1007/s11192-015-1796-y.

Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., & Riccaboni, M. (2013). Is Europe evolving toward an integrated research area? *Science, 339*(6120), 650–651. https://doi.org/10.1126/science.1227970.

Collet, F., Robertson, D. A., & Lup, D. (2014). When does brokerage matter? Citation impact of research teams in an emerging academic field. *Strategic Organization, 12*(3), 157–179. https://doi.org/10.1177/1476127014530124.

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703–722. https://doi.org/10.1177/0306312705055535.

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics, 5*(1), 187–203. https://doi.org/10.1016/j.joi.2010.10.008.

Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics, 77*(3), 385–407. https://doi.org/10.1111/obes.12070.

Fernández, A., Ferrándiz, E., & León, M. D. (2018). Patterns of academic scientific collaboration at a distance: Evidence from Southern European countries. In M. Jibu & Y. Osabe (Eds.), *Scientometrics*. Rijeka: InTech. https://doi.org/10.5772/intechopen.77370.

Freeman, R. B., & Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics, 33*(S1), S289–S318. https://doi.org/10.1086/678973.

Friemel, T. N. (2015). Opinion leadership| influence versus selection: A network perspective on opinion leadership. *International Journal of Communication, 9,* 1002–1022.

Gallotti, R., & De Domenico, M. (2019). Effects of homophily and academic reputation in the nomination and selection of Nobel laureates. *Scientific Reports, 9*(1), 17304. https://doi.org/10.1038/s41598-019-53657-6.

Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: A case study of Harvard University's publications. *Scientometrics, 87*(2), 251–265. https://doi.org/10.1007/s11192-011-0343-8.

Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology, 63*(2), 323–335. https://doi.org/10.1002/asi.21688.

Glänzel, W. (2001). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends, 50*(3), 461–473.

Glänzel, W. (2014). Analysis of co-authorship patterns at the individual level. *Transinformação, 26*(3), 229–238. https://doi.org/10.1590/0103-3786201400030001.

González-Alcaide, G., Pinargote, H., & Ramos, J. M. (2020). From cut-points to key players in co-authorship networks: A case study in ventilator-associated pneumonia research. *Scientometrics, 123*(2), 707–733. https://doi.org/10.1007/s11192-020-03404-w.

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

Guan, J., Yan, Y., & Zhang, J. (2015a). How do collaborative features affect scientific output? Evidences from wind power field. *Scientometrics, 102*(1), 333–355. https://doi.org/10.1007/s11192-014-1311-x.

Guan, J., Zhang, J., & Yan, Y. (2015b). The impact of multilevel networks on innovation. *Research Policy, 44*(3), 545–559. https://doi.org/10.1016/j.respol.2014.12.007.

Guan, J. C., Zuo, K., Chen, K., & Yam, R. C. M. (2016). Does country-level R&D efficiency benefit from the collaboration network structure? *Research Policy, 45*(4), 770–784. https://doi.org/10.1016/j.respol.2016.01.003.

Hâncean, M.-G., & Perc, M. (2016). Homophily in coauthorship networks of East European sociologists. *Scientific Reports, 6,* 36152. https://doi.org/10.1038/srep36152.

Hâncean, M.-G., Perc, M., & Lerner, J. (2020). Data from: The coauthorship networks of the most productive European researchers. *Zenodo*. https://doi.org/10.5281/zenodo.3873772.

Hâncean, M.-G., Perc, M., & Vlăsceanu, L. (2014). Fragmented Romanian sociology: Growth and structure of the collaboration network. *PLoS ONE, 9*(11), e113271. https://doi.org/10.1371/journal.pone.0113271.

Hausman, J., Hall, B., & Griliches, Z. (1984). *Econometric models for count data with an application to the Patents-R&D relationship (No. t0017) (p. t0017).* Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/t0017.

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge, New York: Cambridge University Press.
Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy, 39*(5), 662–673. https://doi.org/10.1016/j.respol.2010.01.012.
Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in scientometrics. *Scientometrics, 75*(2), 189–202. https://doi.org/10.1007/s11192-007-1771-3.
Jeong, H., Néda, Z., & Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters (EPL), 61*(4), 567–572. https://doi.org/10.1209/epl/i2003-00166-9.
Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1.
Kwiek, M. (2016). The European research elite: A cross-national study of highly productive academics in 11 countries. *Higher Education, 71*(3), 379–397. https://doi.org/10.1007/s10734-015-9910-x.
Kwiek, M. (2018). High research productivity in vertically undifferentiated higher education systems: Who are the top performers? *Scientometrics, 115*(1), 415–462. https://doi.org/10.1007/s11192-018-2644-7.
Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology, 66*(7), 1323–1332. https://doi.org/10.1002/asi.23266.
Levitt, J. M., & Thelwall, M. (2011). A combined bibliometric indicator to predict article impact. *Information Processing and Management, 47*(2), 300–308. https://doi.org/10.1016/j.ipm.2010.09.005.
Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy, 42*(9), 1515–1530. https://doi.org/10.1016/j.respol.2013.06.012.
Liao, C. H. (2011). How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics, 86*(3), 747–761. https://doi.org/10.1007/s11192-010-0309-2.
Long, J. A. (2019). *jtools: Analysis and presentation of social scientific data.* R package version 2.0.1. Retrieved May 31, 2020, from https://cran.r-project.org/package=jtools.
Martín-Alcázar, F., Ruiz-Martínez, M., & Sánchez-Gardey, G. (2019). Assessing social capital in academic research teams: A measurement instrument proposal. *Scientometrics, 121*(2), 917–935. https://doi.org/10.1007/s11192-019-03212-x.
McCarty, C., Jawitz, J. W., Hopkins, A., & Goldman, A. (2013). Predicting author *h*-index using characteristics of the co-author network. *Scientometrics, 96*(2), 467–483. https://doi.org/10.1007/s11192-012-0933-0.
Medina, A. M. (2018). Why do ecologists search for co-authorships? Patterns of co-authorship networks in ecology (1977–2016). *Scientometrics, 116*(3), 1853–1865. https://doi.org/10.1007/s11192-018-2835-2.
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 cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213–238. https://doi.org/10.1177/000312240406900204.
Nature. (2019). The top 10 countries that dominate natural-sciences research, *570*. https://doi.org/10.1038/d41586-019-01921-0.
Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E, 64*(2), 025102. https://doi.org/10.1103/PhysRevE.64.025102.
Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5200. https://doi.org/10.1073/pnas.0307545100.
O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6.
Parreira, M. R., Machado, K. B., Logares, R., Diniz-Filho, J. A. F., & Nabout, J. C. (2017). The roles of geographic distance and socioeconomic factors on international collaboration among ecologists. *Scientometrics, 113*(3), 1539–1550. https://doi.org/10.1007/s11192-017-2502-z.
Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics, 4*(4), 475–482. https://doi.org/10.1016/j.joi.2010.04.003.
Perc, M. (2014). The Matthew effect in empirical data. *Journal of the Royal Society Interface, 11*(98), 20140378. https://doi.org/10.1098/rsif.2014.0378.
Petersen, A. M. (2015). Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences, 112*(34), E4671–E4680. https://doi.org/10.1073/pnas.1501444112.

Puuska, H.-M., Muhonen, R., & Leino, Y. (2014). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics, 98*(2), 823–839. https://doi.org/10.1007/s11192-013-1181-7.
Ronda-Pupo, G. A., & Pham, T. (2018). The evolutions of the rich get richer and the fit get richer phenomena in scholarly networks: The case of the strategic management journal. *Scientometrics, 116*(1), 363–383. https://doi.org/10.1007/s11192-018-2761-3.
Rotolo, D., & Messeni Petruzzelli, A. (2013). When does centrality matter? Scientific productivity and the moderating role of research specialization and cross-community ties. *Journal of Organizational Behavior, 34*(5), 648–670. https://doi.org/10.1002/job.1822.
Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science, 3*(1), 9. https://doi.org/10.1140/epjds/s13688-014-0009-x.
Scarazzati, S., & Wang, L. (2019). The effect of collaborations on scientific research output: The case of nanoscience in Chinese regions. *Scientometrics, 121*(2), 839–868. https://doi.org/10.1007/s11192-019-03220-x.
Shoukri, M. M. (2018). *Statistical analysis of health data using SAS and R* (4th ed.). Boca Raton: CRC Press.
Sidone, O. J. G., Haddad, E. A., & Mena-Chalco, J. P. (2017). Scholarly publication and collaboration in Brazil: The role of geography. *Journal of the Association for Information Science and Technology, 68*(1), 243–258. https://doi.org/10.1002/asi.23635.
Sugimoto, C. R., Robinson-Garcia, N., Murray, D. S., Yegros-Yegros, A., Costas, R., & Larivière, V. (2017). Scientists have most impact when they're free to move. *Nature, 550*(7674), 29–31. https://doi.org/10.1038/550029a.
Sun, L., & Rahwan, I. (2017). Coauthorship network in transportation research. *Transportation Research Part A: Policy and Practice, 100,* 135–151. https://doi.org/10.1016/j.tra.2017.04.011.
Tu, J. (2019). What connections lead to good scientific performance? *Scientometrics, 118*(2), 587–604. https://doi.org/10.1007/s11192-018-02997-7.
Uddin, S., Choudhury, N., & Hossain, M. E. (2019). A research framework to explore knowledge evolution and scholarly quantification of collaborative research. *Scientometrics, 119*(2), 789–803. https://doi.org/10.1007/s11192-019-03057-4.
Venables, W. N., & Ripley, B. D. (2010). *Modern applied statistics with S* (4 ed., [Nachdr.].). New York: Springer.
Wagner, C. S., Whetsell, T. A., & Leydesdorff, L. (2017). Growth of international collaboration in science: Revisiting six specialties. *Scientometrics, 110*(3), 1633–1652. https://doi.org/10.1007/s11192-016-2230-9.
Wang, J. (2016). Knowledge creation in collaboration networks: Effects of tie configuration. *Research Policy, 45*(1), 68–80. https://doi.org/10.1016/j.respol.2015.09.003.
Wang, L., Thijs, B., & Glänzel, W. (2015). Characteristics of international collaboration in sport sciences publications and its influence on citation impact. *Scientometrics, 105*(2), 843–862. https://doi.org/10.1007/s11192-015-1735-y.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, New York: Cambridge University Press.
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.
Yin, Z., & Zhi, Q. (2017). Dancing with the academic elite: A promotion or hindrance of research production? *Scientometrics, 110*(1), 17–41. https://doi.org/10.1007/s11192-016-2151-7.
Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology, 69*(1), 72–86. https://doi.org/10.1002/asi.23916.