Journal of Informetrics 11 (2017) 371–388

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Research impact in co-authorship networks: a two-mode analysis

Tahereh Dehdarirad [a], Stefano Nasini [b,*]

[a] University of Barcelona, Barcelona, Spain
[b] IESEG School of Management (LEM CNRS 9221), Lille/Paris, France

ARTICLE INFO

Article history:
Received 14 September 2016
Received in revised form 5 February 2017
Accepted 5 February 2017
Available online 28 February 2017

Keywords:
Bibliometrics
Co-authorship networks
Two-mode networks
Assortativity
Regression modeling

ABSTRACT

In the context of research collaboration and co-authorship, we studied scholars' scientific achievements and success, based on their collection of shared publications. By means of a novel regression model, which exploits the two-mode structure of co-authorship, we translated *paper scientific impact* into *author professional achievement*, to simultaneously account for the effect of paper properties (access status, funding bodies, etc.) as well as author demographic and behavioral characteristics (gender, nationality) on academic success and impact. After a detailed analysis of the proposed statistical procedure, we illustrated our approach with an empirical analysis of a co-authorship network based on 1007 scientific articles.

1. Introduction

The high specialization of scientific research, the interdisciplinary character of most projects and the increased funding of cross-institutional initiatives have made researchers take part in scientific collaboration (Haeussler & Sauermann, 2013; Teixeira, 2011; Wuchty, Jones, & Uzzi, 2007).

Co-authorship networks are among the most tangible forms of collaboration structures (De Stefano, Giordano, & Vitale, 2011; Glänzel & Schubert, 2005; Liu & Xia, 2015; Newman, 2004). They can be seen as two-mode networks, where two types of "nodes" (authors and papers) are connected. Studies on co-authorship networks have mostly focused on structural characteristics, both at the global network level (Newman, 2004) and local node level (Leem & Chun, 2015; Uddin, Hossain, & Rasmussen, 2013). The first one considers global properties of the network, such as the density, the transitivity, or the average path length (Castro & Nasini, 2015). In contrast, the local node level focuses on the analysis of individual nodes and their local neighbors.

To put it into context, local properties of *papers* and *authors* serve as indicators of research quality and can jointly be used to account for the *paper scientific impact* and the *author professional achievement* in the two-mode network of shared publications. In fact, the fundamental dynamics of modern research communities is based on the periodic generation of papers by joint groups of authors (Li, Liao, & Yen, 2013; Ortega, 2014; Yan & Ding, 2009). The research impact or scientific outcome of a paper is reflected by the quality (rank and prestige) of the journal it is published in and by the number of citations it receives after publication. From a co-authorship viewpoint, such a scientific outcome directly translates into an achievement for each author and provides a way to measure authors' scientific productivity and research quality. This

* Corresponding author.
E-mail addresses: tdehdari@gmail.com (T. Dehdarirad), s.nasini@ieseg.fr (S. Nasini).

http://dx.doi.org/10.1016/j.joi.2017.02.002
1751-1577/

duality between *paper scientific impact* and *author professional achievement* encodes the very fundamental dynamics of modern research communities.

Various metrics have been proposed to evaluate and predict research quality at both author and paper levels. A widely used measure of scientific impact at paper level is citations, with a growing body of literature trying to discover their determinant factors (Annalingam, Damayanthi, Jayawardena, & Ranasinghe, 2014; Bornmann & Daniel, 2008; Didegah & Thelwall, 2013; Dong, Johnson, & Chawla, 2015; Tahamtan, Safipour Afshar, & Ahamdzadeh, 2016). Some of the factors that have been mentioned in these studies are: reputation of the author and the research group (Leimu & Koricheva, 2005; Yu, Gu, Zhou, & Han, 2012), the publication venue (conference/journal) (Larivière & Gingras, 2010; Yu et al., 2012), language of the paper (Diekhoff, Schlattmann, & Dewey, 2013; Hurley, Ogier, & Torvik, 2013), the number of co-authors (Vanclay, 2013; Vieira & Gomes, 2010), professional age (Hurley et al., 2013), gender (Lariviere, Ni, Gingras, Cronin, & Sugimoto, 2013; Rigg, McCarragher, & Krmenec, 2012), international collaboration (Low, Ng, Kabir, Koh, & Sinnasamy, 2014; Sooryamoorthy, 2009; Tahamtan et al., 2016) and open access (Eysenbach, 2006; McCabe & Snyder, 2014), amongst others.

Journal impact factor (IF) is another recognized measure of scientific impact, which has acquired a major role in the evaluations of the output of scholars, departments and whole institutions. Typically, papers appearing in journals with a higher IF, receive more citations (Didegah & Thelwall, 2013; van der Pol, McInnes, Petrcich, Tunis, & Hanna, 2015; van Eck, Waltman, van Raan, Klautz, & Peul, 2013) and a high weight in such evaluations (Pan & Fortunato, 2014).

At author level, in the literature, there are several metrics of individual impact, with the *h*-index being the most popular measure by far. The main advantage of the *h*-index is that it combines a measure of quantity and impact in a single indicator (Costas & Bordons, 2007). However, a list of disadvantages of this index have been pointed out by Glänzel (2006), Bornmann and Daniel (2007b) and Jin, Liang, Rousseau, and Egghe (2007). Author Impact Factor (AIF) is another proxy of the authors' professional achievements, which is the extension of the IF to authors (Pan & Fortunato, 2014; Petersen et al., 2014).

Both author and paper level metrics play an important role in how individuals, research groups, journals, academic departments, institutions and countries are evaluated and ranked (Ding, Rousseau, & Wolfram, 2014). These bibliometric indicators are also used as one criterion in the evaluation of grant proposals and research institutes or in hiring committees for faculty positions.

The goal of this work is to provide a unified approach for the analysis of research impact and quality in co-authorship structures. By means of a novel regression model, which exploits the two-mode network of co-authorship, we translate *paper scientific impact* (number of citations and rank of the corresponding journal) into *author professional achievement* (average number of citations and average rank of the corresponding journals). This provides a comprehensive statistical approach that uses paper and author information in a predictive bibliometric analysis. The underlying assumption is that the expected impact of a paper depends not only on the characteristics of a paper (such as access status, funding bodies, etc.), but also on the characteristics of the authors (such as genders and nationalities). Similarly, the expected professional achievement of the author depends on their demographic and behavioral properties, as well as on the characteristics of the papers they coauthor, projected into the author dimension through the two-mode structure. This *two-mode regression* allows to simultaneously account for the effect of paper properties as well as author demographic and behavioral characteristics on academic impact and success.

From a statistical viewpoint, the proposed methodology can be included in the list of regression-like modeling approaches for networked data, i.e. regression models which internalize the cross-section dependencies between statistical unites based on a specified structure of network proximity. In this respect, an early settlement of the problem has been addressed by Doreian (1982), who presented it as a generalization of the spatial effects linear model and the linear model with spatial error terms. In the same year Dow, Burton, and White (1982) provided a simulation study of the consequences of ignoring the network autocorrelated disturbance. Robins, Pattison, and Elliott (2001) designed one of the first modeling approaches in the context of the Exponential Random Graph Models for individual properties and network structure, which has been recently extended by Nasini, de Albeniz, and Dehdarirad (2017). Likewise, Giordano and Vitale (2011) and Giordano and Scepi (2012) proposed a statistical framework in the context of Conjoint Analysis which allows the inclusion of social network data as relational constraints. As an alternative made necessary by our specific data setting, the two-mode regression is based on the projected information of both layers into a common author and paper dimension, rather than on the inclusion of network autocorrelated error terms.

After a detailed analysis of the proposed statistical methods, we illustrate our approach with an empirical study of the co-authorship network based on 1007 scientific articles.

The paper is organized as follows. Section 2 introduces the statistical methodology used for the analysis. The co-authorship data set is described in Section 3. The empirical results, corresponding to the descriptive and the predictive analysis, are presented in Section 4. Section 5 provides a detailed discussion about the bibliometrics insights and conclusions. All the mathematical proofs of propositions are reported in Appendix A.

2. Methodology

The statistical methodology for the analysis of the research impact in co-authorship structures is presented in this section.

Let $\mathcal{P}$ be a set of papers with $m = |\mathcal{P}|$ and $\mathcal{A}$ a set of authors with $n = |\mathcal{A}|$. A two-mode network is defined as a set of connections between a primary and a secondary layer, $\mathcal{P}$ and $\mathcal{A}$, which we also refer to as individual and item dimension, respectively. These connections are represented in matrix form $W \in \{0, 1\}^{n \times m}$. There are different approaches to project one layer into the other to obtain both one-mode structures: a paper–paper network (articles sharing authors), which is associated to the set of connections between papers, $\mathcal{E}_P \subseteq \mathcal{P} \times \mathcal{P}$, and an author–author network (co-authorship), which is associated to a set of connections between authors, $\mathcal{E}_A \subseteq \mathcal{A} \times \mathcal{A}$. Let $D_P \in \mathbb{R}^{m\times m}$ and $D_A \in \mathbb{R}^{n\times n}$ be the diagonal matrices whose components correspond to the number of authors per paper and the number of papers per author, respectively. We define $W_A = D_A^{-1}W$, $W_P = WD_P^{-1}$ and both projections of the two-mode structure into the one-mode paper–paper network $W_{AP} = W_A W_P$ and author–author network $W_{PA} = W_P^{\mathrm{T}} W_A$. The (i, j) component of W_{AP} is a real number between zero and one, denoting the strength of the relation between paper i and j (it is equal to one when i and j have the same authors and zero when both articles share no author). The (i, j) component of W_{PA} can be interpreted in the same way. Additional one–mode structures which can be derived from W are $W_{AA} = W_A^{\mathrm{T}} W_A$ and $W_{PP} = W_P^{\mathrm{T}} W_P$. Matrices W_{AP} and W_{PA} are row-stochastic, whereas matrices W_{AA} and W_{PP} are not.

Finally, for each node in either layer, the following collections of variables are defined: $\mathcal{O}$ is the set of nodal outcomes (in our application $\mathcal{O}$ = {citations, impact factor}); $\mathcal{K}^P$ is the set of paper characteristics (access status, funding bodies, etc.); $\mathcal{K}^A$ is the set of author characteristics (gender, nationality, etc.).[1] In the rest of the paper, the collection of nodal characteristics $\mathcal{K}_P$ and $\mathcal{K}_A$ are considered as exogenous co-variates, whereas the observed nodal outcomes $\mathcal{O}$ are regarded as endogenous response variables in our modeling framework. Precisely, for $i \in \mathcal{P}$ and $o \in \mathcal{O}$, variable $y_i^{(P,o} \in \mathbb{R}$ denotes the oth outcome of the ith item. Similarly, the quantity $x_{ih}^P \in \mathbb{R}$ corresponds to the hth nodal characteristics of the ith item. The vector notation $\mathbf{y}^{(P,o} = [y_1^{(P,o} \ldots y_m^{(P,o}]$ and $\mathbf{x}_h^P = [x_{1h}^P \ldots x_{mh}^P]$ is used, for each $h \in \mathcal{K}^P$. Based on the two-mode structure, item outcomes $\mathbf{y}^{(P,o}$ can be projected into the secondary layer, resulting in the corresponding individual outcomes $\mathbf{y}^{(A,o}$, for each $o \in \mathcal{O}$, i.e. $\mathbf{y}^{(A,o} = (D_A^{-1}W)\mathbf{y}^{(P,o}$. In other words, the scientific impact in the individual dimension is computed by averaging the corresponding impact in the item dimension. Thus, authors' professional achievements are measured in terms of the average citations and the average impact factors of the corresponding journals, allowing to directly translate paper citations and impact factors into authors average citations and average impact factors.[2]

Two levels of statistical analysis of the research impact in co-authorship structures are presented in this section.

2.1. Descriptive analysis: assortative patterns in co-authorship

This level of analysis provides an exploratory description of the associations between nodal properties (either paper and author characteristics X_P and X_A or paper and author scientific outputs $\mathbf{y}^{(P,o}$ and $\mathbf{y}^{(A,o}$) in both structures of connections $\mathcal{E}_P$ and $\mathcal{E}_A$. Network assortativity was first studied by Newman (2001), who proposed a measure which describes how nodes tend to connect with similar others. Generally, the assortativity of a network is determined in terms of the degree (number of direct neighbors) of nodes. However, it may be extended to other nodal characteristics, such as the measures of transitivity and centrality.[3]

In our context, the interest is in the assortative pattern, given by the association between paper and author characteristics, X_P and X_A on the one hand, and paper and author scientific outputs, $\mathbf{y}^{(P,o}$ and $\mathbf{y}^{(A,o}$, on the other. Thus, our interest is in assortativity indexes of the form

$$\text{Assortativity} \equiv \text{Sim}(\mathbf{z}, A\mathbf{z}) \in \mathbb{R}^q$$

for specified vector of properties $\mathbf{z} \in \mathbb{R}^q$, an adjacency matrix $A \in \mathbb{R}^{q\times q}$, and a given specification of the similarity measure Sim( ). Different similarities measures can be considered, such as the absolute difference, Pearson-correlation coefficient, or the inner-product. When multiple variables are taken into account a matrix of cross-similarities can be defined, based on the inner-product specification. The following are *inner-product matrix similarities* between author characteristics and paper characteristics respectively:

$$\text{Sim}(X_P, W_{PP}X_P) = X_P^{\mathrm{T}} W_{PP} X_P \in \mathbb{R}^{m\times m} \quad \text{and} \quad \text{Sim}(X_A, W_{AA}X_A) = X_A^{\mathrm{T}} W_{AA} X_A \in \mathbb{R}^{n\times n}.$$

A global assortativity in the paper–paper network and author–author network is obtained as the traces of $\text{Sim}(X_P, W_{PP}X_P)$ and $\text{Sim}(X_A, W_{AA}X_A)$ respectively.

Sometimes, assortativity measures are calculated with respect to network properties, such as nodal centralities and transitivities. In the co-authorship context, a particular attention is given to the role played by network centrality in both

[1] These paper and author properties have been selected based both on their relevance in the recent literature concerning their effect on citations (see Section 5) and on an exploratory statistical analysis which have been preliminarily conducted to enlighten their association and impact on paper and author outcomes.

[2] Due to its intrinsic dynamic character, the average impact factor is capable to capture trends and variations of the impact of the scientific output of scholars in time, unlike the h-index, which is a growing measure taking into account the whole career path. Furthermore, cumulative measures such as the h-index consider only the best publications and do not penalize low quality work. Given the above-mentioned advantages of average impact factor over h-index and the disadvantages and inconsistencies in the way in which scientists are ranked by the h-index as indicated in as also indicated in Glänzel (2006), Bornmann and Daniel (2007b), Jin et al. (2007) and Waltman and Van Eck (2012), in this paper the average impact factor was preferred over the h-index.

[3] Recently, there have been different attempts to address alternative assortativity measures (Abramson et al., 2014; Piraveenan, Prokopenko, & Zomaya, 2008), which aim to quantify the assortative mixing for individual nodes at a local level.

structures of connections $\mathcal{E}_P$ and $\mathcal{E}_A$. Centrality measures are typically adopted to identify important nodes within a network, based on different definition of the term *importance*. We focus on four measures – degree, eigenvector, betweenness and closeness[4] – and study their pattern of association with other characteristics. The joint bibliometric interpretation of these analysis provides a deeper understanding about how the distribution network centrality translates into paper and author scientific prominence.

2.2. Predictive analysis: two-mode regression

The second level of analysis requires the use of two bivariate linear models, one for each network layer. The idea is to combine the *explanatory power* of both item and individual characteristics, $\mathcal{K}_P$ and $\mathcal{K}_A$, when modeling the scientific outcome of either dimensions. These properties were chosen because they have been significant and effective determinants of citations in many previous studies, as indicated in Tahamtan et al. (2016) comprehensive review of factors affecting number of citations.

Precisely, this regression approach aims to explain scientific impact of either dimensions in their probabilistic relations with item and individual characteristics, i.e.

$$\mathbb{E}[\mathbf{y}^{(P}] = {\boldsymbol{\beta}_P}^{\mathrm{T}} X_P + W_A({\boldsymbol{\beta}_A}^{\mathrm{T}} X_A) \quad \text{and} \quad \mathbb{E}[\mathbf{y}^{(A}] = {\boldsymbol{\beta}_A}^{\mathrm{T}} X_A + W_P^{\mathrm{T}}({\boldsymbol{\beta}_P}^{\mathrm{T}} X_P), \tag{1}$$

where $\mathbf{y}^{(P}$, $\mathbf{y}^{(A}$ are vectors of papers and authors overall outcomes (citations and impact factor), $\mathbb{E}[\,\cdot\,]$ is the expectation operator, and $\boldsymbol{\beta}_P$ and $\boldsymbol{\beta}_A$ are corresponding regression coefficients.

Both expectations rely on paper and author properties through their two-mode projection, W. In the first case, the assumption is that the expected paper scientific impact $\mathbb{E}\left[\mathbf{y}^{(P}\right]$ depends not only on papers characteristics (such as the access status and the number of authors), i.e. ${\boldsymbol{\beta}_P}^{\mathrm{T}} X_P$, but also on characteristics of the their corresponding authors (such as genders and nationalities), projected into the paper dimension through a two-mode structure, i.e. $W_A({\boldsymbol{\beta}_A}^{\mathrm{T}} X_A)$. Similarly, in the second case, the expected author professional achievement $\mathbb{E}\left[\mathbf{y}^{(P}\right]$ depends on their demographic and behavioral properties, ${\boldsymbol{\beta}_A}^{\mathrm{T}} X_A$, as well as on the characteristics of the papers they coauthor, projected into the author dimension through the two-mode structure, i.e. $W_P^{\mathrm{T}}({\boldsymbol{\beta}_P}^{\mathrm{T}} X_P)$.

Our estimates of $\boldsymbol{\beta}_P$ and $\boldsymbol{\beta}_A$ are based on minimizing the sum of squared residuals, resulting in the following error functions associated to the paper and author level regressions respectively[5]:

$$e^{(P}(\boldsymbol{\beta}_P, \boldsymbol{\beta}_A) = (\mathbf{y}^{(P} - {\boldsymbol{\beta}_P}^{\mathrm{T}} X_P - W_A({\boldsymbol{\beta}_A}^{\mathrm{T}} X_A))^{\mathrm{T}} (\mathbf{y}^{(P} - {\boldsymbol{\beta}_P}^{\mathrm{T}} X_P - W_A({\boldsymbol{\beta}_A}^{\mathrm{T}} X_A))$$

$$e^{(A}(\boldsymbol{\beta}_P, \boldsymbol{\beta}_A) = (\mathbf{y}^{(A} - {\boldsymbol{\beta}_A}^{\mathrm{T}} X_A - W_P^{\mathrm{T}}({\boldsymbol{\beta}_P}^{\mathrm{T}} X_P))^{\mathrm{T}} (\mathbf{y}^{(A} - {\boldsymbol{\beta}_A}^{\mathrm{T}} X_A - W_P^{\mathrm{T}}({\boldsymbol{\beta}_P}^{\mathrm{T}} X_P))$$

where $\mathbf{y}^{(P}$, $\mathbf{y}^{(A}$, X_P and X_A are treated as observed response variables and co-variates respectively. The first order conditions entails

$$\begin{bmatrix} \boldsymbol{\beta}_P \\ \boldsymbol{\beta}_A \end{bmatrix} = \begin{bmatrix} (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}} (\mathbf{y}^{(P} - W_A X_A \boldsymbol{\beta}_A) \\ (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}} (\mathbf{y}^{(P} - X_P \boldsymbol{\beta}_P) \end{bmatrix} \tag{2}$$

and

$$\begin{bmatrix} \boldsymbol{\beta}_P \\ \boldsymbol{\beta}_A \end{bmatrix} = \begin{bmatrix} (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}} (\mathbf{y}^{(A} - X_A \boldsymbol{\beta}_A) \\ (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}} (\mathbf{y}^{(A} - W_P X_P \boldsymbol{\beta}_P) \end{bmatrix} \tag{3}$$

correspondingly for $e^{(P}(\boldsymbol{\beta}_P, \boldsymbol{\beta}_A)$ and $e^{(A}(\boldsymbol{\beta}_P, \boldsymbol{\beta}_A)$. It results that the two estimations of $\boldsymbol{\beta}_P$ and $\boldsymbol{\beta}_A$ might not necessarily coincide under the paper and author layout $\mathbf{y}^{(P}$ and $\mathbf{y}^{(A}$.

To characterize the behavior of $\boldsymbol{\beta}_P$ and $\boldsymbol{\beta}_A$ under the two aforementioned layouts, four projection operators are playing an important role: $H_{WA} = W_A X_A (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}}$, $H_P = X_P (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}}$, $H_{WP} = W_P X_P (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}}$, and $H_A = X_A (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}}$.

The two straightforward conditions to guarantee the uniqueness of the estimation under the two different layouts are: (i) papers exogenous characteristics $\mathcal{K}_P$ having no effect on the scientific impact in both dimensions, i.e. $\boldsymbol{\beta}_P = 0$; (ii) authors exogenous characteristics $\mathcal{K}_A$ having no effect on the scientific impact in both dimensions, i.e. $\boldsymbol{\beta}_A = 0$. In the first case, $\mathbb{E}[\mathbf{y}^{(P}] = W_A^{\mathrm{T}}(\boldsymbol{\beta}_A^{\mathrm{T}} X_A)$ and $\mathbb{E}[\mathbf{y}^{(A}] = X_A \boldsymbol{\beta}_A$, so that from (2) and (3) we respectively deduce that $\mathbb{E}[\mathbf{y}^{(P}] = H_{WA}\mathbb{E}[\mathbf{y}^{(P}]$ and $\mathbb{E}[\mathbf{y}^{(A}] = H_A \mathbb{E}[\mathbf{y}^{(A}]$.

---

[4] The nodal degree was the first used index of centrality, defined as the number of links incident upon a node (i.e. the number of ties that a node has). The eigenvector centrality assigns relative scores to all nodes incident upon a node, so that high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. By contrast, betweenness and closeness centrality indexes are both based on the idea of proximity and inclusion of a node within large collections of others.

[5] Note that the ordinary least square estimator coincides with the corresponding maximum likelihood estimators when the error term is normally distributed.

This implies that $\boldsymbol{\beta}_P$ is unique in both dimensions if the paper outcomes $\mathbb{E}[\mathbf{y}^{(P)}]$ belong to the columns space of $X_A^{\mathrm{T}}W_A^{\mathrm{T}}$ – that is to say, they are generated by the projected author characteristics – and the author outcomes $\mathbb{E}[\mathbf{y}^{(A)}]$ belong to the columns space of X_A^{T} – that is to say, they are generated by the author characteristics.

A similar reasoning is valid for the second case, where $\mathbb{E}[\mathbf{y}^{(P)}] = X_P\boldsymbol{\beta}_P$ and $\mathbb{E}[\mathbf{y}^{(A)}] = W_P^{\mathrm{T}}(\boldsymbol{\beta}_P^{\mathrm{T}}X_P)$. This entails $\mathbb{E}[\mathbf{y}^{(A)}] = H_{HP}\mathbb{E}[\mathbf{y}^{(A)}]$ – that is to say, authors outcomes are generated by the projected paper characteristics – and $\mathbb{E}[\mathbf{y}^{(P)}] = H_P\mathbb{E}[\mathbf{y}^{(P)}]$ – that is to say, papers outcomes are generated by the paper characteristics.

The following proposition provides a weaker condition for the consistency of both estimations, based on spectral properties of the co-authorship structure.

Proposition 1. *We claim that the author co-variates X_A and the estimated author effects $\boldsymbol{\beta}_A$ are consistent with the two-mode projections of scientific outcomes $\mathbb{E}[\mathbf{y}^{(A)}] = \hat{W}_P^{\mathrm{T}}\mathbb{E}[\mathbf{y}^{(P)}]$ if and only if $(\boldsymbol{\beta}_A^{\mathrm{T}}X_A)$ is the eigenvector centrality measure in the author–author network.*

This is in line with the idea that the author centrality entirely depends on its demographic and structural properties $\mathcal{K}_A$.[6]

Proposition 2. *Consider a Gaussian error on the paper and author outcomes, with respective expectations $\mu_P = b_P^{\mathrm{T}}X_P + W_A(b_A^{\mathrm{T}}X_A)$ and $\mu_A = b_A^{\mathrm{T}}X_A + W_P^{\mathrm{T}}(b_P^{\mathrm{T}}X_P)$, and variances σ_P^2 and σ_A^2, where b_P and b_A are the true values of the regression coefficients under the two models. Based on the paper layout $\mathbf{y}^{(P}$, the ordinary least square estimators (2) are*

$$\hat{\boldsymbol{\beta}}_P \sim \mathcal{N}\left(B_{WA}\mu_P, \sigma_P^2 B_{WA}B_{WA}^{\mathrm{T}}\right) \quad \text{and} \quad \hat{\boldsymbol{\beta}}_A \sim \mathcal{N}\left(B_P\mu_P, \sigma_P^2 B_P B_P^{\mathrm{T}}\right).$$

where $B_{WA} = \left(I - (X_P^{\mathrm{T}}X_P)^{-1}X_P^{\mathrm{T}}H_{WA}X_P^{\mathrm{T}}\right)^{-1}(X_P^{\mathrm{T}}X_P)^{-1}X_P^{\mathrm{T}}(I - H_{WA})$ and $B_P = \left(I - (X_A^{\mathrm{T}}W_A^{\mathrm{T}}W_AX_A)^{-1}X_A^{\mathrm{T}}W_A^{\mathrm{T}}H_PX_AW_A\right)^{-1}(X_A^{\mathrm{T}}W_A^{\mathrm{T}}W_AX_A)^{-1}X_A^{\mathrm{T}}W_A^{\mathrm{T}}(I - H_P)$. Similarly, based on the author layout (3), we claim that

$$\hat{\boldsymbol{\beta}}_P \sim \mathcal{N}(B_A\mu_A, \sigma_A^2 B_A B_A^{\mathrm{T}}) \quad \text{and} \quad \hat{\boldsymbol{\beta}}_A \sim \mathcal{N}(B_{WP}\mu_A, \sigma_A^2 B_{WP}B_{WP}^{\mathrm{T}}).$$

where $B_A = \left(I - (X_P^{\mathrm{T}}W_P^{\mathrm{T}}W_PX_P)^{-1}X_P^{\mathrm{T}}W_P^{\mathrm{T}}H_AX_PW_P\right)^{-1}(X_P^{\mathrm{T}}W_P^{\mathrm{T}}W_PX_P)^{-1}X_P^{\mathrm{T}}W_P^{\mathrm{T}}(I - H_A)$, and $B_{WP} = \left(I - (X_A^{\mathrm{T}}X_A)^{-1}X_A^{\mathrm{T}}H_{WP}X_A^{\mathrm{T}}\right)^{-1}(X_A^{\mathrm{T}}X_A)^{-1}X_A^{\mathrm{T}}(I - H_{WP})$.

The main consequence of Proposition 2 is the degeneracy of the sampling distributions of $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_A$ under the paper layout $\mathbf{y}^{(P}$, when the paper properties X_P^{T} are collinear to the projected author properties X_AW_A. By analogy, the same degeneracy happens to the sampling distributions of $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_A$ under the author layout $\mathbf{y}^{(A}$, when the author properties X_A^{T} are collinear to the projected paper properties X_PW_P. An immediate implication of this fact is a new level of multicollinearity which might appear in the two-mode regression (see Section 4.2).

A complementary result is the behavior of the sampling distributions of $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_A$ when paper and author information are *fully independent*. Under the paper layout $\mathbf{y}^{(P}$, when the paper properties X_P^{T} are orthogonal to the projected author properties X_AW_A, the variance of the estimators reduce to

$$\mathbb{V}[\hat{\boldsymbol{\beta}}_P] = \sigma_P^2(X_P, X_P^{\mathrm{T}})^{-1} \quad \text{and} \quad \mathbb{V}[\hat{\boldsymbol{\beta}}_A] = \sigma_P^2(\mathrm{Sim}(X_A, W_{AA}X_A))^{-1} \tag{4}$$

Similarly, under the author layout $\mathbf{y}^{(A}$, when the author properties X_A^{T} are orthogonal to the projected paper properties X_PW_P, the variance of the estimators reduce to

$$\mathbb{V}[\hat{\boldsymbol{\beta}}_P] = \sigma_A^2(\mathrm{Sim}(X_P, W_{PP}X_P))^{-1} \quad \text{and} \quad \mathbb{V}[\hat{\boldsymbol{\beta}}_A] = \sigma_A^2(X_A, X_A^{\mathrm{T}})^{-1} \tag{5}$$

Both cases coincide to the well known variance structure of a multiple linear regression $(X_A, X_A^{\mathrm{T}})^{-1}$ and $(X_P, X_P^{\mathrm{T}})^{-1}$ when the co-authorship structure is *trivially disconnected* (each paper is published by a unique author, and each author publishes a unique paper).

This modeling framework will be used in Section 4.2 to predict citation and impact factors at both layers based on the combined item and individual characteristics.

3. Data selection and processing

The data set used in this study contains the scientific publications indexed in the WOS database between 2009 and 2013 in Neuroscience (as a field of subject category WC). A collection of 153.182 research papers were retrieved and a stratified random sampling has been conducted to re-sample from the retrieved set of papers – a 3% sampling error and 95% of level of confidence have been used to determine the sample size. Table 1 shows the total number of publications and the stratified sample size per year in the studied field.

[6] A penalized regression approach can also be conceived based on the proximity between $\boldsymbol{\beta}_A^{\mathrm{T}}X_A$ and $(W^{\mathrm{T}}D_P^{-1})^{\mathrm{T}}(D_A^{-1}W)(\boldsymbol{\beta}_A^{\mathrm{T}}X_A)$. This would allow for a stronger consistency of the two levels of estimations in the individual and item dimensions.

Table 1
The total number of publications and the stratified sample size, 2009–2013.

Year	# publications (%)	Stratified sample size
2009	28,819 (18.81%)	199
2010	30,154(19.69%)	208
2011	31,030 (20.26%)	214
2012	31,265 (20.41%)	218
2013	31,914 (20.83%)	221
Total	153,182	1060

Table 2
Number of authors per gender and nationality in the stratified sample of Table 1.

	# authors (%)
Genders	
Male	3499 (64.96%)
Female	1886 (35.04%)
Nationalities	
USA	1740 (34.31%)
Japan	450 (8.35%)
China	372 (6.96%)
Italy	353 (6.50%)
Germany	344 (6.39%)
UK	265 (4.94%)

Table 3
Number of papers associated with type of access (open access/non-open access) and collaboration (national/international) in the stratified sample of Table 1.

	# papers (%)
Type of access	
Open access	462 (45.87%)
Non-open access	545 (54.12%)
Type of collaboration	
National	788 (78.25%)
International	219 (21.74%)

Table 4
Number of connected components and densities of the two projections of the author–paper bipartite network.

One-mode network	First largest component	Second largest component	Total
Paper–paper	8 (0.25)	5 (0.5)	1007 (0.000)
Author–author	55 (0.34)	53 (0.17)	5385 (0.001)

Authors have been assigned to a gender – using a procedure based on their first names – and to a nationality – based on the authors' country of affiliation. The gender of 5261 (91.80%) authors was directly assigned from either their first names or by checking internet directories or authors' websites. The gender of 124 (2.16%) authors was assigned by contacting them. The gender of 346 (6.04%) authors remained unspecified. They correspond to 53 (5%) papers, which have been eliminated from the sample, resulting in a data set comprising 1007 (95%) of the 1060 papers. These 1007 papers were used as our data set for further analysis. A summarizing description of the distribution of gender and nationality[7] is reported in Table 2.

Similarly, for each paper, the following characteristics were obtained: the number of authors, the number of funding bodies[8] the type of collaboration (national/international) and the type of access (open access/non-open access). In order to check if a paper was published as open access/non-access article, we manually verified internet access to the full text. A summary of the distribution of the type of access and type of collaboration is reported in Table 3.

The two-mode author–paper network was projected into two one-mode networks: author–author network (co-authorship) and paper–paper network (articles sharing authors). By the first projection, a network structure of scientific collaboration between authors was generated by connecting those authors whose names jointly appear in one or more of the 1007 articles. Similarly, by the second projection, a connection structure between papers was generated, where two papers are connected if and only if they share at least a common author. The two resulting one-mode networks comprised 207 and 492 disconnected components, respectively. The size and density of the two largest components are reported in Table 4.

[7] The data set contains 47 nationalities. To save space, Table 2 reports the ones which appear more frequently.
[8] Funding acknowledgment information was extracted from papers and standardized, as the proper standardization of the names of funding bodies is still a problem in WOS.

Table 5
Assortativity coefficient with respect to centrality measures in paper and author dimension.

	One-mode network	First largest component	Second largest component	Global network
Degree	Paper–paper	−0.586	−0.805	0.136
	Author–author	0.196	0.220	0.846
Eigenvector	Paper–paper	0.210	0.251	0.818
	Author–author	0.445	0.875	0.868
Betweenness	Paper–paper	−0.592	−0.667	0.152
	Author–author	−0.052	0.068	0.117
Closeness	Paper–paper	−0.322	−0.462	0.999
	Author–author	−0.261	−0.436	0.999

A description of the two largest components of both network structures have been reported in Appendix B, while a comprehensive statistical analysis of this data set is carried out in the next section, in accordance with the modeling framework described in Section 2.

4. Statistical analysis and results

4.1. Descriptive analysis: assortative patterns in co-authorship

As previously introduced in Section 2, the descriptive analysis focuses on the associations between nodal properties (either paper and author characteristics $\mathcal{K}_P$ and $\mathcal{K}_A$ or scientific outputs $\mathcal{O}$).[9]

Building on the assortative pattern in both structures of connections $\mathcal{E}_P$ and $\mathcal{E}_A$, we focus on the association between network centrality measures of connected nodes, and the association between exogenous nodal characteristics.

To quantify the importance of a node in a network, four centrality measures are considered: *degree*, *eigenvector*, *betweenness* and *closennes*. Table 5 reports the corresponding assortativity coefficients for the first two largest components and the global networks in paper and author dimension. Each coefficient represents the Pearson correlation between the levels of centrality of adjacent nodes.

As can be seen from Table 5, globally at author level, there is a high correlation between connected nodes in terms of the degree and eigenvector centralities. However, poor assortativity is observed with respect to betweenness centralities. Due to the low connectivity of both author and paper networks, the correlations between the closeness centrality of connected nodes reaches remarkably high levels, which do not match with the ones computed with respect to each single connected component.[10]

In the paper dimension, the observed similarity between the centrality measures of papers sharing common authors (connected papers) have different possible interpretations: (i) the assortative behavior of papers varies when different measures of centrality are adopted; (ii) connected papers have similar eigenvector centrality but different degrees (correspondingly 0.818 and 0.136 assortativity coefficient at the global network), which suggests that authors are heterogeneous in terms of their collaboration strategies (some of them are systematically taking part in tightly-knit and highly productive teams, while the remaining follow a more sparse collaboration strategy).

An analogous story can be inferred in author dimension and with regard to largest components. However, in this case, the assortative patterns corresponding to the degree and the eigenvector centralities are both positive. This means that highly central authors (in terms of eigenvector and degree) tend to connect in common research projects (papers).

The presence of frequent connections between central nodes has possible implication for network transitivity and community structure. In fact, 0.821, 0.830, and 0.411 clustering coefficients have been found in the first largest component, second largest component and the global network respectively at the author–author one-mode structure.

The above mentioned finding is also supported by the assortative pattern gained by the exogenous nodal characteristics. This reveals the tendency (bias) of papers and authors to be connected with others with similar exogenous nodal characteristics, such as the ones in $\mathcal{K}_A$ and $\mathcal{K}_P$.

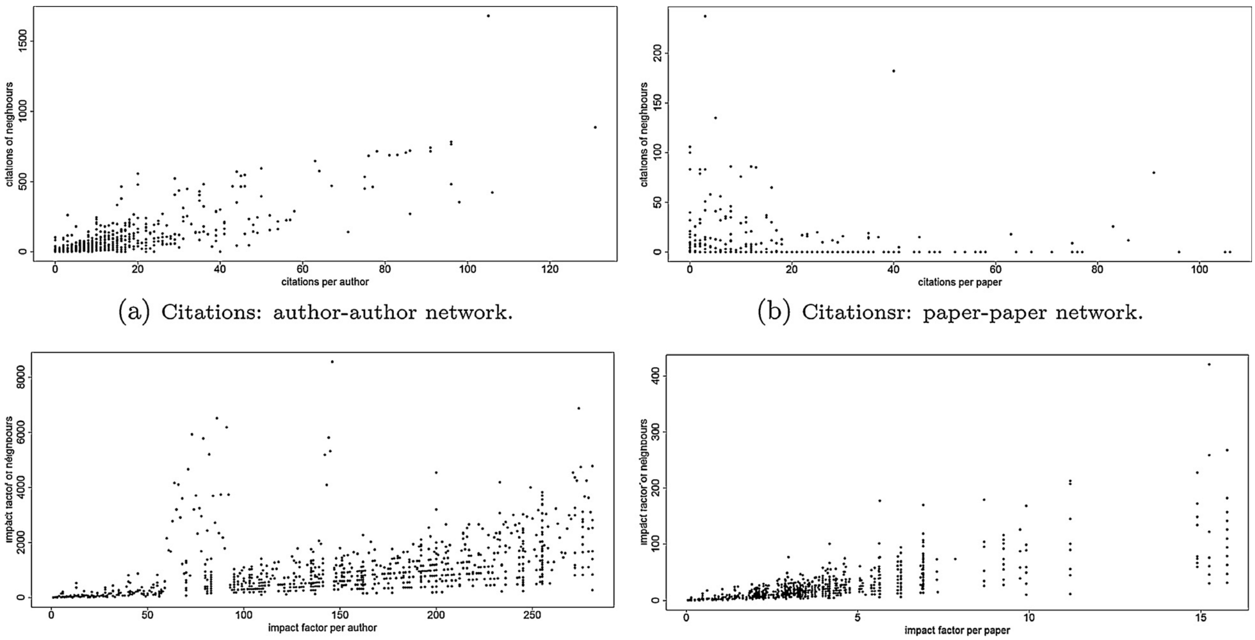
Table 6 suggests a positive propensity of authors to cooperate within the same national scientific community (assortativity coefficients of 0.25, 0.62 and 0.44, for the first largest component, second largest component and the global network respectively). By contrast, both female and male neuroscientists showed a similar propensity to take part in mixed-gender collaborative projects.

From the paper viewpoint, international papers (articles co-authored by scholars from the different nationalities) more frequently share authors with other international ones, while national papers more frequently share authors with other

[9] Due to the poor connectedness of the one-mode paper–paper projection, the main results of this subsection refer to the one-mode author–author projection.
[10] In highly disconnected networks, the closeness centrality score of most of nodes would be approximately $1/(n-1)^2$.

Table 6
Assortativity coefficient to measure the similarity between connected authors (in the author–author network) and the similarity between connected papers (in the paper–paper network), with respect to four exogenous nodal properties.

	First largest component	Second largest component	Global network
Paper–paper network			
Type of access	−0.17	−0.25	0.17
Type of collaboration	0.16	0.43	0.34
Author–author network			
Gender	−0.02	0.06	0.06
Nationality	0.25	0.62	0.44



(a) Citations: author-author network. (b) Citationsr: paper-paper network.

(c) Impact factor: author-author network. (d) Impact factor: paper-paper network.

Fig. 1. Assortativity with respect to the number of citations.

Table 7
Assortativity coefficient with respect to the scientific outcomes in paper and author dimension.

	Citations	Impact factors
Author–author network	0.9188	0.912
Paper–paper network	0.0150	0.961

national ones (assortativity coefficients of 0.16, 0.43 and 0.34, for the first largest component, second largest component and the global network respectively).

We expect this conservative pattern of network self-similarity to give rise to extensive associations between the scientific outcomes of papers and authors at a short network distance. The graphical illustration in Fig. 1 shows these type of associations: each point in the scatter-plots correspond to the scientific outcome (either citations or impact factor) of a node (either paper or author) and the average outcomes of its neighbors.

As trivially expected the scientific impact of an author is strongly associated to the ones of his/her co-authors, as shown in the plot (a) of Fig. 1. However, plot (b) of Fig. 1 reveals that citations of a paper are not related to the ones of other papers sharing the same authors (i.e. highly cited papers might be authored by scholars who also published other less cited ones). Connected papers are published in journals with similar impact factors (or possibly the same journal) (see plot (d) in Fig. 1).

The graphical illustration of Fig. 1 can again be summarized by the assortative patterns of citations and impact factors at both network layers, as reported in Table 7.

In the next subsection, the observed dependency of author and paper properties on each other will be entirely attributed to other levels of association: assortativity with respect to the nodal centrality and exogenous paper and author characteristics.

Table 8
Two multiple regressions of paper and author citations. This is based on the paper and author information introduced separately in each regression model.

	Paper outcome		Author outcome	
	Estimate	*p*-Value	Estimate	*p*-Value
Paper characteristics $\mathcal{K}_P$				
Access status	1.070	0.175	–	–
Number of funding	0.013	0.946	–	–
Number of authors	0.252	0.0166*	–	–
National status	−3.223	7.98e−07***	–	–
Impact factor	2.170	<2e−16***	–	–
Author characteristics $\mathcal{K}_A$				
Gender	–	–	−0.801	0.057
Number of papers	–	–	1.380	0.275
Centrality	–	–	−1.488	0.739
Argentina	–	–	3.207	0.437
Australia	–	–	6.688	4.9e−04***
Italy	–	–	6.078	1.3e−04***
Japan	–	–	6.205	7.0e−05***
Russia	–	–	1.940	0.684
UK	–	–	11.371	7.8e−12***
USA	–	–	10.858	1.9e−15***

* a *p*-value smaller than 0.05; ** a *p*-value smaller than 0.01; *** a *p*-value smaller than 0.001.

4.2. Predictive analysis: two-mode regression

Scientific outcomes are studied in this section from the dual perspective of the two-mode regression model (1), which is independently applied to citations and impact factors. This is based on the same specification of the collections of explanatory variables $\mathcal{K}_P$ and $\mathcal{K}_A$.

To empirically assess the advantages of including paper and author information through their two-mode projection, we compare the proposed modeling approach with two separated regressions at paper and author level.

As already mentioned in the previous subsection, an underlying assumption in this regression analysis is that the observed similarity between scientific outcomes of connected nodes can be attributed to paper and author exogenous characteristics. This entails that all systemic similarities (as the ones between connected nodes observed in Fig. 1) should be explained by these collections of co-variates.

The two scientific outcomes are not assumed to be independent, i.e. given the collection of co-variates, papers citations might depends on the impact factors of the corresponding journal.[11]

Focusing on citations, Table 8 reports the estimated coefficients and corresponding *p*-values, for the two separated models (multiple linear regression) based on paper and author information (without their simultaneous inclusion). In contrast, Table 9 is based on the simultaneous inclusion of the paper and author information (two-mode regression). Both regression models are carried out at both dimensions, corresponding to (1).

Similarly, when analyzing impact factor, Tables 10 and 11 report the respective estimated coefficients and corresponding *p*-values, without the simultaneous inclusion of the paper–author information (multiple linear regression). In other words, they are based on the simultaneous inclusion of the paper–author information (two-mode regression).

With a few exceptions the four model specifications in Tables 8–10 and 11 are consistent in terms of coefficient signs, but different in terms of significance. The number of stars of the *p*-values refers to the significance level.[12]

All regression models support the positive effect of open-access on the number of received citations and the impact factors.

The co-variates ‘number of authors’ and ‘number of papers’ refer to the number of authors of the corresponding paper and the number of papers of the corresponding author, respectively (calculated as row and column sum of the two-mode matrix W). The number of author per paper has a small effect on the number of citations and a slightly larger impact on impact factors.

By contrast, the number of funding bodies has different effects on citation and impact factor. In the first case it has no impact, whereas in the latter case it plays a significantly positive role (i.e. papers with several funding bodies are often found in high impact factor journals).

As relevant demographic features, special attention should be given to the impact of gender and nationality.[13]

[11] As widely argued by Didegah and Thelwall (2013), van Eck et al. (2013), and van der Pol et al. (2015), papers appearing in journals with a higher IF, receive more citations. Specifically, the expected number of citations is assumed to be linear with respect to the impact factor of the corresponding journal.
[12] One star means “a *p*-value smaller than 0.05”; two stars means “a *p*-value smaller than 0.01”; three stars means “a *p*-value smaller than 0.001”.
[13] The reason why we picked these seven nationalities as representative of the rest of the collection is the intriguing mismatch between their effect in terms of citations and impact factor (i.e. while Australia has a big effect on citations, it has a smaller one on impact factor).

Table 9
Two-mode regressions of paper and author citations. This is based on the paper and author information introduced jointly thought their connection.

	Paper outcome		Author outcome	
	Estimate	*p*-Value	Estimate	*p*-Value
Paper characteristics $\mathcal{K}_P$				
Access status	0.249	<2e−16***	0.764	<2e−16***
Number of funding	−0.131	0.552	0.085	0.348
Number of authors	0.049	2.01e−04**	−0.093	0.077
National status	−6.151	<2e−16***	−7.097	<2e−16***
Impact factor	1.993	<2e−16***	2.074	<2e−16
Author characteristics $\mathcal{K}_A$				
Gender	−2.298	<2e−16***	−0.714	0.007**
Number of papers	2.325	<2e−16***	1.975	<2e−16***
Centrality	−31.291	<2e−16***	−28.970	<2e−16***
Argentina	2.387	2.38e−08***	2.804	<2e−16***
Australia	9.291	<2e−16***	7.251	<2e−16***
Italy	7.142	<2e−16***	6.534	<2e−16***
Japan	7.295	<2e−16***	1.217	<2e−16***
Russia	7.160	<2e−16***	6.332	<2e−16***
UK	7.032	<2e−16***	6.119	<2e−16***
USA	8.718	<2e−16***	7.668	<2e−16***

* a *p*-value smaller than 0.05; ** a *p*-value smaller than 0.01; *** a *p*-value smaller than 0.001.

Table 10
Two multiple regressions of paper and author impact factor. This is based on the paper and author information introduced separately in each regression model.

	Paper outcome		Author outcome	
	Estimate	*p*-Value	Estimate	*p*-Value
Paper characteristics $\mathcal{K}_P$				
Access status	1.339	<2e−16***	–	–
Number of funding	0.298	5.16e−15***	–	–
Number of authors	0.211	<2e−16***	–	–
National status	1.105	<2e−16***	–	–
Author characteristics $\mathcal{K}_A$				
Gender	–	–	0.013	0.881
Number of papers	–	–	−0.379	0.148
Centrality	–	–	11.129	<2e−16***
Argentina	–	–	4.071	2.1e−06***
Australia	–	–	4.333	<2e−16***
Italy	–	–	3.984	<2e−16***
Japan	–	–	3.574	<2e−16***
Russia	–	–	1.550	0.118
UK	–	–	5.630	<2e−16***
USA	–	–	5.742	<2e−16***

* a *p*-value smaller than 0.05; ** a *p*-value smaller than 0.01; *** a *p*-value smaller than 0.001.

The last rows of each table correspond to seven selected nationalities. We noticed that nationalities that publish papers in high impact factor journals are not necessarily those that publish highly cited papers. This disagreement supports the substantial heterogeneity of scientific production.

To support the consistency of the different model specifications, Fig. 2 reports the pairwise comparisons between the estimated nationalities coefficients of Tables 8 and 9, for the citation analysis, and Tables 10 and 11, for the impact factor analysis.[14]

The national/international status also has different effects on citation and impact factor. On the one hand, the estimated parameters in Tables 8 and 9 indicate that national collaborations receive less citations than their international counterparts. On the other hand, Tables 10 and 11 show that national collaboration tends to have a slightly higher chance to appear in journals with high impact factors.

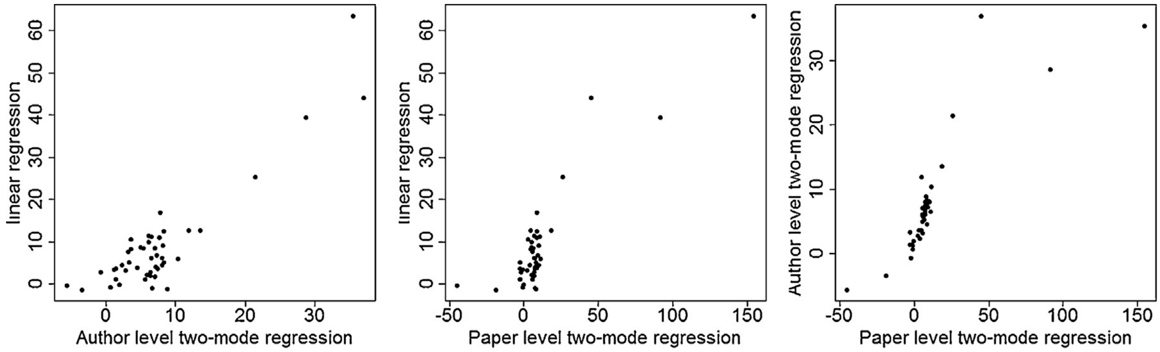
As far as the gender is concerned we observe a negative propensity of women neuroscientists to receive citations, compared to men. However, when it comes to the quality of the journals, women do not exhibit any significant tendency to publish articles in journals with higher or lower impact factors.

[14] Bulgaria is associated to the unique outlier which appears with a large nationality coefficient in the six plots in Fig. 2. This is due to the fact that Bulgaria has a unique paper in the data set, which resulted to be highly cited.

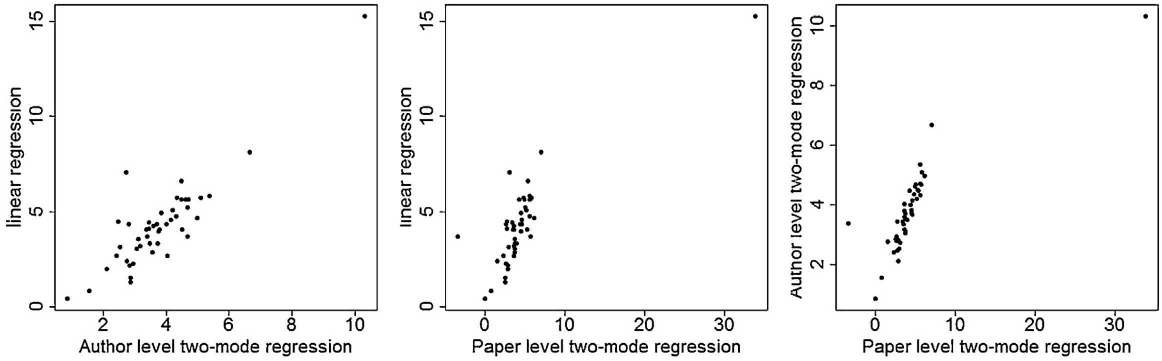
Table 11
Two-mode regressions of paper and author impact factors. This is based on the paper and author information introduced jointly thought their connection.

	Paper outcome		Author outcome	
	Estimate	p-Value	Estimate	p-Value
Paper characteristics $\mathcal{K}_P$				
Access status	1.036	3.1e−12***	1.071	<2e−16***
Number of funding	0.263	<2e−16***	0.219	0.157
Number of authors	0.134	0.001**	0.186	0.002*
National status	0.433	7.9e−04**	0.555	0.003*
Author characteristics $\mathcal{K}_A$				
Gender	0.096	0.237	0.001	0.499
Number of papers	−0.742	4.1e−05***	−0.585	2.3e−06***
Centrality	6.696	0.060	6.630	<2e−16***
Argentina	5.233	<2e−16***	4.500	<2e−16***
Australia	4.319	<2e−16***	4.017	<2e−16***
Italy	4.528	<2e−16***	3.753	<2e−16***
Japan	3.739	<2e−16***	3.117	<2e−16***
Russia	2.471	5.0e−04**	2.856	<2e−16***
UK	4.897	<2e−16***	4.610	<2e−16***
USA	4.826	<2e−16***	4.353	<2e−16***

* a p-value smaller than 0.05; ** a p-value smaller than 0.01; *** a p-value smaller than 0.001.



(a) Nationalities coefficients of the citation models in tables 8 and 9.



(b) Nationalities coefficients of the impact factor models in tables 10 and 11.

Fig. 2. The nationalities coefficients of the estimated regression models are graphically compared to asses consistency. The scatter-plots provide a collection of pairwise comparisons between the 47 nationalities coefficients for each pair of model specifications.

4.3. Model fit and diagnostics

A collection of regression diagnostics have been taken into account to assess the adequacy of the model specification, and also to compare the proposed two-mode regression with the two separated multiple linear regressions. In particular the multicollinearity issue and the residual distribution are monitored for each model specification.

As discussed in Section 2, the main consequence of Proposition 2 is the degeneracy of the sampling distributions of $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_A$ when paper and author information are *fully dependent* through the co-authorship structure. The immediate implication of this fact is a new level of multicollinearity which might appear in the two-mode regression.

Table 12
Correlations between co-variates.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Paper layout	−0.452	−0.009	−0.004	−0.007	−0.002	0.292
Author layout	−0.372	−0.014	−0.006	−0.009	−0.002	0.349

Table 13
Goodness of fit of the estimated number of citations.

	Paper outcome		Author outcome	
	Linear regr. (Table 8)	Two-mode regr. (Table 9)	Linear regr. (Table 8)	Two-mode regr. (Table 9)
R^2	0.523	0.539	0.233	0.580
MAPE	0.510	0.499	0.488	0.486

Table 14
Goodness of fit of the estimated impact factors.

	Paper outcome		Author outcome	
	Linear regr. (Table 10)	Two-mode regr. (Table 11)	Linear regr. (Table 10)	Two-mode regr. (Table 11)
R^2	0.650	0.653	0.360	0.683
MAPE	0.224	0.213	0.247	0.218

Table 12 summarizes the information of the correlation coefficients between every pair of co-variates in terms of paper and author characteristics. This supports the independence of co-variates from each other and the correctness with respect to the multicollinearity issue.

As measures of goodness of fit we consider the coefficient of determination (R^2) and the mean absolute proportion error (MAPE). They can both be interpreted as the amount of the variation in paper and author outcomes that are predictable based on $\mathcal{K}_P$ and $\mathcal{K}_A$. For the citation analysis, Table 13 reports the corresponding values of R^2 and MAPE for the multiple linear regression and the two-mode regression results of Tables 8 and 9, respectively. Likewise, Table 14 reports an analogous goodness of fit for impact factor analysis, as resulting from the multiple linear regression and the two-mode regression results of Tables 10 and 11 respectively.

We observe an improvement in the fitted models for all regressions, when paper and author information is jointly included in the model. The results indicate a slight improvement at the paper level, whereas the two-mode regression appears to out-perform the linear regression at the author level. This can be partially explained by the initial lack of fit of the linear regression models at the author level, which magnify the impact of the inclusion of paper properties.

Fig. 3 provides a comparison between the observed citations and impact factors, and the one predicted by the model, for Tables 9 and 11 respectively. The prediction is reported on the vertical axis, while the observation is reported on the horizontal axis.

Although a quite accurate prediction is observed for the impact factors at both paper and author levels, the number of citations seem to be difficult to characterize in relation to the aforementioned co-variates $\mathcal{K}_P$ and $\mathcal{K}_A$ (see Fig. 3).

5. Discussion

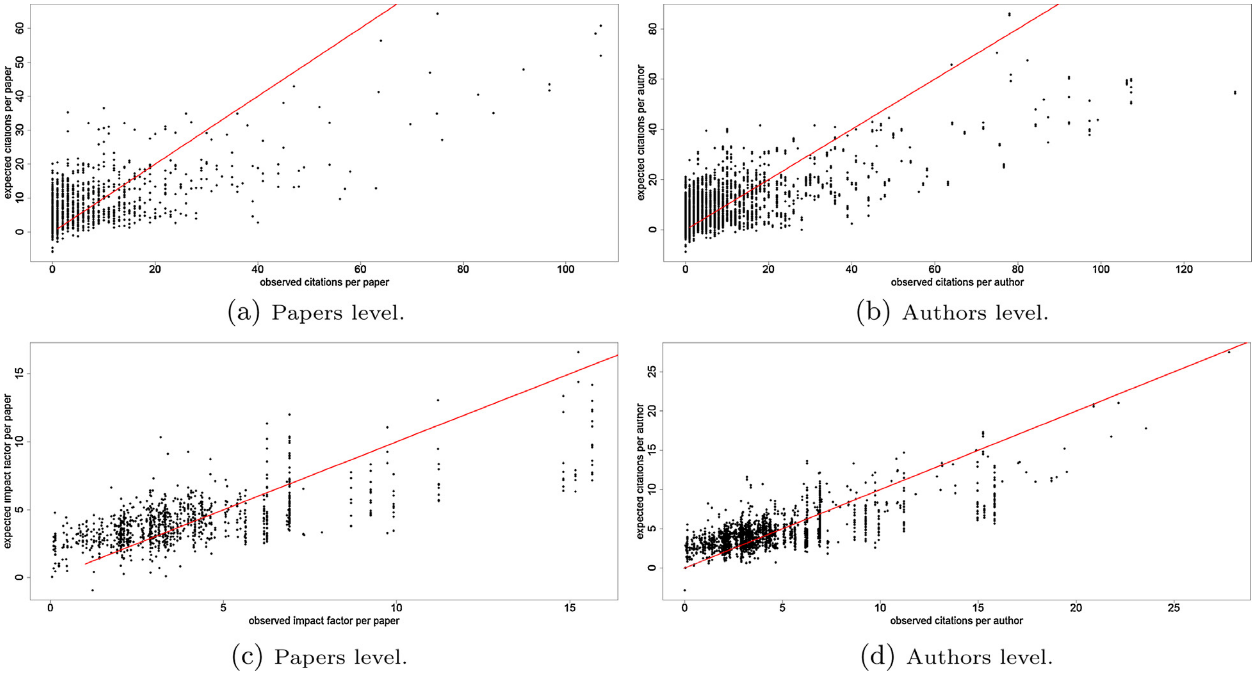
The paper provides an insight into the research impact of scholars involved in scientific collaborations, based on the number of citations and the impact factor of published papers.

To do this, we first described the neuroscience community, based on the assortativity of network structural properties, exogenous nodal characteristics and nodal outcomes (scientific impact). Then, a novel regression approach which exploits the information associated to both layers of the two-mode co-authorship structure has been described and applied to the analyzed data set.

Regarding assortative patterns in co-authorship, three relevant conclusions can be inferred: first, high assortative pattern of nodal centrality for both paper–paper network and author–author network, second, transitive co-authorship structures and attachment to similar others (homophily), and finally, heterogeneous scientific outcomes at paper and author level (the distribution of scientific outcomes presents high variability).

Regarding the structural properties, the results revealed that authors in the neuroscience community are heterogeneous in terms of their collaboration strategies. That is to say, while some of them are systematically taking part in tightly-knit group collaboration, others follow a more sparse collaboration strategy. Furthermore, highly central authors (in terms of their co-authorship connection pattern) tend to collaborate within same projects.

Similarly, when studying nodal exogenous properties (nationality and type of collaboration), it has been observed that authors tend to collaborate in small densely connected national communities, with very few international relations.The presence of frequent connections between central nodes in paper and author dimensions, suggests possible implications for the network transitivity and community structure. This finding is in agreement with a study conducted by Nasini et al.



(a) Papers level. (b) Authors level.

(c) Papers level. (d) Authors level.

Fig. 3. Estimation versus observation for the four estimated models. The red line is the identity line $x = y$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

(2017). Similar results are also reported by Cummings and Kiesler (2005) in their study on multidisciplinary collaborations. They argued that dispersed forms of collaboration might be obstructed by physical distance between scientists, which not only reduce the likelihood of collaboration, but also have a negative impact on success.

What can be globally suggested from the assortative pattern of author centralities and the appearance of these small densely connected communities, is the existence of a so-called conservative collaboration strategy (Gilbert, Ahrweiler, & Pyka, 2014) in the neuroscience community, which limits the inclusion and incorporation of researchers in collaborative projects.

From another point of view, it represents a scientific community in which a few authors are involved in a very specific research topic (Velden, Haque, & Lagoze, 2010). Due to the high level of expertise and specialization, the number of researchers collaborating on a specific research problem is often very limited. Therefore authors in neuroscience authorship community might have lower selection choice of collaborators (Ghiasi, Larivière, & Sugimoto, 2015).

Regarding nodal outcomes at author level, it was observed that neuroscientists tend to collaborate with those who have similar research impact. Additionally, the research impact of an author is strongly associated to those of its co-authors. This homophily of authors to collaborate with authors with higher level of similarity in terms of research impact can be explained by selectivity of researchers and Mathew effect i.e. that advantage tends to beget further advantage. Thus, in our case, accomplished authors (i.e. those with higher number of citation and impact factor) might prefer to collaborate with similar others to produce high-impact work. As also indicated in Guimera, Uzzi, Spiro, and Amaral (2005) study, collaborations between experienced authors is more likely to result in a publication in a high impact journal than in collaborations between unseasoned authors.

At paper level, an interesting insight was that the citations of a paper are not related to the ones of other papers sharing the same authors. This suggests that authors in neuroscience community might have signed papers with different number of citations (i.e. highly cited papers might have been authored by scholars who also published other less cited ones). By contrast, connected papers are published in journals with similar impact factors (or possibly the same journal). This might have three possible explanations: i) scholars might be conservative in the choice of the target journal, ii) editors might be conservative in the acceptance of papers by already known scholars, iii) a combination of both.

As a consequence, the global picture suggests that sharing authors might results in different scientific impact and visibility, when analyzed in terms of citations, but it certainly entails similar outcomes in terms of journal rankings.

In the next step, by using a two-mode regression model we were able to simultaneously account for the effect of paper properties (access status, number of funding bodies, number of authors, type of collaboration, centrality) and author characteristics (gender, nationality, number of papers, centrality, transitivity) on research success at both paper and author levels.

With regard to the two-mode regression analysis, different factors and paper properties appear to be responsible for boosting and increasing the number of received citations.

The open access status shows to increase the number of received citations at both paper and author dimensions. It is completely in line with the existing literature (Antelman, 2004; Eysenbach, 2006; Hajjem, Harnad, & Gingras, 2005; Lawrence, 2001). For instance, Lawrence (2001) found that openly accessible computer science articles were cited more. Hajjem et al. (2005) replicated the same analysis in ten disciplines (Biology, Psychology, Sociology, Health, Political Science, Economics, Education, Law, Business, Management), by testing 1,307,038 articles published across 12 years 1992–2003.

The number of funding sources has a controversial effect on the number of received citations and the quality of the journal in which the paper is published. Rigby (2011) claimed that *the number of funding sources which a paper cites is not invariably a reliable predictor of increased impact, and where such an impact can be found, the strength of the effect is very small.* This finding is however contradicted by successive works of Wang and Shapira (2015) on nanotechnology publications. Their study showed that sponsored research exhibits higher impacts in terms of both journal ranking and citation. Rigby (2013) suggested the reason for such a link is the fact that research supported by more funding bodies undergoes more peer review. This then leads to higher quality in terms of a greater number of citations received by publications. However, this argument cannot be extended to the neuroscience community as analyzed in our work, where we found a non-significant effect of funding bodies on citations.

The number of authors has previously been found to be correlated with the paper's impact and the number of received citations (Biscaro & Giupponi, 2014; Frosch et al., 2010; Gazni & Didegah, 2011; Tahamtan et al., 2016). In our study, we also found the same results with regard to paper and author impact factors as well as the number of citations received at paper level. However, at author level, authors number shows to have a negative impact on the number of received citations. This finding is in accordance with some other studies that found no such a relationship (Bornmann & Daniel, 2007a; Collet, Robertson, & Lup, 2014; Ruano-Ravina & Álvarez-Dardet, 2012).

The negative effect of national status on the number of citations received should be interpreted in accordance with the previously discussed tendency of authors to collaborate in small densely connected national communities, with little international collaboration. This reveals a clear consistency between the heterogeneity of paper scientific impacts and the disperse presence of international collaborations (Cummings & Kiesler, 2005). By contrast, paper resulting from international collaborations appear significantly more often in high impact factor journals, than their national counterparts.

From the author viewpoint, other factors are responsible for boosting the number of citations received and journal quality.

As far as gender is concerned, we observe a negative propensity of women neuroscientists to receive citations, compared to men. However, there is no significant difference between their journals quality compared to men. This was previously found by Ghiasi et al. (2015), in the context of engineering, a highly male-dominated field. The authors showed that women engineers publish their papers in journals with similar citations rates, while their work receives fewer citations from the engineering community. This allows for an extensive interpretation within the framework of the *Matilda effect* (Rossiter, 1993), by which women's publications receive less recognition than what is expected (in this case, expected from the journal in which their discoveries were published). Regarding nationality, our results suggested that the countries that publish in lower-ranked journals tend to over-perform in terms of the number of citations they receive. This result is in accordance with (Smith, Weinberger, Bruna, & Allesina, 2014). Smith et al. (2014) referred to academic culture (i.e., the "biased author" effect) as a potential reason for this finding. If the academic environment of a country is such that submission to to*p*-tier journals is discouraged or is not associated with professional advancement, then papers that could potentially be published in top journals may frequently be submitted to lower-tier ones, resulting in a mismatch between placement and performance. It could also be due to a negative bias at the peer-review stage (Opthof, Coronel, & Janse, 2002).

The proposed methodology and bibliometric application allow for different streams of future research.

From a statistical viewpoint, Proposition 1 motivates the use of a penalized regression approach, based on the proximity between the author characteristics of authors $\boldsymbol{\beta}_A^{\mathrm{T}} X_A$ and their one step propagation though network $(W^{\mathrm{T}} D_P^{-1})^{\mathrm{T}} (D_A^{-1} W)(\boldsymbol{\beta}_A^{\mathrm{T}} X_A)$. This would allow for a stronger consistency at both author and paper level. The two-mode regression analysis might be generalized to allows for exponential probabilistic models, as described by Nasini et al. (2017). A similar statistical framework would allow for the joint inference and prediction of scientific impact and collaborations. The time dimension might also be included in this generalized modeling framework, with the aim of assessing the dynamic change of research impact across paper and author scientific life cycles.

From a bibliometric viewpoint, this analysis might be replicated using data from other fields. This would require the definition of an automatized sampling approach, in particular for the selection of author genders and nationalities. Substantial insights might result from applying an analogous two-mode regression approach to the data presented by Lawrence (2001) and by Hajjem et al. (2005). This would help to provide a comparative analysis and a comprehensive picture of the role played by co-authorship and the role played by papers and authors characteristics on research impact.

Acknowledgements

The second author has been supported by the European Research Council – ref. ERC-2011-StG 283300-REACTOPS. The authors would like to thank the two anonymous reviewers for their valuable comments, suggestions and insights that helped improving the manuscript.

Author contributions

Conceived and designed the analysis: Tahereh Dehdarirad and Stefano Nasini contributed to define the problem and the question to be answered in the analysis.

Collected the data: Tahereh Dehdarirad contributed to carry out the sampling procedure and to collect the data set.

Contributed data or analysis tools: Stefano Nasini contributed to program the required codes to carry out the data manipulation.

Performed the analysis: Stefano Nasini has carried out the statistical analysis.

Wrote the paper: Tahereh Dehdarirad and Stefano Nasini contributed to write the manuscript.

Appendix A. Proofs

Proposition 1.

Proof. From (1) we see that $\mathbb{E}[\mathbf{y}^{(A)}] = (W^{\mathrm{T}} D_P^{-1})^{\mathrm{T}} \mathbb{E}[\mathbf{y}^{(P)}]$ is verified if and only if

$$\boldsymbol{\beta}_A^{\mathrm{T}} X_A = (W^{\mathrm{T}} D_P^{-1})^{\mathrm{T}} (D_A^{-1} W)(\boldsymbol{\beta}_A^{\mathrm{T}} X_A) \tag{A.1}$$

Note that $(W^{\mathrm{T}} D_P^{-1})^{\mathrm{T}} (D_A^{-1} W)$ is a row stochastic matrix, so that its largest eigenvalue is one. This means that $(\boldsymbol{\beta}_A^{\mathrm{T}} X_A)$ is the eigenvector centrality measure in the author–author network. □

Proposition 2.

Proof. From the first order condition (2) we have

$$\begin{aligned}
\boldsymbol{\beta}_P &= (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}} (\mathbf{y}^{(P} - W_A X_A (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}} (\mathbf{y}^{(P} - X_P \boldsymbol{\beta}_P)) \\
\boldsymbol{\beta}_A &= (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}} (\mathbf{y}^{(P} - X_P (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}} (\mathbf{y}^{(P} - W_A X_A \boldsymbol{\beta}_A))
\end{aligned}$$

by replacing H_{WA} and H_P with few algebraical operations we we find

$$\begin{aligned}
\boldsymbol{\beta}_P &= (I - (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}} X_P^{\mathrm{T}} H_{WA} X_P^{\mathrm{T}})^{-1} (X_P^{\mathrm{T}} X_P)^{-1} X_P^{\mathrm{T}} (I - H_{WA}) \mathbf{y}^{(P} \\
\boldsymbol{\beta}_A &= (I - (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}} H_P X_A W_A)^{-1} (X_A^{\mathrm{T}} W_A^{\mathrm{T}} W_A X_A)^{-1} X_A^{\mathrm{T}} W_A^{\mathrm{T}} (I - H_P) \mathbf{y}^{(P}
\end{aligned}$$

Similarly, from the first order condition (3) we have

$$\begin{aligned}
\boldsymbol{\beta}_P &= (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}} (\mathbf{y}^{(A} - X_A (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}} (\mathbf{y}^{(A} - W_P X_P \boldsymbol{\beta}_P)) \\
\boldsymbol{\beta}_A &= (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}} (\mathbf{y}^{(A} - W_P X_P (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}} (\mathbf{y}^{(A} - X_A \boldsymbol{\beta}_A))
\end{aligned} \tag{A.2}$$

by replacing H_{WP} and H_A with few algebraical operations we we find

$$\begin{aligned}
\boldsymbol{\beta}_P &= (I - (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}} H_A X_P W_P)^{-1} (X_P^{\mathrm{T}} W_P^{\mathrm{T}} W_P X_P)^{-1} X_P^{\mathrm{T}} W_P^{\mathrm{T}} (I - H_A) \mathbf{y}^{(A} \\
\boldsymbol{\beta}_A &= (I - (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}} H_{WP} X_A^{\mathrm{T}})^{-1} (X_A^{\mathrm{T}} X_A)^{-1} X_A^{\mathrm{T}} (I - H_{WP}) \mathbf{y}^{(A}
\end{aligned} \tag{A.3}$$

□

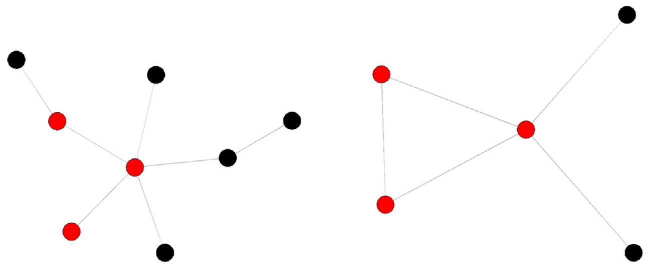
Appendix B. Largest components of both one-mode projections

Figs. 4 and 5 show the network plots of the two largest components of the paper–paper network, associated with two nodal characteristics: type of access (open access, non-open access) and type of collaboration (national, international).

Figs. 6 and 7 show the network plots of the two largest components of the author–author network, associated with nodal characteristics of gender and nationality.

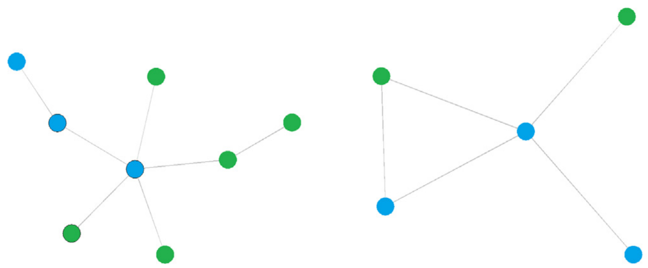
As previously mentioned, both network layers are associated with nodal outcomes, which can be measured in terms of citations and impact factor in the context of bibliometric analysis. They are originally measured at item level (paper) and then projected to the individual level (author), based on the two mode structure, i.e., $\mathbf{y}^{(A,o} = (W^{\mathrm{T}} D_P^{-1})^{\mathrm{T}} \mathbf{y}^{(P,o}$, for each $o \in \mathcal{O}$. Thus, authors' number of citations and impact factor is calculated with respect to their papers in the stratified sample.

Table 15 reports the mean and standard deviation of the number of citations and impact factor corresponding to both network layers.



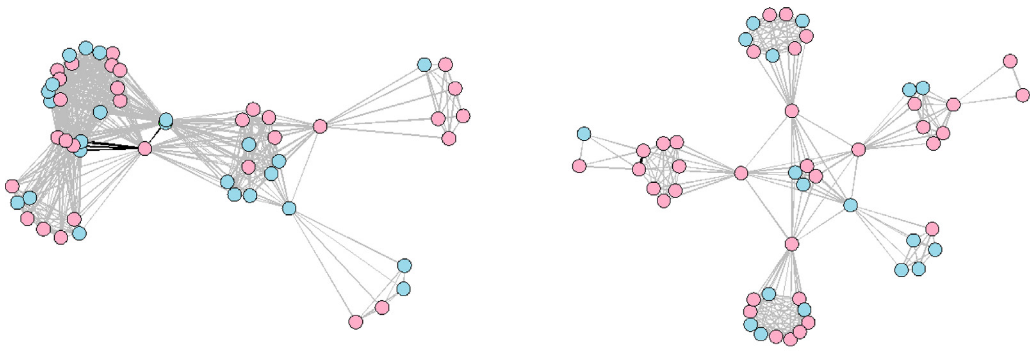
(a) First largest component. (b) Second largest component.

Fig. 4. The two largest components of the paper–paper network with the type of access (red for open, black for non-open access). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



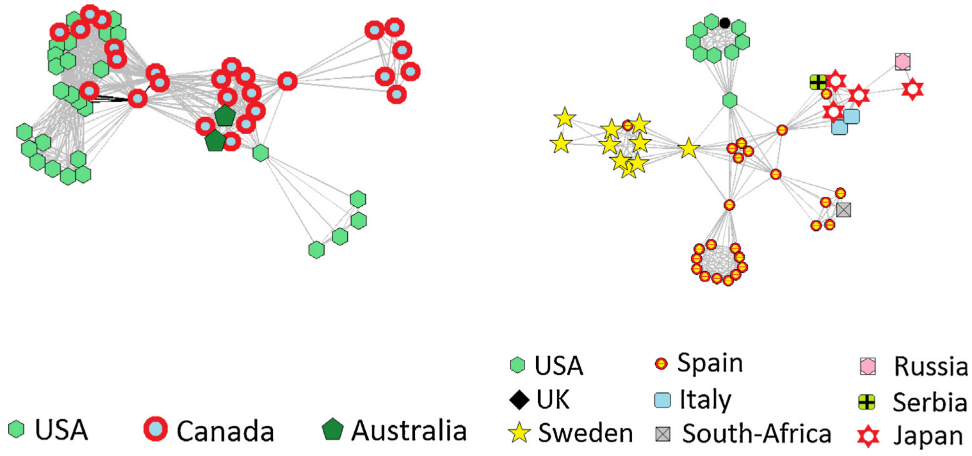
(a) First largest component. (b) Second largest component.

Fig. 5. The two largest components of the paper–paper network with the type of collaboration (blue for national, green for international). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



(a) First largest component. (b) Second largest component.

Fig. 6. The three largest components with nodal genders (blue for men, pink for women). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



(a) First largest component.

(b) Second largest component.

Fig. 7. The three largest components with nodal nationalities.

Table 15
Mean and standard deviation of the two nodal outcomes in both paper and author dimension.

Network layer	Average citations (sd)	Average impact factor (sd)
Paper	4.14 (2.96)	4.13 (2.95)
Author	4.63 (3.41)	4.63 (3.41)

References

Abramson, D., Lees, M., Krzhizhanovskaya, V., Dongarra, J., Sloot, P. M., Thedchanamoorthy, G., et al. (2014). Node assortativity in complex networks: An alternative approach. *Procedia Computer Science*, *29*, 2449–2461.

Annalingam, A., Damayanthi, H., Jayawardena, R., & Ranasinghe, P. (2014). Determinants of the citation rate of medical research publications from a developing country. *SpringerPlus*, *3*(1), 140.

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, *65*(5), 372–382.

Biscaro, C., & Giupponi, C. (2014). Co-authorship and bibliographic coupling network effects on citations. *PLOS ONE*, *9*(6), 1–12.

Bornmann, L., & Daniel, H.-D. (2007a]). 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.

Bornmann, L., & Daniel, H.-D. (2007b]). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.

Castro, J., & Nasini, S. (2015). Mathematical programming approaches for classes of random network problems. *European Journal of Operational Research*, *245*(2), 402–414.

Collet, F., Robertson, D. A., & Lup, D. (2014). *When does brokerage matter? Citation impact of research teams in an emerging academic field*. Strategic Organization.

Costas, R., & Bordons, M. (2007). The *h*-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, *1*(3), 193–203.

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

De Stefano, D., Giordano, G., & Vitale, M. P. (2011). Issues in the analysis of co-authorship networks. *Quality & Quantity*, *45*(5), 1091–1107.

Didegah, F., & Thelwall, M. (2013). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, *64*(5), 1055–1064.

Diekhoff, T., Schlattmann, P., & Dewey, M. (2013). Impact of article language in multi-language medical journals – A bibliometric analysis of self-citations and impact factor. *PLOS ONE*, *8*(10).

Ding, Y., Rousseau, R., & Wolfram, D. (2014). *Measuring scholarly impact: Methods and practice*. Springer.

Dong, Y., Johnson, R. A., & Chawla, N. V. (2015). Will this paper increase your *h*-index? Scientific impact prediction. In *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 149–158). ACM.

Doreian, P. (1982). Maximum likelihood methods for linear models spatial effect and spatial disturbance terms. *Sociological Methods & Research*, *10*(3), 243–269.

Dow, M. M., Burton, M. L., & White, D. R. (1982). Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks*, *4*(2), 169–200.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, *4*(5), e157.

Frosch, D. L., Saxbe, D., Tomiyama, A. J., Glenn, B. A., Low, C. A., Hanoch, Y., Motivala, S. J., & Meeker, D. (2010). Assessing the scholarly impact of health psychology: A citation analysis of articles published from 1993 to 2003. *Health Psychology*, *29*(5), 555.

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.

Ghiasi, G., Larivière, V., & Sugimoto, C. R. (2015). On the compliance of women engineers with a gendered scientific system. *PLOS ONE*, *10*(12), e0145931.

Gilbert, N., Ahrweiler, P., Pyka, A., et al. (2014). *Simulating knowledge dynamics in innovation networks*. Springer.

Giordano, G., & Scepi, G. (2012). Network data as contiguity constraints in modeling preference data. In *Challenges at the interface of data analysis, computer science, and optimization* (pp. 233–241). Berlin, Heidelberg: Springer.

Giordano, G., & Vitale, M. P. (2011). On the use of external information in social network analysis. *Advances in Data Analysis and Classification*, *5*(2), 95–112.

Glänzel, W. (2006). On the opportunities and limitations of the *h*-index. *Science Focus*, *1*(1), 10–11.

Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research*. pp. 257–276. Springer.

Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702.

Haeussler, C., & Sauermann, H. (2013). Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. *Research Policy*, *42*(3), 688–703.

Hajjem, C., Harnad, S., & Gingras, Y. (2005, December). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data Engineering Bulletin*, *28*(4), 39–47.

Hurley, L. A., Ogier, A. L., & Torvik, V. I. (2013). Deconstructing the collaborative impact: Article and author characteristics that influence citation count. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1–10.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The *R*- and AR-indices: Complementing the *h*-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, *61*(2), 424–427.

Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213.

Lawrence, S. (2001). Online or invisible. *Nature*, *411*(6837), 521.

Leem, B.-H., & Chun, H. (2015). Measuring the influence of efficient ports using social network metrics. *International Journal of Engineering Business Management*, *7*.

Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology and Evolution*, *20*(1), 28–32.

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.

Liu, P., & Xia, H. (2015). Structure and evolution of co-authorship network in an interdisciplinary research field. *Scientometrics*, *103*(1), 101–134.

Low, W. Y., Ng, K. H., Kabir, M., Koh, A. P., & Sinnasamy, J. (2014). Trend and impact of international collaboration in clinical medicine papers published in Malaysia. *Scientometrics*, *98*(2), 1521–1533.

McCabe, M. J., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry*, *52*(4), 1284–1300.

Nasini, S., de Albeniz, V. M., & Dehdarirad, T. (2017). Conditionally exponential random models for individual properties and network structures: Method and application. *Social Networks*, *48*, 202–212.

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

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

Opthof, T., Coronel, R., & Janse, M. J. (2002). The significance of the peer review process against the background of bias: Priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovascular Research*, *56*(3), 339–346.

Ortega, J. L. (2014). Influence of co-authorship networks in the research impact: Ego network analyses from Microsoft Academic Search. *Journal of Informetrics*, *8*(3), 728–737.

Pan, R. K., & Fortunato, S. (2014). Author Impact Factor: Tracking the dynamics of individual scientific impact. *Scientific Reports*, *4*.

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., et al. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), 15316–15321.

Piraveenan, M., Prokopenko, M., & Zomaya, A. (2008). Local assortativeness in scale-free networks. *EPL (Europhysics Letters)*, *84*(2), 28002.

Rigby, J. (2011). *Systematic grant and funding body acknowledgment data for publications: An examination of new dimensions and new controversies for bibliometrics. Manchester Business School Research Paper 611*.

Rigby, J. (2013). Looking for the impact of peer review: Does count of funding acknowledgements really predict research impact? *Scientometrics*, *94*(1), 57–73.

Rigg, L. S., McCarragher, S., & Krmenec, A. (2012). Authorship, collaboration, and gender: Fifteen years of publication productivity in selected geography journals. *The Professional Geographer*, *64*(4), 491–502.

Robins, G., Pattison, P., & Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, *66*(2), 161–189.

Rossiter, M. W. (1993). The Matthew Matilda effect in science. *Social Studies of Science*, *23*(2), 325–341.

Ruano-Ravina, A., & Álvarez-Dardet, C. (2012). Evidence-based editing: Factors influencing the number of citations in a national journal. *Annals of Epidemiology*, *22*(9), 649–653.

Smith, M. J., Weinberger, C., Bruna, E. M., & Allesina, S. (2014). The scientific impact of nations: Journal placement and citation performance. *PLOS ONE*, *9*(10), 1–6.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, *81*(1), 177.

Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225.

Teixeira, d. S. J. (2011). The ethics of collaborative authorship. More realistic standards and better accountability are needed to enhance scientific publication and give credit where it is due. *EMBO Reports*, *12*(9), 889.

Uddin, S., Hossain, L., & Rasmussen, K. (2013). Network effects on scientific collaborations. *PLOS ONE*, *8*(2), e57546.

van der Pol, C. B., McInnes, M. D., Petrcich, W., Tunis, A. S., & Hanna, R. (2015). Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? *PLOS ONE*, *10*(3), e0119892.

van Eck, N. J., Waltman, L., van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLOS ONE*, *8*(4), 1–6.

Vanclay, J. K. (2013). Factors affecting citation rates in environmental science. *Journal of Informetrics*, *7*(2), 265–271.

Velden, T., Haque, A.-u., & Lagoze, C. (2010). A new approach to analyzing patterns of collaboration in co-authorship networks: Mesoscopic analysis and interpretation. *Scientometrics*, *85*(1), 219–242.

Vieira, E. S., & Gomes, J. A. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13.

Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology*, *63*(2), 406–415.

Wang, J., & Shapira, P. (2015). Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *PLOS ONE*, *10*(2), e0117727.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, *60*(10), 2107–2118.

Yu, X., Gu, Q., Zhou, M., & Han, J. (2012). *Citation prediction in heterogeneous bibliographic networks*. pp. 1119–1130. *SDM* (Vol. 12) SIAM.