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

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