Journal of Informetrics 11 (2017) 892–903

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Detecting the emergence of new scientific collaboration links in Africa: A comparison of expected and realized collaboration intensities

Raf Guns [a,*], Lili Wang [b]

[a] University of Antwerp, Centre for R&D Monitoring (ECOOM), Middelheimlaan 1, 2020 Antwerpen, Belgium
[b] UNU-MERIT, Boschstraat 24, 6211 AX Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 24 July 2017
Accepted 26 July 2017
Available online 30 August 2017

Keywords:
Scientific research
Collaboration
Networks
Link prediction
Africa
Geographic groups
Outside-Africa collaboration
Within-Africa collaboration

ABSTRACT

The evolution of a collaboration network is to some extent steered by the network topology itself. This is the reason behind the success of network evolution models and approaches to link prediction. At the same time, some changes are due to exogenous factors (i.e., factors external to the network itself). In this paper, we explore changes in the collaboration network of African countries (2000–2014), with a focus on detecting emergent links beyond endogenous network features. Using link prediction and machine learning, we generate an 'expected' (predicted) collaboration network based on past data and compare it with the actual network that evolved in later years. The results show that the intensity of collaboration with non-African countries is higher than expected, especially for countries that are scientifically more active. To a lesser extent, we also find an increase in collaboration within Africa, which seems mostly due to the scientifically less developed countries. Emergent collaborations are mostly found in the first half of the studied period; in the second half the network structure appears to stabilize.

1. Introduction

Lack of research capabilities is a well-known factor impeding the growth of scientific research in Africa. Given that most African countries have difficulties in reaching their R&D targets, i.e. 1% of their respective gross domestic product (AOSTI, 2014), relying on project funding alone may hamper the scientific progress in Africa. International collaborations have become one of the most efficient means for capacity building (AOSTI, 2014). International human resources (i.e. international collaborating scholars) are believed to be of great importance in shaping the research agendas in Africa (Chataway, Smith, & Wield, 2005; Toivanen & Ponomariov, 2011). At the same time, a research agenda that is dictated mainly by international collaboration may have little societal relevance (Kreimer & Meyer, 2008). As pointed out by Arvanitis, Waast, and Gaillard (2000), due to the lack of science policies and the "privatization" of science, scientists in many African countries spend much more time in writing contract proposals and doing contracting work rather than conducting original and fundamental scientific research.

Moed and Halevi (2014) propose a four-stage model of scientific development in which the second stage – building up – is especially characterized by a large share of international collaboration. In the building-up phase of a country's

* Corresponding author.
E-mail address: raf.guns@uantwerpen.be (R. Guns).

http://dx.doi.org/10.1016/j.joi.2017.07.004
1751-1577/

scientific development, international scientific collaborations are often established with developed countries (Moed, 2016). The international links between Africa and other countries are often mediated through cooperative health and agricultural programs (Adams, Gurney, Hook, & Leydesdorff, 2014).

Different collaboration structures indicate different types of knowledge flows. Hence the evolution of a network signals the changes of knowledge resources in the process of collaborating and learning. Although the development of a network is sometimes believed to result from earlier connections (Albert and Barabási, 2000; Powell, White, Koput, & Owen-Smith, 2005), exogenous factors can also direct future collaboration links (Boshoff, 2010; Schubert & Sooryamoorthy, 2010). This raises the question to what extent the evolution of the collaboration network of African countries can be explained by endogenous factors (network topology in the past) and how exogenous factors affect the network at the level of (1) the continent as a whole, (2) geographical groups within Africa, and (3) African countries.

This study aims to explore the changes in the collaboration network of African countries, with a focus on detecting emergent links beyond endogenous network features. Using link prediction and machine learning, we generate an 'expected' (predicted) collaboration network based on past data and compare it with the actual network that evolved in later years. Hence we show to what extent the network has developed based on existing links. Where there are clear deviations from our predictions, we examine where the main emerging collaboration links are located.

This study makes the following methodological contributions:

- Combination of predictors through random forest regression for predicting link weights;
- Use of link prediction techniques to generate 'expected' networks, as a baseline against which to compare actual networks;
- Definition and use of international collaboration intensity as an indicator at the level of countries and geographical groups of countries (geo-groups for short).

2. Literature

2.1. Network prediction

Social networks grow and evolve according to certain regularities. Mechanisms like assortativity, triadic closure, and cumulative advantage (De Solla Price, 1976) have given rise to the existence of different network (growth) models, such as the preferential attachment model (Barabási & Albert, 1999). This indicates that the development of a network is not entirely random and future interactions can often be explained by current network topology, without further information on the actors or social or geographical context. We will refer to these features as *intrinsic* or *endogenous* features of the network, by which we mean that the network structure itself 'encourages' the development of the network in a certain direction (e.g., the establishment of a new link).

Predicted and recommended network connections have been discussed in earlier literature (e.g., Guns & Rousseau, 2014; Liben-Nowell & Kleinberg, 2007; Yan & Guns, 2014; Zhou, Lü, & Zhang, 2009). Liben-Nowell and Kleinberg (2007) propose to use proximity features for the prediction of future links. Guns and Rousseau (2014) present a method for recommending research collaboration partners which significantly improves the accuracy of recommendations. Comparing author-, institution-, and country-level collaboration networks, Yan and Guns (2014) find that collaboration at higher levels of aggregation, especially country-level collaboration networks, tend to yield more accurate prediction outcomes than lower-level ones.

Despite the importance of network topology to the evolution of a network, there are also exogenous factors influencing connections. Provided there are new resources and opportunities, actors are expected to alter their positions and reconfigure the networks (Powell et al., 2005). The propensity of one country's engagement in collaboration can be steered by government mediation, strength of the partners, professional recognition, funding possibilities and infrastructure (Boshoff, 2010; Schubert & Sooryamoorthy, 2010; Wagner & Leydesdorff, 2005).

The existence of such exogenous factors implies that link prediction has inherent limits. Our use of link prediction in this paper is somewhat similar to that of Rattigan and Jensen (2005), who propose exploiting these limits to find 'anomalous' links. However, our contribution moves in several ways beyond that of Rattigan and Jensen (2005): we propose a method to aggregate multiple predictions, take link weight into account (see the methods section for details) and explore exogenous factors at the level of nodes (countries) or groups of nodes (geo-groups or even the whole of Africa).

2.2. Scientific collaboration in Africa

Science production in Africa is characterized by its high intensity of international collaborations and unbalanced partnerships (AOSTI, 2014; Boshoff, 2010; Confraria, Godinho, & Wang, 2017; Jacobs & Pichappan, 2006; Mêgnigbêto, 2013; Sooryamoorthy, 2009) in particular with a low collaboration frequency between African countries and a high collaboration frequency with non-African countries.

Collaborative networks are widely recognized to be helpful in providing knowledge spillovers and for African countries to advance their innovation systems (Toivanen & Ponomariov, 2011). Using a set of publications of South African scientists, Sooryamoorthy (2009) reports that collaboration has a positive impact on the productivity of South-African scientists. The collaborative partners from other geographical locations form sources of knowledge for the African countries to learn from.