Journal of Informetrics 7 (2013) 371–378

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

On time-varying collaboration networks

Matheus P. Viana[a,*], Diego R. Amancio[b], Luciano da F. Costa[b]

[a] Developmental and Cell Biology, University of California Irvine, Irvine, CA, United States
[b] Institute of Physics of São Carlos, University of São Paulo, P.O. Box 369, Postal Code 13560-970, São Carlos, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form
14 November 2012
Accepted 10 December 2012
Available online 4 February 2013

Keywords:
Collaborative networks
Time-varying networks
Complex networks
Pattern recognition

ABSTRACT

The patterns of scientific collaboration have been frequently investigated in terms of complex networks without reference to time evolution. In the present work, we derive collaborative networks (from the arXiv repository) parameterized along time. By defining the concept of affine group, we identify several interesting trends in scientific collaboration, including the fact that the average size of the affine groups grows exponentially, while the number of authors increases as a power law. We were therefore able to identify, through extrapolation, the possible date when a single affine group is expected to emerge. Characteristic collaboration patterns were identified for each researcher, and their analysis revealed that larger affine groups tend to be less stable.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The progressive and inexorable informatization of scientific publishing has implied several important consequences, including the possibility to quantify and analyze the patterns characterizing scientific collaborations. For instance, many efforts have been dedicated to the identification of citations between articles (e.g. Amancio, Nunes, Oliveira Jr., & Costa, 2012; Amancio, Oliveira Jr., Costa, 2012a,b; Chen & Redner, 2010; Persson, 2010; Roth, Wu, & Lozano, 2012). Another well-developed approach involves mapping and studying collaborations between researchers (e.g. Liljeros, Edling, Amaral, Stanley, and Aaberg (2001), Newman (2001, 2004), Shrum, Chompalov, and Genuth (2001), Hsiang and Rebecca (2012)). Such works are often done by using complex networks (Costa, da, Rodrigues, Travieso, & Villas Boas, 2007). In the case of collaboration networks, each researcher is mapped as a node, while the joint authorships establish the links between those nodes. However, most such efforts disregards time, in the sense that the citation and collaborations are taken along long periods of time. By doing so, important information about transient patterns of collaboration are overlooked. For instance, some collaborations are more likely to follow an intermittent pattern, while others would be expected to proceed along continuous periods of time.

The current work aims precisely at addressing this important issue, which has been accomplished by parameterizing the collaboration networks explicitly along time. So, instead of a single network, we derive a sequence of networks defined from a starting time up to the present moment (i.e. our networks are cumulative). For each node i in each of such parameterized networks, we define its respective *affine group*, corresponding to two sets of nodes. First, we identify those nodes that are directly attached to i, as they are co-authors. The second set of nodes corresponds to those that belong to the same community (Girvan & Newman, 2002) as node i, and therefore represents those authors that are more closely interrelated. Having obtained the time-parameterized networks and the respective affine groups, we proceed to analyze the evolution of the latter along time. More specifically, we calculate the mean size of the affine groups along time for three different collaboration networks extracted from the arXiv repository (http://www.arXiv.org). Remarkably, we found that these sizes

* Corresponding author.
E-mail address: vianamp@gmail.com (M.P. Viana).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.12.005

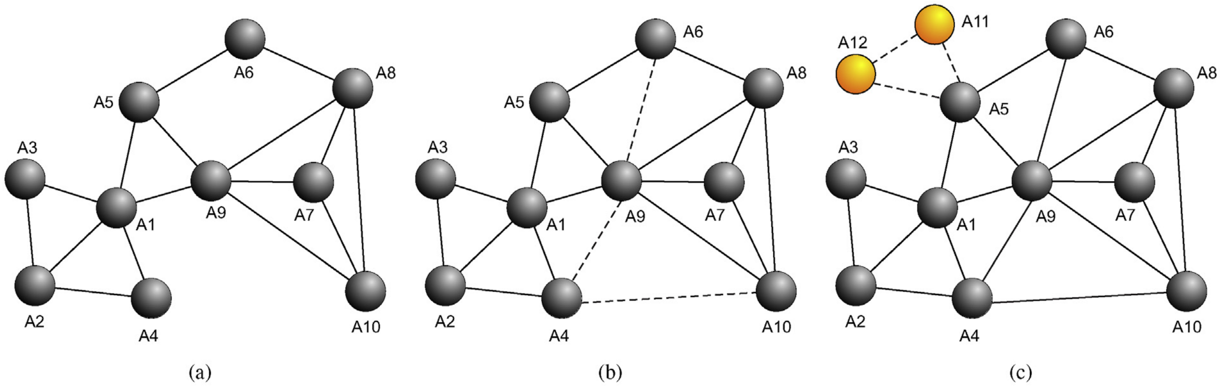

Fig. 1. Example of growing of collaborative networks from $t = t_0$ to $t = t_0 + 2\Delta t$. The toy database comprises 12 authors and 10 papers. (a) Collaborative network at $t = t_0$ built from the following list of 7 papers and 9 authors (A*X* represents Author *X*): (i) paper 1 (A1, A2 and A3); (ii) paper 2 (A1 and A2); (iii) paper 3 (A1, A2 and A4); (iv) paper 4 (A5 and A6); (v) paper 5 (A7, A8, A9 and A10); (vi) paper 6 (A1, A5 and A9); (vii) paper 7 (A6 and A8). (b) Collaborative network at $t = t_0 + \Delta t$ when paper 8 (A6 and A9) and paper 9 (A9, A4 and A10) are included. New edges are represented as dotted lines. (c) Collaborative network at $t = t_0 + 2\Delta t$ when paper 10 (A5, A11 and A12) is included. New nodes are represented as orange nodes. (For interpretation of references to color in this figure legend, the reader is referred to the web version of this article.)

scale as an exponential with different exponents, while the number of authors in the respective networks grows slower, as a power law. We also found that different affine groups tend to exhibit rather distinct intermittence patterns, which suggested a classification of the authors according to their time-dependent collaboration patterns. So, for each author, we calculated the maximum size of the affine groups to which it belonged, as well as the average duration of the respective collaborations. These findings suggest that authors who collaborate with more people also tend to have shorter collaborations.

2. Methodology

2.1. The time-varying collaboration network

The following procedure was applied in order to represent the relationship between authors in a specific topic. Let $A = \{a_{ij}\}$ be the matrix representing the undirected and unweighted network. If authors *i* and *j* collaborate on at least one paper from the database, then a link between them is established so that $a_{ij} = 1$. Otherwise, $a_{ij} = 0$. Fig. 1 serves as a gist of how the collaborative networks are constructed. Note that at every instant of time, new edges and new nodes might be included in the network.

We built three collaboration networks using the arXiv repository. Each network was built based on papers about an specific topic. We adopted the criteria employed in (Amancio, Nunes, et al., 2012; Amancio et al., 2012a,b: given a keyword, we selected all papers in arXiv which contain this keyword in title or abstract. The keywords chosen were *complex networks*, *graphene* and *topological insulator*. For simplicity's sake we call the respective networks of COMPNET, GRAPHENE, and TOPINSU. These three topics have been chosen for they represent modern topics of current interest in the area of Physics. Specifically, one network was obtained for each year of the aforementioned networks and the evolution of collaborative groups of authors was studied in terms of the time-varying collaboration networks. Details regarding the networks are given in Table 1.

3. The affine group

Here we define the main concept in this paper, i.e. the *affine group*. For each author *i* belonging to the set Λ of authors we aim at identifying the subset of authors which are potentially interested in the same subject of research. The most natural choice of authors to belong to the affine group of *i* are the current or previous collaborators of *i*, i.e., the set $V_i(t) = \{j \in \Lambda \mid a_{ij}(t) > 0\}$. Obviously, authors possibly interested in the same subject may never have collaborated in the past. To consider this case we used the concept of community (Girvan & Newman, 2002) in networks. A community is a subnetwork (i.e., a group of nodes) that is more densely connected internally than with the other nodes of the network. Formally, a

Table 1
Database and network statistics. *P* represents the set of papers, *N* represents the number of authors and *m* is the number of edges. λ corresponds to the value of the parameter used to fit the exponential growth of the average size of affine groups.

Network	*P*	*N*	*m*	λ
COMPNET	1316	2013	5342	0.56
GRAPHENE	4468	6490	24,956	1.09
TOPINSU	778	1436	5537	1.87

Download English Version:

https://daneshyari.com/en/article/523233

Download Persian Version:

https://daneshyari.com/article/523233

Daneshyari.com