Physica A 413 (2014) 220–229

Contents lists available at ScienceDirect

Physica A

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

Patterns of collaboration in four scientific disciplines of the Turkish collaboration network

Abdullah Çavuşoğlu [a], İlker Türker [b,*]

[a] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Yıldırım Beyazıt University, Ankara, Turkey
[b] Department of Computer Engineering, Faculty of Engineering, Karabük University, Karabük, Turkey

HIGHLIGHTS

- We constructed a nationwide scientific collaboration network in four scientific disciplines.
- By the years pass, all the subsets display the tendency of co-authoring with more authors per paper.
- The perfect power-law consistence of the interdisciplinary network in degree distribution is best copied by the mathematics subset.
- The surgery subset tends to deviate from the exponential growth rate, while others are in good consistency with exponential fits.
- The Matthew Effect is observed in career longevity distributions.

ARTICLE INFO

Article history:
Received 14 December 2013
Received in revised form 9 April 2014
Available online 8 July 2014

Keywords:
Scientific collaboration
Networks
Small worlds
Turkey
Career longevity

ABSTRACT

Scientific collaboration networks, as a prototype of complex evolving networks, are studied in many aspects of their structure and evolving characteristics. The organizing principles of these networks also vary in different scientific disciplines, demonstrating that each discipline has specific connecting rules. Retrieving the co-authorship data from the ISI Web of Science, we constructed networks of four disciplines (engineering, mathematics, physics and surgery) as a subset of the Turkish scientific collaboration network spanning 33 years' data. To provide a comparative perspective on the network topologies, we studied some statistical and topological properties such as the number of authors, degree distributions, authors per paper and papers per author histograms and distributions. These properties yield that the rapid growth of high education in Turkey (i.e. doubling of the number of universities and students within the last decade) had boosted the number of publications and increased the level of collaborations in the scientific collaboration networks. We showed the occurrence of Matthew effect in career longevity distributions, and also outlined the Heaps' law relation in the scaling of the collaborations as well. We outlined the prominent properties of each subset, while the similarities and deviations from the interdisciplinary networks are also evaluated.

1. Introduction

Scientific collaboration networks are accepted as close prototypes of complex evolving networks. They are recently studied not only for bibliographic means or social sciences, but also for understanding the underlying mechanisms governing the structure of complex networks [1–3]. While traditionally complex networks have been modeled as random graphs, it is

* Corresponding author. Tel.: +90 5317936507; fax: +90 3704336604.
E-mail addresses: abdullah.cavusoglu@ybu.edu.tr (A. Çavuşoğlu), iturker@karabuk.edu.tr, ilkerturker@gmail.com (İ. Türker).

http://dx.doi.org/10.1016/j.physa.2014.06.069

increasingly recognized that the topology and evolution of real networks are governed by robust organizing principles [4]. These organizing principles are responsible for some variety of outputs such as the fast availability of information within the world-wide-web [5,6], the fast spreading of epidemics [7–9,3], evolution of cooperation [10,11] and some other aspects of natural relations.

Among the featuring results of the studies that involve these organizing principles, the Matthew "rich get richer" effect is a significant one outlined by testing whether the past experience and the career longevity lead an advantage for the further steps of one's career [12]. Preferential attachment is an expected result of this effect, that nodes link with higher probability to those nodes that already have a larger number of links [1,10]. Another aspect of organization is the need of the network for the new nodes. Studying the word population in millions of books published in the past two centuries, it is outlined that the need for new words decreases while the languages grow [13], yielding that the network velocity would decrease in a complex network that maturates. In fact, this is evident in many real-world networks at which the rate of links added is different from the rate at which nodes are added [14], and is proposed to be the indicator of a network's functionally organized by the integrated activity of its nodes [15]. Beyond the interactions between nodes, an impressive amount of work has been done on community detection in complex networks, which in turn enables the realistic modeling of networks' evolution [16,17]. These communities also interact with each other where high numbers of between-group links simply unify the two groups and make them act as one, while too rare between-group links facilitate the information flow between the two groups [18,19].

The growth and evolution characteristics of complex networks generally obey the Zipf's law [13,20,21] quantifying the degree distribution of nodes. Also the Heaps law [13,22–24], relating the size of a corpus to the number of nodes (words), is observed in such systems of expanding languages where the words account for the nodes of the complex network. These classic parameters are the common ingredients of the complex network studies.

To uncover the structure of complex networks, several datasets are available including the world-wide-web, Internet, movie actor collaboration networks, scientific collaboration networks, electric power networks etc. [4]. Among these datasets, the scientific collaboration databases stand out with the feature that every link between the nodes (authors) are captured in the time domain by the publication date of the relevant paper they co-authored together. This leads tracking the dynamic evolution of the network explicitly [1]. Also, the fact that authors choose their collaborators with their own decisions makes these networks more analogous with the natural networks [1,4].

Recently, the scientific collaboration networks are studied not only for whole national databases [10,25,26], but also for distinct scientific disciplines [1,2,27–29]. In the case study of Turkish scientific collaboration network, we have uncovered the network dynamics of the whole national scientific collaboration network for a 30 years' time span and stated out the similarities and differences with the other studies in the literature [26]. To provide an inter-disciplinary perspective to that whole network study, we supplied the scientific collaboration network data from the "Web of Science" for four separate scientific disciplines; engineering, mathematics, physics and surgery for 33 years of time span covering the years from 1980 to 2012. Starting with the growing rate of these networks, we uncovered the *degree distributions*,*authors per paper* and *papers per author* distributions as well. We showed that the Matthew effect emerges in our dataset that the cumulative experience turns into an advantage favoring to move forward in the career. We also calculated the growth fluctuations of the number of nodes which has a decreasing trend in two of four databases as the number of collaborations increase, indicating a slowdown in node generation while the networks expand. To display disciplinary deviations, we presented the mentioned distributions/values comparatively with each other and the available ones for the whole network as well.

2. Presentation

We obtained the scientific collaboration dataset from the Web of Science consisting of the publications addressed from Turkey between years 1980 and 2012. We constructed the scientific collaboration networks for either the interdisciplinary (whole) network data or the four disciplines separately. The raw data was parsed into a structural database, having each author as a separate node and preserving co-authorships as defined links between node pairs. Since the links are generated using the data including the publication date, the evolution of the network in time is traceable, providing a timeline view of the desired parameters rather than its final state.

In our recent study mentioned above, we obtained a valuable view to the whole scientific collaboration network of Turkey, stating out a perfect power-law degree distribution, accelerating growth property, increasing collaboration trend etc. [26]. These results motivated us to investigate whether the subsets of the whole network (i.e. scientific disciplines) have differential effects on this exceptional growth rate or degree distribution, or whether some of these subsets play an attractive-point role on some outputs. (The reader is referred to our recent study in Ref. [26] for the comparisons with the interdisciplinary network parameters.)

We first present the network size as in Fig. 1, where the right panel shows the log-linear plot of the same data. Fig. 1(a) shows that all subsets are in increasing phase while physics and surgery subsets reach approximately the same number of authors. In the right panel, applied exponential fit with exponent 0.17 shows similarity to the recently reported characteristics of the whole network (slope: 0.16) [26].

The major indicator about the network topology, the degree distribution, is presented in Fig. 2. Some complex structures exhibit degree distributions consistent with the Zipf's Law [20,21,30], that the frequency of a given node is inversely