Chaos, Solitons and Fractals 91 (2016) 386–392

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Detailing the co-authorship networks in degree coupling, edge weight and academic age perspective

İlker Türker[a,*], Abdullah Çavuşoğlu[b,1]

[a] Dept. of Computer Engineering, Karabük University, Demir Çelik Campus, Karabük, Turkey

[b] Member of the Council for Higher Education (YÖK), Council for Higher Education, Bilkent, Ankara, Turkey

ARTICLE INFO

Article history:
Received 26 May 2016
Revised 26 June 2016
Accepted 27 June 2016

Keywords:
Complex networks
Scientific collaboration networks
Link strength
Academic age
Degree coupling

ABSTRACT

Scientific collaboration networks are good resources for understanding self-organizing systems, reflecting the main generic properties like clustering, small-world and scale-free degree distribution. Beyond discovering the evolution of main parameters, we aimed to uncover the microscopic wiring properties in this study. We focused on the degree circumstances of pairing nodes together with degree differences, academic age differences and link weights. Analyzes are visualized by single distribution plots of the network parameters together with the 2D coupling characteristics of these parameters with a logarithmic colorbar as a third dimension, drawing visual perspective presenting "who prefers connecting to whom" during the network evolution. We showed that majority of the edges in the co-authorship network connects the nodes of comparable degrees and academic ages, featuring that strong collaboration activities occur between comparable academic careers. We also stated out that beyond the node degree distributions, power-law regimes are also observed in link weight and degree difference distributions.

1. Introduction

Complex networks as an emerging framework of science, is being studied across many fields [35]. Describing real systems with interactions between its components provide a view for understanding the evolution of their structural and behavioral characteristics. The elements of these systems are named as vertices (or nodes) and the edges represent the interactions between these vertices [1,4]. Some variety of real systems modeled with complex networks are biological oscillators, neural networks, spatial games, genetic control networks, food webs, computer networks, power grid networks, linguistic networks, social networks like scientific collaboration networks, the network of film actors, epidemic networks and many other self-organizing systems [5,28,40]. These type of real networks generally display the same generic properties like small-world property (one can find a short path between most pairs of nodes) or scale-free degree distribution (the node degree distribution follows a power-law decay) [2,5,9,38]. Strong consistencies in main statistical parameters indicate that these diverse systems in fact display similar organizing principles, which should be encoded in their topology [1].

Imitating a self-organizing system is only performable with a broad survey of several real networks. Among these networks, scientific collaboration networks (SCN) are noteworthy because of the nodes (individuals) decide to collaborate with one other by his/her own choice, and the collaboration records are available in time domain with the resolution defined by publications [22]. By this view, one can say that SCNs are a good resource for learning about self-organizing systems that evolve in time.

SCNs are dynamic systems. They expand by the addition of new nodes (authors), where the existing nodes also interact via new links by coauthoring a new paper together. Focusing on this evolving property in time, Barabasi et al. [4,5] and Newman [22,23] stated out the clustering and preferential attachment properties of these systems as well as the other generic properties mentioned above. Clustering means that the neighbors of a specific node are also neighbors of each other when the links to the initial node are removed. On the other hand the nodes prefer to connect to high-degree (popular) nodes rather than the low-degree ones, comprising the preferential-attachment property that also plays role on power-law degree distribution.

Like the pioneering studies about SCNs mentioned above, a number of studies focus on co-authorship data spanning some specific disciplines like mathematics, engineering, physics, surgery, neuroscience, genetics etc. [1,4,5,8,22–26,34,36] or interdisciplinary properties of SCNs [19]. Narrowing the dataset to a specific journal database is another approach [14]. But spanning a nationwide

* Corresponding author: Fax: +90 370 4333290.
E-mail address: iturker@karabuk.edu.tr (İ. Türker).

[1] All of the authors equally contributed to the paper.

http://dx.doi.org/10.1016/j.chaos.2016.06.023

collaboration network is attractive in both statistical physics and bibliographic view. Studies in this scope cover national co-authorship networks of Slovenia [19,29,7], Turkey [8,11], Europe [13], China [20] etc. Like the national co-authorship networks, the international SCNs [3,18,27,39] are also point of interest by the view of monitoring the effect of geographic borders and linguistic differences on wiring properties.

As mentioned above, the massive studies about SCNs firstly aim to understand self-organizing systems. The output of these complex network analyzing studies are evaluated as the input of modeling studies, improving the consistency of these models with the real world networks. While the growing property driven by preferential attachment is imitated, the internal link generation processes should empower the clustering property as well. Such a modeling study of Li et al. [17] uses econophysicists collaboration network as a resource for determining network dynamics, then use an evolving model with internal link generator that imitates evolution of link weights. While linear preferential attachment is successful for power-law degree distribution as a global rule, Vazquez [37] designed a growing network model based on local rules. This study originated from the detailed analysis of five real networks (including a mathematicians SCN), especially the distributions of some network parameters as a function of vertex degree. These detailed distributions provide a broad view of microscopic mechanisms that take role in network evolution like degree correlations of internal links. These internal links either improve an existing cluster of nodes, or form new clusters as microscopic actors. Lee et.al outlined that the co-authorship networks evolve through three common major processes in time: the nucleation of small isolated components, the formation of a treelike giant component through cluster aggregation, and the entanglement of the network by large-scale loops [15]. This ingredients show us that macroscopic properties like the giant component must also be tracked during network evolution. Another interesting discovery is that power-law consistency is not only observed in degree distributions, but also observed in link strength distributions for three distinct co-authorship networks [16]. Lastly, we must mention the Matthew effect (first mover advantage) that takes role in complex networks, described as "rich gets richer" or "advantage begets further advantage" ([8,10,21,30–33]).

All these studies conduct that beyond defining a global preferential attachment, there are something special about generating new links through imitating a real network. More realistic network models need more accurate analyzing procedures. Motivated by this view, we aimed a comprehensive analysis of a national (Turkish) co-authorship network of 33 years through a scientific discipline (engineering). Charged with our previous work about general network parameters [7,8], we focus on the wiring principles that drive evolution of a real network. In this context, we studied degree circumstances in link construction processes as weighted and unweighted separately, link strength distributions, first mover advantage taking role in wiring procedures, source and target node degree correlations (who links to whom), academic age differences in link, together with the evolution of basic network parameters in time.

2. Methods

Our network is based on data collected from ISI Web of Science, filtered for publications which have Turkish authors, spanning years 1980 to 2012 in engineering field. The publications having Turkish authors also include a small numbers of foreign co-authors that we also included in the network. Parsing the raw data that consists of 37,403 publications, we constructed a co-authorship network of 29,548 authors (nodes) that are interconnected with 131,822 links. Since the "Turkey" addressed publications are collaborated by also foreign authors, 8041 foreign authors participate in the network, corresponding to 27.2% of the whole node set. Some basic distributions about this data are outlined in our previous work [8] together with the emergence of growing property and the Matthew effect. In this study, we detailed the wiring properties and degree correlations between the authors to outline "who prefers connecting to whom" in the network.

We constructed the co-authorship network by defining the authors as nodes, and the collaborating actions within a scientific paper as links. We used Gephi [6] for straight analyzing procedures, and also developed custom software to perform the unusual computations in C# programming language. By the way we could perform the unusual computations like finding the academic age or degree differences of the pairing nodes, or calculating the average degrees for the nodes with respect to the first moving year.

Since there are some known issues that affect the data quality in the ISI Web of Science, we processed a hand-driven procedure to assign the abbreviations of the authors to the best suitable one, regarding the previous publications of the authors and the research areas as well. The nationalities of the authors are also determined with this procedure, since an automated data retrieval procedure did not give appropriate results for this attribute.

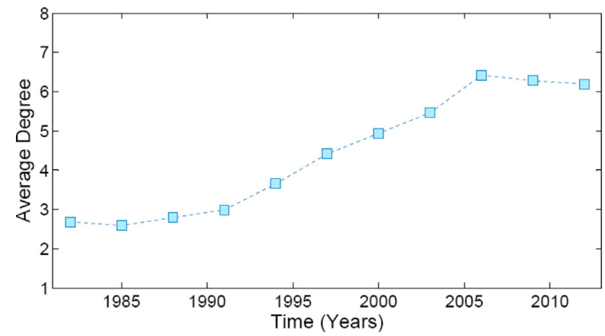

Fig. 1. Time evolution of average degree for the co-authorship network.

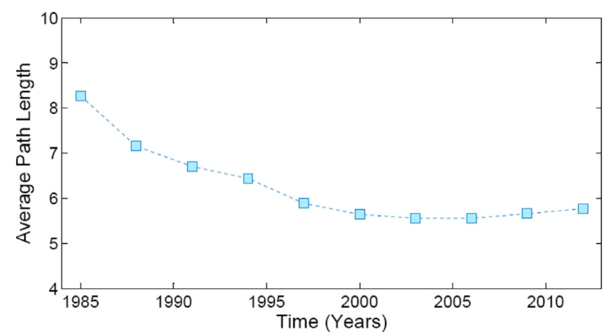

Fig. 2. Time evolution of average separation for the co-authorship network.

3. Results

We start by drawing the evolution of the basic network parameters in time. Figs. 1–3 show the time dependencies of average degree, average separation and average clustering coefficient respectively.

Fig. 1 shows that average degree tends to increase linearly, while the right side implies a saturation range. Average path length tends to converge to 6 in Fig. 2, consistent with the small-world phenomena known as six-degrees of separation. The global clustering coefficient displays a rising trend in Fig. 3, yielding that the network gets more inter-connected in time. All these parameters and time evolution characteristics conform to the previous work mentioned above.