Available online at www.sciencedirect.com
ScienceDirect
Physica A 372 (2006) 333–339
PHYSICA A
www.elsevier.com/locate/physa

A topological analysis of scientific coauthorship networks

Alessio Cardillo[a], Salvatore Scellato[b], Vito Latora[a,*]

[a]Dipartimento di Fisica e Astronomia, Università di Catania, and INFN sezione di Catania, Via S.Sofia 64, 95125 Catania, Italy
[b]Scuola Superiore di Catania, Via S. Paolo 73, 95123 Catania, Italy

Available online 18 September 2006

Abstract

We study coauthorship networks based on the preprints submitted to the Los Alamos cond-mat database during the period 2000–2005. In our approach two scientists are considered connected if they have coauthored one or more cond-mat preprints together in the same year. We focus on the characterization of the structural properties of the derived graphs and on the time evolution of such properties. The results show that the cond-mat community has grown over the last six years. This is witnessed by an improvement in the connectivity properties of coauthorship graphs over the years, as confirmed by an increasing size of the largest connected component, of the global efficiency and of the clustering coefficient. We have also found that the graphs are characterized by long-tailed degree and betweenness distributions, assortative degree–degree correlations, and a power-law dependence of the clustering coefficient on the node degree.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Structure of complex networks; Social systems

A social network is defined by a set of *actors*, mostly individuals or organizations, and a set of *ties* between couples of actors. It describes how the actors are connected through various social relationships ranging from casual acquaintance to close family bonds [1,2]. Social network analysis has emerged as a key technique in modern sociology, anthropology, social psychology and organizational studies, as well as a popular topic of speculation and study. Research in a number of academic fields has demonstrated that social networks operate on many levels, from families up to nations, and play a critical role in determining the way problems are solved, organizations are run, and the degree to which individuals succeed in achieving their goals. The shape of the social network helps determining a network's usefulness to its individuals. Networks with many weak ties [3] are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties. That is to say that tight groups of friends share the same knowledge and opportunities, while a group of individuals with connections to other social worlds is likely to have access to a wider range of information. It is better for individual success to have few connections to a variety of networks rather than many connections within a single network. Similarly, individuals can exercise influence or act as brokers within their social networks by bridging two networks that are not directly linked [4].

These considerations can be interestingly applied to *scientific collaboration networks* (referred as SCNs from now on), a particular kind of social networks whose actors are scientists and the investigated relationships are

*Corresponding author.
E-mail address: Vito.Latora@ct.infn.it (V. Latora).

0378-4371/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2006.08.059

scientific collaborations. One way to define the existence of a scientific collaboration is through scientific publications: two scientists are considered connected if they have coauthored one or more publications together. As indicated in Refs. [5–8], this appears to be a useful and reasonable definition of scientific acquaintance, for people who have been working together will know each other quite well and are more likely to set up a continuative collaboration and therefore contribute to a knowledge spread, particularly if two related scientists belong to different fields (e.g. physics and computer science). Furthermore, data related to coauthorships can be easily found on the huge publication records that are now accessible on the Internet, and offer one of the largest and most precise database to date on social networks. Focusing on SCNs by using data extracted from the publication records is not a new topic: one of the most famous result of this interest is the Erdös number (see, for instance, the Erdös Number Project [9]), which is a number assigned to each mathematician indicating the number of steps in the shortest path to the incredibly prolific Hungarian mathematician Paul Erdös on the relative SCN.

Here, we present a study of a SCN constructed by using data drawn from the Los Alamos e-Print cond-mat Archive at the website `http://xxx.lanl.gov/archive/cond-mat`. Following some previous works by Newman [5–8], and Barabási et al. [10], we construct the network by considering two scientists connected if they have coauthored one or more cond-mat preprints together in the same year. In particular, we focus on the cond-mat database in the period from 2000 to 2005, inclusive, in order to study how the pattern of collaborations have changed over time in the most recent years.

In Table 1 we report, year by year, the number of papers submitted to the archive and the average number of authors per paper. An important thing to be noticed is that the databases include also the *cross listings* papers (and authors). Such papers do not belong directly to the cond-mat archive, but they are listed there and so they have been included in the analysis. The basic properties of the six graphs under study are reported in Table 2. We notice that both the number of nodes N (the number of different authors per year), and the number of links K in the graph, increase over the years. The number of scientists who submitted at least one paper to the cond-mat archive has almost doubled in the period considered. This number is, in fact, equal to $N = 9077$ in year 2000, and equal to $N = 15\,964$ in 2005. A similar monotonic increase over the years has been observed in the number of edges, and also in the average degree, i.e., the number of different collaborators per author. This might be due to an effective growth of the cond-mat community in the past six years (as denoted for instance by the increasing number of authors per paper shown in Table 1), but also to the fact that an increasing number of scientists are getting used to submit their manuscripts to the e-Print archives. The six graphs under study show a rather high value of efficiency E [11,12] and clustering coefficient C [4], and a rather small value of the characteristic path length $\langle l \rangle$ [13]. The values of C we have found are about two times larger than the values reported by Newman in Ref. [6]. This is probably due to the fact we are considering smaller graphs. The improvement in the connectivity properties of coauthorship networks over the years is confirmed by the increasing size of the largest component [14], and by the increasing value of global efficiency. It is worth to notice that the value of the efficiency increases of about 90% in the period considered (from $E = 0.043$ in 2000 to $E = 0.071$ in 2005). Conversely, a measure of the local properties of the graph, such as the clustering coefficient C, exhibits only a slight increase by less than 5%. The behavior of $C(k)$ is shown in Fig. 1 for year 2000 and year 2005, and indicates that authors with few collaborators are more likely to work within groups in which all the scientists collaborate together (high clustering) than authors with a large degree, usually collaborating with a large number of scientists (eventually belonging to different universities and research groups) not having direct scientific collaborations one with the other. Moreover, as indicated in Fig. 1, the clustering coefficient averaged over nodes of the same degree decays approximatively as

Table 1
Some fundamental information on the Los Alamos cond-mat Archive year by year in the period 2000–2005

	2000	2001	2002	2003	2004	2005
Total number of papers	6581	7616	8395	9096	9882	10 220
(Cross listings)	(556)	(600)	(627)	(728)	(862)	(985)
Mean authors per paper	2.94	3.20	3.11	3.23	3.32	3.37

Download English Version:

https://daneshyari.com/en/article/975817

Download Persian Version:

https://daneshyari.com/article/975817

Daneshyari.com