Journal of Informetrics 4 (2010) 475–482

Contents lists available at ScienceDirect

Journal of Informetrics

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

Growth and structure of Slovenia's scientific collaboration network

Matjaž Perc*

Department of Physics, Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia

ARTICLE INFO

Article history:
Received 17 March 2010
Received in revised form 18 April 2010
Accepted 22 April 2010

Keywords:
Scientific collaboration
Networks
Preferential attachment
Small worlds
Slovenia

ABSTRACT

We study the evolution of Slovenia's scientific collaboration network from 1960 till present with a yearly resolution. For each year the network was constructed from publication records of Slovene scientists, whereby two were connected if, up to the given year inclusive, they have coauthored at least one paper together. Starting with no more than 30 scientists with an average of 1.5 collaborators in the year 1960, the network to date consists of 7380 individuals that, on average, have 10.7 collaborators. We show that, in spite of the broad myriad of research fields covered, the networks form "small worlds" and that indeed the average path between any pair of scientists scales logarithmically with size after the largest component becomes large enough. Moreover, we show that the network growth is governed by near-liner preferential attachment, giving rise to a log-normal distribution of collaborators per author, and that the average starting year is roughly inversely proportional to the number of collaborators eventually acquired. Understandably, not all that became active early have till now gathered many collaborators. We also give results for the clustering coefficient and the diameter of the network over time, and compare our conclusions with those reported previously.

1. Introduction

The structure of social networks is paramount for understanding the spread of knowledge, cultural traits, disease, as well as many other entities and attributes that can be associated with individuals living in groups or societies. As such it has been the subject of intense investigation, both theoretical as well as empirical, for at least half a century (Barabási, 2002; Christakis & Fowler, 2009; Wasserman & Faust, 1994; Watts, 1999). Primarily, and in many ways not really surprisingly, however, these investigations were in the domain of social rather than natural sciences. Probably best know in this context is the study by Milgram (1967), who studied how many steps it took, on average, to get a letter from a randomly selected person to a stockbroker in Boston, who was a friend of Milgram's. The result was six – a number that has since been reused outside of science for a number of purposes, one of the latest examples being the launch of SixDegrees.org seeking to exploit the "small-world phenomenon" for charitable purposes. A shortcoming of the study of Milgram, as well as that of many others conducted in a similar fashion, is that the size and structure of social networks mapped in such a direct and labor intensive way is rather small and receptive to bias. The advent of large-scale online portals made it possible to test the "six degrees of separation" hypothesis more thoroughly. Remarkably though, a study performed by Leskovec and Horvitz (2008), encompassing some 30 billion conversations from 240 million people, reported that the average path length among Microsoft Messenger users is 6.6. Although being closer to seven than six, the number is nevertheless in a strikingly good agreement with the result by Milgram obtained over 40 years ago.

* Tel.: +386 41 981 460.
E-mail address: matjaz.perc@uni-mb.si.
URL: http://www.matjazperc.com/.

1751-1577/$ – see front matter
doi:10.1016/j.joi.2010.04.003

In natural sciences the attention to networks was sparked by works such as those of Watts and Strogatz (1998) and Barabási and Albert (1999), making ground-breaking advances with regard to our understanding of the "small-world phenomenon" and the emergence of scaling via growth and preferential attachment, respectively. The two works, along with subsequent refinements of the concepts they introduced (Barthélemy & Amaral, 1999; Newman, Moore, & Watts, 2000; Krapivsky, Redner, & Leyvraz, 2000; Dorogovtsev, Mendes, & Samukhin, 2000; Dorogovtsev & Mendes, 2000; Amaral, Scala, Barthélemy, & Stanley, 2000; Krapivsky & Redner, 2001; Krapivsky, Rodgers, & Redner, 2001), spawned an impressive number of studies on networks, as evidenced by the many reviews (Newman, 2000; Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002; Newman, 2003; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Dorogovtsev, Goltsev, & Mendes, 2008) and books (Dorogovtsev & Mendes, 2003; Pastor-Satorras & Vespignani, 2004; Newman, Watts, & Barabási, 2006; Barrat, Barthélemy, & Vespignani, 2008) dedicated either specifically to this field of research or its many interdisciplinary variations. This is all the more impressive since, at least within the hard sciences, prior to the late 1990s a paper on network theory is hard to come by (Newman, 2009). For a field this young the volume of insightful findings that have accumulated until now is something to be reckoned with. For example, the structure of networks has been found crucial for their resilience to error and attack (Albert, Jeong, & Barabási, 2000; Cohen, Erez, ben Avraham, & Havlin, 2000, 2001; Callaway, Newman, Strogatz, & Watts, 2000; Pietsch, 2006), for the fast availability of information within the world-wide-web (Albert, Jeong, & Barabási, 1999; Pastor-Satorras, Vázquez, & Vespignani, 2001), uninterrupted supply with electricity (Albert, Albert, & Nakarado, 2004), fast spread of epidemics and viral infections (Pastor-Satorras & Vespignani, 2002; Zanette & Kuperman, 2002; Barthélemy, Barrat, Pastor-Satorras, & Vespignani, 2004; Colizza, Barrat, Barthélemy, Valleron, & Vespignani, 2007), robust and near flawless reproduction of organisms (Hartwell, Hopfield, Leibler, & Murray, 1999), the evolution of cooperation (Santos & Pacheco, 2005; Szabó & Fáth, 2007; Perc, 2009) and coevolution (Gross & Blasius, 2008; Perc & Szolnoki, 2010), the dynamics of social systems (Castellano, Fortunato, & Loreto, 2009), and surely many other aspects of everyday life.

An interesting and potentially very revealing subset of complex networks are the social networks, of which scientific collaboration networks are a beautiful example (Newman, 2001d,b,c, 2004). Notably, for a social network to be representative for what it stands – an account of human interaction – a consistent definition of acquaintance is important. And while it may be challenging to *define* a friendship or an enemy in a consistent and precise manner (Moody, 2001; Moody & White, 2003), scientific collaboration is accurately documented in the final product and thus fairly straightforward to assess. Also amenable to a precise definition of connectedness are movie actors (Amaral et al., 2000), electric grids (Watts & Strogatz, 1998; Albert et al., 2004) and airports (Guimera, Mossa, Turtschi, & Amaral, 2005), for example, yet these are either approximations of social networks in that they don't really document real human contact or the level of acquaintance between people forming them is difficult to determine. As argued convincingly by Newman (2001d), considering scientific collaboration networks alleviates these problems to a large extend.

Here we study the evolution of a scientific collaboration network, namely that formed by Slovenia's scientists, from its very beginnings in the 1960s until the present time. Covering a time span of 50 years, the data are unique in that they provide an excellent testing ground for the "small-world" and preferential attachment hypotheses in growing social networks. We tackle these issues similarly as outlined in previous studies on growing scientific collaboration networks (Newman, 2001a; Jin, Girvan, & Newman, 2001; Barabási, Jeong, Néda, Ravasz, Schubert, & Vicsek, 2002; Jeong, Néda, & Barabási, 2003; Moody, 2004), where it has been reported, for example, that the growth is governed by linear or sublinear preferential attachment, and that as the networks grow their average degree increases while the average distance between individuals decreases. Evidences for strong clustering and models describing the growth of social networks have been presented in this context as well. Notably, for a set of different yet static scientific collaboration networks, Newman (2001d) has shown that the average distance between different authors scales logarithmically with size. We come to results that are in agreement with these earlier observations, but for a single growing scientific collaboration network. Moreover, we show that the observed near-linear preferential attachment rate translates into the expected log-normal degree distribution fairly accurately; a detail that was previously a source of some discrepancy not just in the context of scientific collaboration networks (Jeong et al., 2003; Redner, 2005). In the continuation we first give information on the raw data and network construction, while the results are presented and summarized in Sections 3 and 4, respectively.

2. Preliminaries

Slovenia is a small country located at the heart of Europe with a population of two million.[1] It has a well-documented research history, which is made possible by SICRIS – Slovenia's Current Research Information System[2] – hosting up-to-date publication records of all Slovene scientists. At present, there are 30,630 registered, including young and non-active scientists as well as laboratory personnel, which boils down the initial number to 8402 of those that are truly active research-wise or have been so in the past. By this we mean those that, to date, have at least one bibliographic unit indexed by the Web of Science. This criterium may be somewhat stringent, but it is the only one we could apply consistently. Moreover, since the publication data contain records not just of research but also of professional work and many other activities not necessarily concerning research, it is important to define a threshold for when two scientists are considered connected. Having given

[1] The official Web page of Slovenia is accessible via: http://www.slovenia.si/.
[2] The SICRIS Web page is accessible via: http://sicris.izum.si/.