Available online at www.sciencedirect.com
ScienceDirect
Social Networks 29 (2007) 70–80
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff

Trevor Fenner*, Mark Levene, George Loizou

School of Computer Science and Information Systems, Birkbeck College, University of London, London WC1E 7HX, UK

Abstract

Recently several authors have proposed stochastic evolutionary models for the growth of complex networks that give rise to power-law distributions. These models are based on the notion of preferential attachment leading to the "rich get richer" phenomenon. Despite the generality of the proposed stochastic models, there are still some unexplained phenomena, which may arise due to the limited size of networks such as protein, e-mail, actor and collaboration networks. Such networks may in fact exhibit an exponential cutoff in the power-law scaling, although this cutoff may only be observable in the tail of the distribution for extremely large networks. We propose a modification of the basic stochastic evolutionary model, so that after a node is chosen preferentially, say according to the number of its inlinks, there is a small probability that this node will become inactive. We show that as a result of this modification, by viewing the stochastic process in terms of an urn transfer model, we obtain a power-law distribution with an exponential cutoff. Unlike many other models, the current model can capture instances where the exponent of the distribution is less than or equal to two. As a proof of concept, we demonstrate the consistency of our model empirically by analysing the Mathematical Research collaboration network, the distribution of which has been shown to be compatible with a power law with an exponential cutoff.

Keywords: Collaboration networks; Power-law distribution; Exponential cutoff; Preferential attachment; Evolutionary models; Complex networks

1. Introduction

Power-law distributions taking the form:

$$f(i) = Ci^{-\tau}, \tag{1}$$

* Corresponding author.
E-mail addresses: trevor@dcs.bbk.ac.uk (Trevor Fenner), mark@dcs.bbk.ac.uk (Mark Levene), george@dcs.bbk.ac.uk (George Loizou).

0378-8733/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2005.12.003

where C and τ are positive constants, are abundant in nature (Sornette, 2000). The constant τ is called the *exponent* of the distribution. Examples of such distributions are: *Zipf's law*, which states that the relative frequency of a word in a text is inversely proportional to its rank, *Pareto's law*, which states that the number of people whose personal income is above a certain level follows a power-law distribution with an exponent between 1.5 and 2 (Pareto's law is also known as the 80:20 law, stating that about 20% of the population earn 80% of the income), and *Gutenberg-Richter's law*, which states that over a period of time the number of earthquakes of a certain magnitude is roughly inversely proportional to the magnitude. Recently, several researchers have detected power-law distributions in the topology of several networks such as the World-Wide-Web (Broder et al., 2000; Kumar et al., 2000) and author citation graphs (originally Price (1965) and more recently Redner (1998)).

The motivation for the current research is two-fold. First, from a complex network perspective, we would like to understand the stochastic mechanisms that govern the growth of a network. This has lead to fruitful interdisciplinary research by a mixture of Computer Scientists, Mathematicians, Statisticians, Physicists and Social Scientists (Albert and Barabási, 2002; Dorogovtsev and Mendes, 2000; Kleinberg et al., 1999; Krapivsky et al., 2000; Levene et al., 2002; Newman, 2001; Pennock et al., 2002; Solé et al., 2002), who are actively involved in investigating various characteristics of complex networks such as the degree distribution of the nodes, the diameter, and the relative sizes of various components. These researchers have proposed several stochastic models for the evolution of complex networks; many of these have the common theme of *preferential attachment* (originally Price (1976)) – which results in the "rich get richer" phenomenon – for example, where new links to existing nodes are added in proportion to the number of links to these nodes currently present. (We note that an alternative model for the evolution of complex networks involves a copying mechanism, whereby new links to an existing node are added by copying these links from the set of links of a node chosen uniformly at random (Kleinberg et al., 1999; Solé et al., 2002).

An extension of the preferential attachment model, proposed in Dorogovtsev and Mendes (2000), takes into account the ageing of nodes so that a link is connected to an old node, not only preferentially, but also depending on the age of the node: the older the node is the less likely it is that other nodes will be connected to it. It was shown in Dorogovtsev and Mendes (2000) that if the ageing function is a power law then the degree distribution has a phase transition from a power-law distribution, when the exponent of the ageing function is less than one, to an exponential distribution, when the exponent is greater than one. A different model of node ageing was proposed in Amaral et al. (2000) with two alternative ageing functions. With the first function the time a node remains 'active', i.e. may acquire new links, decays exponentially, and with the second function a node remains active until it has acquired a maximum number of links. Both functions were shown by simulation to lead to an exponential cutoff in the degree distribution, and for strong enough constraints the distribution appeared to be purely exponential. Another explanation of the cutoff, offered in Mossa et al. (2002), is that when a link is created the author of the link has limited information processing capabilities and thus only considers linking to a fraction of the existing nodes, viz. those that appear to be "interesting". It was shown by simulation that, when the fraction of "interesting" nodes is less than one, there is a change from a power-law distribution to one that exhibits an exponential cutoff, leading eventually to an exponential distribution when the fraction is much less than one.

The second motivation for this research is that the viability and efficiency of network algorithmics are affected by the statistical distributions that govern the network's structure. For example, the discovered power-law distributions in the web have recently found applications in local search