Discrete Applied Mathematics 162 (2014) 78–89

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

A random model of publication activity

Ágnes Backhausz *, Tamás F. Móri

Department of Probability Theory and Statistics, Eötvös Loránd University, Pázmány P. s. 1/c, Budapest, H-1117, Hungary

ARTICLE INFO

Article history:
Received 17 August 2012
Received in revised form 19 August 2013
Accepted 23 August 2013
Available online 12 September 2013

Keywords:
Scale free
Random graphs
Martingales
Renewal equation

ABSTRACT

We examine a random structure consisting of objects with positive weights and evolving in discrete time steps. It generalizes certain random graph models. We prove almost sure convergence for the weight distribution and show scale-free asymptotic behaviour. Martingale theory and renewal-like equations are used in the proofs.

1. Introduction

In this paper we examine a dynamic model inspired by scientific publication activity and networks of coauthors. However, the model contains many simplifying assumptions that are not valid in reality. We still use the terminology of publications for the sake of simplicity.

The model consists of a sequence of researchers. Each of them has a positive weight which is increasing in discrete time steps. The weights reflect the number and importance of the researcher's publications. One can think of cumulative impact factor for instance.

We start with a single researcher having a random positive weight. At the nth step a new publication is born. The number of its authors is randomly chosen. Then we select the authors, that is, one of the groups of that size; the probability that a given group is chosen is proportional to the sum of the weights of its members. After that the weights of the authors of the new publication are increased by random bonuses. Finally, a new researcher is added to the system with a random initial weight.

This is a preferential attachment model; one can see that authors with higher weights have higher chance to be chosen and increase their weights when the new publication is born.

We are interested in the weight distribution of the model. That is, for fixed $t > 0$, we consider the ratio of authors of weight larger than t, and study the asymptotic behaviour of this quantity as the number of steps goes to infinity.

Our main results (Section 3) include the almost sure convergence of the ratio of authors of weight larger than t under suitable conditions; first, when all weights are integer valued, then assuming that these random variables have continuous distribution. In both cases we describe the limiting sequence or function and determine its asymptotics. They are polynomially decaying under suitable conditions, thus our model shows scale-free behaviour.

The proofs of the almost sure convergence are based on the methods of martingale theory, while the polynomial decay of the asymptotic weight distribution follows from the results of [1] about renewal-like equations. See Section 4 for the details.

* Corresponding author. Tel.: +36 13722500, +36 13728529; fax: +36 13812152.
E-mail addresses: agnes@cs.elte.hu, agnes@math.elte.hu (Á. Backhausz), mori@math.elte.hu (T.F. Móri).

0166-218X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dam.2013.08.034

This model generalises some random graph models. To see this, assume that every publication has only one author, and at each step, when a publication is born, connect its author to the new one with an edge. We get a random tree evolving in time.

In the particular case where the initial weights and author's bonuses are always equal to 1, and the new paper always has one author, we get the Albert–Barabási random tree [2]. The neighbour of the new vertex is chosen with probabilities proportional to the degrees of the old vertices. Similarly, if the initial weights and the bonuses are fixed, but they are not necessarily equal to each other, we get random trees with linear weights [9], sometimes called generalised plane oriented recursive trees. In these cases the asymptotic degree distribution is well-known.

2. Notation and assumptions

2.1. Notation

- Let the label of the only researcher being present in the beginning be 0; the label of the researcher coming in the nth step is n.
- X_i is the initial weight of researcher i for $i = 0, 1, \ldots$. We suppose that $X_0, X_1, \ldots$ are independent, identically distributed positive random variables.
- ν_n is the number of coauthors at step n. This is an integer valued random variable for each n. For technical reasons we assume that $\nu_n \geq 1$ for all $n \geq 1$. Since the authors' weights are not necessarily increased, this may be supposed without loss of generality. Therefore ν_n has a distribution on $\{1, \ldots, n\}$.

  Given that $\nu_n = k$, a group of size k is chosen randomly from researchers $0, \ldots, n-1$. The probability that a given group is chosen is proportional to the total weight of the group. The selected researchers will be the authors of the nth paper.
- At step n, if the chosen authors are $i_1, \ldots, i_{\nu_n}$, we generate independently of the past random variables (bonuses) $Y_{n,i_1}, \ldots, Y_{n,i_{\nu_n}}$ to be given respectively to the authors $i_1, \ldots, i_{\nu_n}$. We will assume that the row vector $(Y_{n,i_1}, \ldots, Y_{n,i_{\nu_n}})$ is distributed like a generic row vector $(Y_{n,1}, \ldots, Y_{n,\nu_n})$.
- Let Z_n be the total weight of the nth paper; that is, $Z_n = Y_{n,1} + Y_{n,2} + \cdots + Y_{n,\nu_n}$ for $n \geq 1$.
- $W(n, i)$ denotes the weight of author i after step n for $i = 0, \ldots, n$. This is equal to X_i plus the sum of all bonuses $Y_{j,\ell}$ for which author i is the ℓth author of the jth paper $\big(\ell = 1, \ldots, \nu_j, j = 1, 2, \ldots, n\big)$.
- Let S_n be the total weight after n steps; namely,

$$S_n = W(n, 0) + \cdots + W(n, n) = X_0 + \cdots + X_n + Z_1 + \cdots + Z_n.$$

- X, ν, Y_n, Y and Z are random variables. X is equal to X_0 in distribution, and Y_n is equal to $Y_{n,1}$ in distribution for $n \geq 1$. The other random variables will be determined later by the assumptions. Finally, $\mathcal{F}_n$ is the σ-algebra generated by the first n steps; $\mathcal{F}_n^+ = \sigma\{\mathcal{F}_n, \nu_{n+1}\}$.

Throughout this paper $\mathbb{I}(A)$ denotes the indicator of event A. We say that two sequences (a_n), (b_n) are asymptotically equal ($a_n \sim b_n$), if they are not equal to zero except finitely many terms, and $a_n/b_n \to 1$ as $n \to \infty$. A sequence (a_n) is exponentially small if $|a_n| \leq q^n$, for all sufficiently large $n \in \mathbb{N}$ for some $0 < q < 1$.

2.2. Assumptions

Now we list the assumptions on the model.

Assumption 1. $X_0, X_1, \ldots$ are independent, identically distributed. The initial weights X_n, and the triplets $(\mathcal{F}_{n-1}, (Y_{n,1}, \ldots, Y_{n,\nu_n}), \nu_n)$ are independent $(n = 1, 2, \ldots)$.

Assumption 2. X has finite moment generating function.

Assumption 3. ν_n and $\big(Y_{n,1}, \ldots, Y_{n,\nu_n}\big)$ are independent of $\mathcal{F}_{n-1}$ for $n \geq 1$.

Assumption 4. $\nu_n \to \nu$ in distribution as $n \to \infty$; in addition, $\mathbb{E}\nu_n \to \mathbb{E}\nu < \infty$ and $\mathbb{E}\nu_n^2 \to \mathbb{E}\nu^2 < \infty$ hold.

Recall that $\nu_n \leq n$. Assumption 4 trivially holds if ν is a fixed random variable with finite second moment, and the distribution of ν_n is identical to the distribution of $\min(n, \nu)$, or to the conditional distribution of ν with respect to $\{\nu \leq n\}$.

Assumption 5. The conditional distribution of $\big(Y_{n,1}, \ldots, Y_{n,\nu_n}\big)$, given $\nu_n = k$, does not depend on n. Moreover, the components are conditionally interchangeable, given $\nu_n = k$.

Assumption 6. Z_n has finite expectation.

Now we know that $(\nu_n, Y_n, Z_n) \to (\nu, Y, Z)$ in distribution as $n \to \infty$, where Y and Z are random variables. We need that they also have finite moment generating functions, and they are not degenerate.