

A Stochastic Generative Model for Citation Networks Among Academic Papers

Junji Nakano

Department of Global Management, Chuo University, Tokyo, Japan

Abstract

We propose a stochastic generative model to represent a directed graph constructed by citations among academic papers, where nodes and directed edges represent papers with discrete publication time and citations respectively. The proposed model assumes that a citation between two papers occurs with a probability based on the type of the citing paper, the importance of cited paper, and the difference between their publication times, like the existing models. We consider the out-degrees of citing paper as its type, because, for example, survey paper cites many papers. We approximate the importance of a cited paper by its in-degrees. In our model, we adopt three functions: a logistic function for illustrating the numbers of papers published in discrete time, an inverse Gaussian probability distribution function to express the aging effect based on the difference between publication times, and an exponential distribution (or a generalized Pareto distribution) for describing the out-degree distribution. We consider that our model is a more reasonable and appropriate stochastic model than other existing models and can perform complete simulations without using original data. In this paper, we first use the Web of Science database and see the features used in our model. By using the proposed model, we can generate simulated graphs and demonstrate that they are similar to the original data concerning the in- and out-degree distributions, and node triangle participation. In addition, we analyze two other citation networks derived from physics papers in the arXiv database and verify the effectiveness of the model.