Physica A 465 (2017) 141–152

Contents lists available at ScienceDirect

Physica A

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

Parameter specification for the degree distribution of simulated Barabási–Albert graphs

Fairul Mohd-Zaid [a,*], Christine M. Schubert Kabban [a], Richard F. Deckro [b], Edward D. White [a]

[a] Department of Mathematics and Statistics, Air Force Institute of Technology, WPAFB, OH, United States

[b] Department of Operational Sciences, Air Force Institute of Technology, WPAFB, OH, United States

HIGHLIGHTS

- Least squares method is not appropriate for estimating β of Barabási–Albert graphs.
- A modified version of Clauset et al. (2009) is more preferable for estimation.
- Simulated graphs show that exponent is not the theoretical $\beta = 2$ for small graphs.
- Value of β is also dependent on number of edges added, *m*, for small graphs.
- Smaller value of m results in larger value of β for all graph sizes.

ARTICLE INFO

Article history:
Received 25 January 2016
Received in revised form 25 May 2016
Available online 10 August 2016

Keywords:
Barabási–Albert
Degree distribution
Parameter estimation
Pareto distribution
Yule–Simon distribution
Goodness of fit

ABSTRACT

The degree distribution of a simulated Barabási–Albert graph under linear preferential attachment is investigated. Specifically, the parameters of the power law distribution are estimated and compared against the theoretical values derived using mean field theory. Least squares method and MLE-nonparametric method were utilized to estimate the distribution parameters on 1000 simulated Barabási–Albert graphs for edge parameter $m \in \{2, 4, 6\}$ and size $n \in \{2^k : k = 5, 6, \ldots, 14, 15\}$. Goodness of fit metrics were computed on a second set of simulated graphs for the median of the estimated parameters and other hypothetical values for the distribution parameters. The results suggest that the distribution of the parameters from simulated graphs are significantly different from the theoretical distribution and is also dependent on m. Further results confirm the finding that the parameter of the power law distribution, β, increases as m increases.

Published by Elsevier B.V.

1. Introduction

Many fields of research produce relational data that can be represented in the form of graphs [1,2]. Representing relational data, or networks, in graph form provides a forum with which to study the relationships within the dataset, in particular, graphical properties that may be used to characterize the network. One such property is referred to as the scale-free property [3,4], and networks that are scale-free have a degree distribution which follows the *power law*. The scale-free property governs many physical, biological, and man-made phenomena [5–8].

One of the more well known graph models that exhibits the scale-free property was proposed by Barabási and Albert [5]. This model is governed by the concept of *preferential attachment* where new entities are more likely to form connections

* Correspondence to: 2950 Hobson Way, Bldg 641, Wright-Patterson AFB, OH 45433, United States.
E-mail address: fairul.mohd-zaid@us.af.mil (F. Mohd-Zaid).

http://dx.doi.org/10.1016/j.physa.2016.07.024
0378-4371/Published by Elsevier B.V.

with those that are already well connected in the network. Although there have been more complicated models for random graphs established in the literature [9], the properties of the Barabási–Albert model make it highly desirable.

It has been established that the degree distribution of the Barabási–Albert model follows a specific power law distribution, namely, the Pareto distribution, with an *exponent* of $\beta = 2$ [5]. The assumption that $\beta = 2$ was derived based on asymptotic theory and thus holds true for very large networks, although how well small networks are represented by $\beta = 2$ is unknown. Further, the degree distribution does not have a continuous support, suggesting that, for small networks, a discrete distribution may be more appropriate. Both the *Zipf's law* [10] and the Yule–Simon distribution [11,12] are discretized versions of the power law. Newman [6] stated that the Yule–Simon distribution is easier to work with than the Zipf's law since sums of the probability mass function (PMF) of the former can frequently be performed in closed form whereas sums of the PMF of the latter can only be written in terms of special functions.

Despite these findings, little has been done to extend the use of the degree distribution of the Barabási–Albert graph under the linear preferential attachment model although this model has been in use for over a decade [9]. One exception is shown by the work of Wang et al. [13] where they studied the degree distribution of the line graph of the Barabási–Albert graph and showed that it also follows the power law. However, a simulation in their study showed that the empirical distribution of the line graph deviates from the expected theoretical distribution. It is widely accepted in the literature that the power law exponent of the Barabási–Albert graph is $\beta = 2$ based on its derivation, although questions have been raised as to whether or not this is true [14,6,15]. Newman [6] claimed that $\beta = 1.2$, and even by their own admission, Albert and Barabási [14] showed that $\beta = 1.9$ provided the best fit on their simulated Barabási–Albert model for very large n.

One of the factors affecting the value of β is the choice of method for parameter estimation. Dating back to the 19th century, the classical method of parameter estimation for the power law has been conducted by fitting a least square line on the doubly logarithmic histogram [16]. A doubly logarithmic histogram is a technique of plotting a histogram using a skewed logarithmic scale on the vertical and horizontal axes. The exponent of the power law can then be estimated using the absolute slope of the line from the doubly logarithmic histogram. However, some have argued that there are issues with this technique. Newman [6] argued that the survivor function (SF) should instead be used because the behavior of the histogram is very dependent on the chosen binning, whereas Clauset et al. [7] argued for the use of a combination of maximum likelihood and nonparametric techniques. Clausetet al. [7] proposed that their method is more robust than least square fitting due to its statistical framework. The purpose of properly estimating the degree distribution for the Barabási–Albert network is so that we can represent complex phenomena in a well described, manageable, and well known visual graphical representation. This is especially important for small graphs which are well suited for visual analysis. This work seeks to determine the most appropriate parameter value for representing a Barabási–Albert network of various sizes.

We begin by describing the Barabási–Albert model and the representation of its degree distribution via the power law. Then, we will estimate the true distribution of the nodal degree for the Barabási–Albert graph for when the size of the graph, n, and the number of links added, m, are relatively small and compare these to the theoretical value of $\beta = 2$ provided by Barabási and Albert [5]. Considering the disparity in opinions of the proper way to estimate the parameters of the power law, we use both approaches of fitting a least squares on the SF and the MLE-nonparametric approach as proposed by Clauset et al. [7]. Comparisons are then made by looking at the goodness of fit of both methods by computing the $-$*loglikelihood* and mean square error (MSE) of their respective estimated distribution. We conclude with a discussion on the implication of our results on estimating the parameters of the Pareto distribution for the degree distribution.

2. Background

2.1. Barabási–Albert graph model

The Barabási–Albert model is based on two mechanisms that govern the scale-free property of real world networks: (1) networks expand continuously by the addition of new nodes, and (2) new nodes attach preferentially to other nodes that are already well connected. The model operates by first starting with an initial number of nodes, $m \leq m_0$, with no edges. This is followed by an iterative process of adding a single node with m edges. Each edge is connected to an existing node i with degree k_i based on the linear preferential attachment probability, $\pi(k_i)$, where

$$\pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{1}$$

Here, $\pi(k_i)$ is the probability that node i will be attached to the new node.

Thus, the nodal degree of the Barabási–Albert scale-free graph can be derived by using the mean field theory as described in Ref. [5] and reproduced, in part, here. Let m_0 be the number of nodes initially included in the graph and m be the number of edges added at each iteration over t iterations, where $m \leq m_0$. The size of the graph is then $n = m_0 + t$, the total number of edges $E = mt$, and consequently the total degree of the graph $\sum k_i = 2mt$. Given the preferential attachment in Eq. (1), the rate of change for k_i over t iterations could be written as

$$\frac{\partial k_i}{\partial t} = m\pi(k_i) = \frac{k_i}{2t}. \tag{2}$$

Download English Version:

https://daneshyari.com/en/article/974204

Download Persian Version:

https://daneshyari.com/article/974204

Daneshyari.com