Physica A 433 (2015) 182–197

Contents lists available at ScienceDirect

Physica A

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

Exactly scale-free scale-free networks

Linjun Zhang [a,b], Michael Small [b,*], Kevin Judd [b]

[a] Department of Statistics, The Wharton School, University of Pennsylvania, USA
[b] School of Mathematics and Statistics, The University of Western Australia, Crawley, WA, 6009, Australia

HIGHLIGHTS

- A maximum entropy process for generating random networks with specific properties.
- Applied to scale free networks we show that widely held properties are not typical.
- Random scale-free networks are typically not “robust-yet-fragile”.
- Rather, this is a property of the hub-centric nature of preferential attachment.
- Surrogate networks—to determine which properties of a specific network are typical.

ARTICLE INFO

Article history:
Received 13 November 2014
Received in revised form 11 March 2015
Available online 7 April 2015

Keywords:
Scale-free networks
Power-law
Complex networks

ABSTRACT

Many complex natural and physical systems exhibit patterns of interconnection that conform, approximately, to a network structure referred to as scale-free. Preferential attachment is one of many algorithms that have been introduced to model the growth and structure of scale-free networks. With so many different models of scale-free networks it is unclear what properties of scale-free networks are typical, and what properties are peculiarities of a particular growth or construction process. We propose a simple maximum entropy process which provides the best representation of what are typical properties of scale-free networks, and provides a standard against which real and algorithmically generated networks can be compared. As an example we consider preferential attachment and find that this particular growth model does not yield typical realizations of scale-free networks. In particular, the widely discussed “fragility” of scale-free networks is actually found to be due to the peculiar “hub-centric” structure of preferential attachment networks. We provide a method to generate or remove this latent hub-centric bias—thereby demonstrating exactly which features of preferential attachment networks are atypical of the broader class of scale-free networks. We are also able to statistically demonstrate whether real networks are typical realizations of scale-free networks, or networks with that particular degree distribution; using a new surrogate generation method for complex networks, exactly analogous the widely used surrogate tests of nonlinear time series analysis.

1. Introduction

The notion of scale-free networks has been around for a while [1]. The introduction of the *preferential attachment* (PA) algorithm for generating random scale-free graphs was a significant step in understanding the properties of scale-free networks, and the physical processes that create them [2]. The PA algorithm has spawned a good deal of subsequent

* Corresponding author.
E-mail addresses: linjunz@wharton.upenn.edu (L. Zhang), michael.small@uwa.edu.au (M. Small).

http://dx.doi.org/10.1016/j.physa.2015.03.074
0378-4371/

algorithms and analysis. The purpose of this paper is to highlight the straightforward fact that not all scale-free networks have the same properties, and that algorithms, like PA (for example), do not capture the full richness of scale-free networks, nor do they necessarily display properties that may be termed typical of all scale-free networks. We then generalize this process to show how it can be used to answer questions of typicality for experimentally observed networks.

To these aims, we first briefly recall in this introduction the principal processes that have been proposed to describe and generate scale-free networks, and indicate deficiencies of these processes when employed as models of typical scale-free networks. We then propose a *maximum entropy process* that provides an unbiased sample of the set of all scale-free networks. A *maximum entropy process* provides a better representation of expected properties of scale-free networks, both in terms of richness and typicality. A *maximum entropy process* also provides a unbiased standard against which other processes that generate scale-free networks can be compared. While a great deal of this maximum entropy process reduces to simple edge-switching, it is the application of this process to achieve maximum entropy realizations that is important.

In Section 2 we make a careful comparison of PA with our unbiased standard to illustrate how PA has a significant bias in the structural properties of the networks it generates. We demonstrate it has a *hub-centric* bias. In Section 3 we use a surrogate data approach to examine a selection of real-world networks claimed to be scale-free, and analyze in what sense these networks are typical of scale-free networks and how they differ. In nonlinear time series analysis, surrogate data are Monte-Carlo resamplings of observed time series that are also consistent with a given null hypothesis. Hence, they provide a statistical test for consistency of the original data with that hypothesis. Here we apply the same randomization ideas to generate new random network realizations from an experimentally observed network.

The main results of this paper are four-fold. *First*, we provide an algorithm to generate random networks with a given degree distribution (Section 1.3). Second we apply this algorithm to networks generated via preferential attachment and show that preferential attachment is not typical (Section 2.3)—and in particular the "robust-yet-fragile" property of preferential attachment networks is not generic to typical scale-free networks (Section 2.2). *Third*, we observe that the defining feature that makes preferential attachment networks special is their "hub-centric" character (Section 2.5), which we define algorithmically (Appendix B). *Fourth*, we introduce a surrogate test for networks (Section 3)—this random resampling of the network allows one to test which properties of a particular network are typical of all random networks with that given degree distribution, and which properties are unique for that particular observed network.

1.1. Scale-free networks from processes

In this section we briefly review the notion of scale-free networks, and the principal models of the physical processes that generate scale-free networks. These models can be broadly divided into *growth* models and *configuration* models.

Scale-free networks are usually identified by the histogram of node degrees having a power-law tail. However, many naturally occurring networks[1] have been identified as being scale-free and having a power-law tail: citation and collaboration networks [3,4], airline networks [5], protein–protein interaction [6], metabolic pathways [7], the world-wide web and internet [8].

To understand the formation of scale-free networks various models have been proposed to mimic the physical or conceptual processes that build and shape these networks. One of the first, proposed by Barabási and Albert, is *preferential attachment* [2], which is a restatement of the process described by de Solla Price [1] as a model of observed scale-free networks of citations [3].

Preferential attachment is an *unchanging*, *additive* growth process, where nodes of a fixed degree are added to the network, with links preferentially attached to existing nodes depending on their degree; usually proportional to the degree. Although this was suggested as a model of the process that grows the world-wide web (seen as hyperlinks between webpages), it was recognized that there is an additional *aging* process (AOL is replaced by Facebook, yahoo looses popularity to Google) so that attachment preferences *change* as the network growth proceeds [9–11].

Other processes can be introduced into a network growth model, such as *shuffling* parts of the network, and *deletion* of links and nodes [9,12], such that these actions do not change the underlying scale-free property of the networks produced. Moreover, scale-free networks can be produced by processes that are not growth processes. The most commonly considered of these are the so-called *configuration models*. Configuration models proceed by choosing nodes to have prescribed degrees, then connecting these together to form a scale-free network [13]; although care is needed to ensure the networks obtained satisfy other expected properties, such as, being simple (no self-links or multiple edges between nodes) and connected [14,15,12]. In addition to the problems with multiple edges and self-links [12], configuration models do not provide a well-founded sampling—in the sense of achieving maximum entropy realizations [16].

1.2. Not all scale-free networks are the same

With so many different processes generating scale-free networks, the question arises: do they generate the same type of networks? The simple answer is no. Many differences in the scale-free networks generated by different processes have been noted; some particular differences are outlined in the following.

[1] For a sample of the data we use here, see http://vlado.fmf.uni-lj.si/pub/networks/data/.