Exploration of scale-free networks

Do we measure the real exponents?

T. Petermann and P. De Los Ríos

Laboratoire de Biophysique Statistique, ITP-FSB, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Received 16 October 2003
Published online 17 February 2004 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2004

Abstract. The increased availability of data on real networks has favoured an explosion of activity in the elaboration of models able to reproduce both qualitatively and quantitatively the measured properties. What has been less explored is the reliability of the data, and whether the measurement technique biases them. Here we show that tree-like explorations (similar in principle to traceroute) can indeed change the measured exponents of a scale-free network.

PACS. 89.75.-k Complex systems – 87.23.Ge Dynamics of social systems – 05.70.Ln Nonequilibrium and irreversible thermodynamics

1 Introduction

In recent years networks have become one of the most promising frameworks to describe systems as diverse as the Internet and the WWW, email and social communities, distribution systems, food-webs, protein interaction, genetic and metabolic networks [1]. The collected data have allowed the discovery of many important properties: in particular two of them have become prominent, namely the small-world [2] and scale-free features [3]. Small-world implies that the average distance between nodes of the network increases at most logarithmically with the number of nodes, and formalizes the concept of “six degrees of separation” typical in social contexts. Scale-free refers to the lack of an intrinsic scale in some of the properties of the network. In particular, the quantity that has been most thoroughly studied is the degree (or connectivity) distribution: the degree \( k \) of a node is the number of other nodes it has links to (here we do not distinguish between directed and undirected links), and the degree distribution \( P(k) \) is simply the histogram of the number of nodes with a given degree \( k \). Scale-free networks exhibit a power-law behavior of the distribution \( P(k) \sim k^{-\gamma} \), with \( \gamma \) values often between 2 and 3 [1]. The small-world and scale-free properties turned out being quite ubiquitous and some general, qualitative, models to explain their appearance have been put forward. At the same time various versions of these models have been also proposed in order to capture also the detailed values of some quantities, such as the exponent \( \gamma \). Yet, as this new field is slowly coming of age, and as a consequence it is also becoming more quantitative, an analysis of the data, and of their reliability, is due.

The main problem that should be addressed is whether the data we are using have been skewed somehow by the detection method. In the lack of such analysis on real data and methods, we propose to work on synthetic models and data to explore their robustness in some simple test case. In the next section we address the tree-like exploration technique and discuss how it can bias the measurements, and in the third section we show that a random graph can be distorted by the exploration so to look like a SF one: in this case the exponent \( \gamma \) is completely spurious.

2 Tree-like exploration of scale-free networks

Scale-free (SF) networks can be explored in many different ways. One of the most popular methods, that has been extensively used for example for the Internet, is a sort of tree-like exploration implemented by the recursive use of the traceroute command. In short, traceroute finds a path (usually a short one, but not necessarily the shortest) from the node where the command is executed to another given node. By repeating the procedure asking traceroute to find paths to all other possible nodes (addressed by their IP number), one ends up with a representation of the Internet that shows just a small amount of loops. This is due to the fact that traceroute mostly uses the same paths: if a node \( D \) can be reached from \( A \) through both \( B \) and \( C \), traceroute most of the times detects only one of them. Actually, chances are that traceroute can find more than a single path if traffic over an already discovered one is so high that it becomes more convenient to switch to a different path. Data collected with this technique have shown that degrees in the Internet are distributed according to a
power-law with exponent $\gamma \simeq 2.2 \pm 0.1$ [4]. In order to analyze the effects of a tree-building exploration algorithm on SF networks, we have synthesized our own networks according to two different models: the Barabasi-Albert (BA) model [3], and the hidden variable model [5].

The BA model describes the growth of a network as new nodes are added at a constant rate, and they connect to older nodes in the network according to the preferential attachment rule. Preferential attachment means that an old node has a probability proportional to its degree of acquiring a connection from a new one. It is useful to recall a simple derivation of the degree distribution starting from these two simple rules, growth and preferential attachment. The rate of change of the degree $k_i$ of node $i$ is

$$\frac{dk_i(t)}{dt} = \frac{k_i(t)}{2m}$$

(1)

where $m$ is the number of connections that a new node establishes with older ones, and the denominator in the right hand side of equation (1) represents the sum over all the degrees of the network. Equation (1) has the simple solution $k_i(t) = m(t/\tau_i)^{1/2}$, where $\tau_i$ is the time at which node $i$ entered the network. Since the relation between $k_i$ and $\tau_i$ is monotonous we can classify nodes according either to their degree $k_i$ or to their age $\tau_i$. As a consequence we can apply the usual formula to transform probabilities into densities.

The first step in our analysis is to build a BA network, whose degree distribution is shown in Figure 1. Then, the density of reachable nodes at time $t$ is given by

$$\frac{dN(t)}{dt} = 1 - \left(1 - p \int_0^t \frac{dN(t')}{dt'} q(t')dt'\right)^m$$

(2)

where $q(t')$ is the probability to choose a node introduced in the network between $t'$ and $t' + dt'$: the preferential attachment rule translates to $q(t') = 1/(2(t' - t)^{1/2})$ (this trick is similar to assigning to each node a hidden variable corresponding to the time $t'$ at which it entered the network, with a connection probability that depends on the hidden variables of both the new and old nodes; for more details see below [7,8]). Since $N(t)$ can grow at most linearly, we make the assumption that $\frac{dN(t)}{dt} \sim t^\alpha$ with $\alpha$ expected to be negative. After some algebra, and keeping only the leading terms, we find $\alpha = (mp - 1)/2$ as long as $mp < 1$ the density of reachable nodes decreases in time. Then, the measured degree distribution can be again obtained from the relation $P_m(k)dk = \rho(\tau)d\tau$, with $\rho(\tau) \sim \tau^\alpha$, from which we obtain $P_m(k) \sim k^{-\gamma}$, with $\gamma_m = 2 + mp$ For $m = 1$, $p = 0.5$ we have $\gamma_m = 2.5$. 

![Degree distribution for a Barabási-Albert network grown with $m = 1$, with $10^5$ nodes. Circles: original network; stars: explored network with $p = 0.5$. The best fit to the original network is with $\gamma \simeq 3$, and to the explored network with $\gamma \simeq 2.5$. Inset: rescaled degree distribution $k^\rho P(k)$, such that the data for the original network are constant, and the residual power-law behavior of the explored network is more evident.](image-url)