PHYSICAL REVIEW E **73**, 046102 (2006)

Distribution of edge load in scale-free trees

Attila Fekete* and Gábor Vattay†
Department of Physics of Complex Systems, Eötvös University, Pázmány P. sétány 1/A., H-1117 Budapest, Hungary

Ljupco Kocarev‡
Institute for Nonlinear Sciences, University of California, 9500 Gilman Drive, La Jolla, San Diego, California 92093, USA
(Received 27 October 2005; published 3 April 2006)

Node betweenness has been studied recently by a number of authors, but until now less attention has been paid to edge betweenness. In this paper, we present an exact analytic study of edge betweenness in evolving scale-free and non-scale-free trees. We aim at the probability distribution of edge betweenness under the condition that a local property, the in-degree of the "younger" node of a randomly selected edge, is known. En route to the conditional distribution of edge betweenness the exact joint distribution of cluster size and in-degree, and its one-dimensional marginal distributions have been presented in the paper as well. From the derived probability distributions the expectation values of different quantities have been calculated. Our results provide an exact solution not only for infinite, but for finite networks as well.

DOI: 10.1103/PhysRevE.73.046102 PACS number(s): 89.75.Da, 02.50.Cw, 05.50.+q

I. INTRODUCTION

In recent years, the statistical properties of complex networks have been extensively investigated by the physics community [1–4]. With the increasing computing power of modern computers, analysis of large-scale networks and databases has become possible. It has been shown that the degree statistics of many natural and artificial networks follow power law. Examples for such networks vary from social interconnections and scientific collaborations [5] to the world-wide web [6] and the Internet [7,8]. These networks are usually referred to as *scale-free* networks, since the power law distribution indicates that there is no characteristic scale in these systems.

In the early 1960s Erdős and Rényi (ER) introduced random graphs that served as the first mathematical model of complex networks [9]. In their model, the number of nodes is fixed and connections are established randomly, with probability p_{ER}. Although the ER model leads to rich theory, it fails to predict the power law distributions observed in scale-free networks. Barabási and Albert (BA) proposed a more suitable evolving model of these networks [10,11]. The BA model is also based on the random graph theory, but involves two key principles in addition: (a) *Growth*, that is, the size of the network is increasing during development, and (b) *preferential attachment*, that is, new network elements are connected to higher degree nodes with higher probability. In the BA model every new node connects to the core network with a fixed number of links m.

The study of complex networks usually deals with the structural properties of networks, like degree distribution [12], shortest path distribution [13], degree-degree correlations, clustering [14], etc. For complex networks which involve a transport mechanism *betweenness* is the matter of importance. Roughly speaking, betweenness is the number of shortest paths passing through a certain network element. For example, in communication networks information flows between remote hosts via intermediate stations and in the Internet data packets are transmitted through routers and cables. The expected traffic flowing through a link or a router is proportional to the particular edge or node betweenness, respectively. News and rumors spread in social networks, and node betweenness measures the importance or centrality of an individual in society.

Node betweenness has been studied recently by Goh, Kahng, and Kim [15], who argued that it follows power law in scale-free networks, and the exponent $\delta \approx 2.2$ is independent from the degree distribution in a certain range. Szabó, Alava, and Kertész [13] used rooted deterministic trees to model scale-free trees, and have found scaling exponent $\delta_t = 2$. The same scaling exponent has been found experimentally by Goh, Kahng, Kim for scale-free trees. The rigorous proof of the heuristic results of [13] has been provided by Bollobás and Riordan in [16].

Until recently, less attention has been paid to edge betweenness, even though edge betweenness is often essential for estimating the load on links in complex networks. For example, the edge betweenness can measure the "importance" of relationships in social networks, or it can measure the expected amount of data flow on links in computer networks. The probability distribution of edge betweenness gives a rough statistical description of links and it characterizes the network as a whole. Therefore, it is an important tool for an overall description of links in complex networks.

In some cases, a local property of the network is known as well. For instance, if the number of friends of any individual can be counted, then it is reasonable to ask the importance of a relationship (i.e., an edge in a social network) under the condition that the number of friends of the related individuals is known. In this case, the *conditional* probability distribution of edge betweenness provides a much finer description of links than the total distribution.

*Electronic address: fekete@complex.elte.hu
†Electronic address: vattay@complex.elte.hu
‡Electronic address: lkocarev@ucsd.edu

In this paper we focus on how additional local information could be used to describe links. In particular, we aim at deriving the probability distribution of edge betweenness in evolving scale-free trees, under the condition that the in-degree of the "younger" node of any randomly selected link is known. For the sake of simplicity we consider the in-degree of the younger node only. Whether a node is younger than another node or not can be defined uniquely in evolving networks, since nodes attach to the network sequentially. Note that the in-degree is considered instead of total degree for practical reasons only. The construction of the network implies that the in-degree is less than the total degree by one for every younger node.

To obtain the desired conditional distribution we calculate the exact joint distribution of cluster size and in-degree for a *specific* link first. Then, the joint distribution of a *randomly selected* link is derived, which is comparable with the edge ensemble statistics obtained from a network realization. The exact marginal distributions of cluster size and in-degree follow next. After that, we give the distribution and mean of cluster size under the condition that in-degree is known. For the sake of completeness the conditional in-degree distribution is presented as well. Finally, the distribution and mean of edge betweenness is derived under the condition that the corresponding in-degree is known. Note that all of our analytic results are *exact even for finite networks*, which is valuable since the size of the real networks are often much smaller than the valid range of asymptotic formulae. Moreover, *exact results for unbounded networks* are provided as well.

As a model of evolving scale-free trees we consider the BA model with parameter $m=1$, extended with initial attractiveness [17,18]. With the initial attractiveness the scaling properties of the network can be finely tuned. Note, that in the limit of initial attractiveness to infinity the preferential attachment disappears, and new nodes are connected to the old ones with uniform probability. In this limit the network looses its scale-free nature and becomes similar to an ER network with $p_{ER}=2/N$. Therefore, scale-free and nonscale free networks can be compared within one model. For the sake of simplicity the infinite limit of initial attractiveness is referred to as the "ER limit" throughout this paper.

We restrict our model to trees, that is to connected loopless graphs. The simplicity of trees allows analytic results for edge betweenness, since the shortest paths in trees are unique between any pair of nodes. Although trees are special graphs, a number of real networks can be modeled by trees or by tree-like graphs with only a negligible number of shortcuts. Important examples of such networks are the Autonomous Systems in the Internet [19].

The rest of this paper is organized as follows: In Sec. II, a short introduction to the construction of BA trees is given. Then, a master equation for the joint distribution of cluster size and in-degree of a specific edge is derived and solved in Secs. III and IV, respectively. The total joint distribution of cluster size is calculated in Sec. V. The marginal and conditional distributions of cluster size and in-degree are derived in Secs. VI and VII, respectively. In Sec. VIII, the conditional distribution of edge betweenness follows. Finally, we conclude our work and discuss future directions in Sec. IX.

II. THE NETWORK MODEL

The concepts of graph theory are used throughout this paper. A graph consists of *vertices* (nodes) and *edges* (links). Edges are ordered or un-ordered pairs of vertices, depending on whether an ordered or un-ordered graph is considered, respectively. The *order* of a graph is the number of vertices it holds, while the *degree* of a vertex counts the number of edges adjacent to it. *Path* is also defined in the most natural way: it is a vertex sequence, where any two consecutive elements form an edge. A path is called a *simple path* if none of the vertices in the path are repeated. Any two vertices in a *tree* can be connected by a unique simple path. The graph is called connected if for any vertex pair there exists a path which starts from one vertex and ends at the other.

The construction of the network proceeds in discrete time steps. Let us denote time with $\tau \in \mathbb{N}$, and the developed graph with $G_\tau=(V_\tau,E_\tau)$, where V_τ and E_τ denote the set of vertices and the set of edges at time step τ, respectively. Initially, at $\tau=0$, the graph consists only of a single vertex without any edges. Then, in every time step, a new vertex is connected to the network with a single edge. The edge is *directed*, which emphasize that the two sides of the edge are not symmetric. The newly connected node, which is the source of the edge, is always younger than the target node. The term "younger node of a link" is used in this sense below. Note that the initial vertex is different from all the others, since it has only incoming connections; we refer to it as the *root vertex*.

The target of every new edge is selected randomly from the present vertices of the graph. The probability that a new vertex connects to an old one is proportional to the attractiveness of the old vertex v, defined as

$$A(v) = a + q, \tag{1}$$

where parameter $a>0$ denotes the initial attractiveness and q is the in-degree of vertex v. It has been shown in [18] that the in-degree distribution is asymptotically $\mathbb{P}(q) \simeq (1+a)\frac{\Gamma(2a+1)}{\Gamma(a)}(q+a)^{-(2+a)}$. We will improve this result and derive the exact in-degree distribution below. Note that in the special case $a=0$ the attractiveness of every node is zero except of the root vertex. It follows that every new vertex is connected to the initial vertex in this case, which corresponds to a star topology. The special case $a=1$ practically returns the original BA model. Indeed, except for the root vertex, the attractiveness of every vertex becomes equal to its degree if $a=1$; this is exactly the definition of the attractiveness in the BA model [10]. Finally, if $a\to\infty$, then preferential attachment disappears in the limit, and the model tends to a Poisson-type graph, similar to an ER graph.

The attractiveness of sub-graph S is the sum of the attractiveness of its elements:

$$A(S) = \sum_{v' \in S} A(v'). \tag{2}$$

We refer to a connected sub-graph as a *cluster*. The attractiveness of cluster C can be given easily:

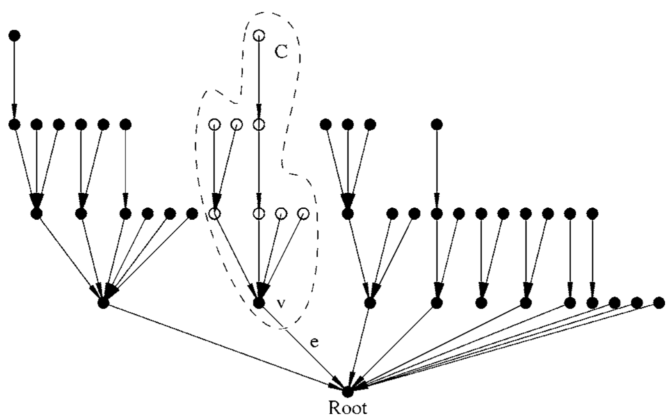

FIG. 1. Schematic illustration of the evolving network at time τ. Vertex v, connected to the network at τ_e, denotes the root of cluster C. Variables q and $n=|C|-1$ denote the in-degree of vertex v and the number of nodes in C without v (marked by circles), respectively.

$$A(C)=(1+a)|C|-1, \tag{3}$$

where $|C|$ denotes the size of the cluster. It is obvious that the overall attractiveness of the network at time step τ is

$$A(V_\tau)=(1+a)(\tau+1)-1. \tag{4}$$

III. MASTER EQUATION FOR THE JOINT DISTRIBUTION OF CLUSTER SIZE AND IN-DEGREE

Let us consider the size of the network N, an arbitrary edge e, which connected vertex v to the graph at time step $\tau_e>0$, and let us denote by C the cluster which has developed on vertex v until $\tau>\tau_e$ (Fig. 1). The calculation of betweenness of the given edge is straightforward in trees, since the number of shortest paths going through the given edge, that is the betweenness of the edge, is obviously $L=|C|(N-|C|)$. Therefore, it is sufficient to know the size of the cluster on the particular edge to get edge betweenness.

The development of cluster C can be regarded as a Markov process. The states of the cluster are indexed by (n,q), where $n=|C|-1$ denotes the number of vertices in cluster C without v. The in-degree of vertex v is denoted by q. Transition probabilities can be obtained from the definition of preferential attachment

$$W_{\tau,n,q}=\frac{A(C_\tau\setminus v)}{A(V_\tau)}=\frac{n-\alpha q}{\tau+1-\alpha} \tag{5}$$

$$W'_{\tau,q}=\frac{A(v)}{A(V_\tau)}=\frac{\alpha q+1-\alpha}{\tau+1-\alpha}, \tag{6}$$

where $\alpha=1/(1+a)\in]0,1]$ and $W_{\tau,n,q}$ denotes the transition probability $(n,q)\to(n+1,q)$, and $W'_{\tau,q}$ denotes the transition probability $(n,q)\to(n+1,q+1)$, respectively.

The Master-equation, which describes the Markov process, follows from the fact that cluster C can develop to state (n,q) obviously in three ways: A new vertex can be connected:

(1) To cluster C but not to vertex v, and the cluster was in state $(n-1,q)$;

(2) to vertex v, and the cluster was in state $(n-1,q-1)$; or

(3) to the rest of the network, and the cluster was in state (n,q).

Therefore, the conditional probability $\mathbb{P}_\tau(n,q|\tau_e)$ that the developed cluster on edge e is in state (n,q) satisfies the following Master-equation

$$\begin{aligned}\mathbb{P}_\tau(n,q|\tau_e)&=W_{\tau-1,n-1,q}\mathbb{P}_{\tau-1}(n-1,q|\tau_e)\\&\quad+W'_{\tau-1,q-1}\mathbb{P}_{\tau-1}(n-1,q-1|\tau_e)\\&\quad+[1-W_{\tau-1,n,q}-W'_{\tau-1,q}]\mathbb{P}_{\tau-1}(n,q|\tau_e),\end{aligned} \tag{7}$$

Since the process starts with $n=0$, $q=0$ at $\tau=\tau_e$, the initial condition of the above Master equation is $\mathbb{P}_{\tau_e}(n,q|\tau_e)=\delta_{n,0}\delta_{q,0}$, where $\delta_{i,j}$ is the Kronecker-delta symbol.

IV. THE SOLUTION OF THE MASTER EQUATION

After substituting the above transition probabilities into (7), the following first order linear partial difference equation is obtained

$$\begin{aligned}(\tau-\alpha)\mathbb{P}_\tau(n,q|\tau_e)&=(n-1-\alpha q)\mathbb{P}_{\tau-1}(n-1,q|\tau_e)\\&\quad+(\alpha q+1-2\alpha)\mathbb{P}_{\tau-1}(n-1,q-1|\tau_e)\\&\quad+(\tau-n-1)\mathbb{P}_{\tau-1}(n,q|\tau_e),\end{aligned} \tag{8}$$

Let us seek a particular solution of (8) in product form: $f(\tau)g(n)h(q)$. The following equation is obtained after substituting the probe function into (8)

$$\begin{aligned}(\tau-\alpha)\frac{f(\tau)}{f(\tau-1)}-\tau&=(n-1-\alpha q)\frac{g(n-1)}{g(n)}-n-1\\&\quad+(\alpha q+1-2\alpha)\frac{g(n-1)}{g(n)}\frac{h(q-1)}{h(q)}.\end{aligned}$$

The above partial difference equation can be separated into a system of three ordinary difference equations. The solutions of the separated equations are

$$f(\tau)=\frac{\Gamma(\tau+\lambda_1)}{\Gamma(\tau-\alpha+1)}, \tag{9}$$

$$g(n)=\frac{\Gamma(n+\lambda_2)}{\Gamma(n+\lambda_1+1)}, \tag{10}$$

$$h(q)=\frac{\Gamma(q+1/\alpha-1)}{\Gamma(q+\lambda_2/\alpha+1)}, \tag{11}$$

where λ_1 and λ_2 are separation parameters.

The solution of (7), which fulfils the initial conditions, is constructed from the linear combination of the above particular solutions:

$$\mathbb{P}_\tau(n,q|\tau_e)=\sum_{\lambda_1,\lambda_2}C_{\lambda_1,\lambda_2}f(\tau)g(n)h(q), \tag{12}$$

where C_{λ_1,λ_2} coefficients are independent of τ, n and q.

To obtain coefficients C_{λ_1,λ_2}, the initial condition of (7) is expanded on the bases of $g(n)$ and $h(q)$. The detailed calculation is presented in Appendix A.

The solution of (7) is

$$\mathbb{P}_\tau(n,q|\tau_e) = \frac{\Gamma(\tau-\tau_e+1)}{\Gamma(\tau_e)\Gamma(n+1)}\frac{\Gamma(\tau-n)}{\Gamma(\tau-\tau_e-n+1)} \times \frac{\Gamma(\tau_e+1-\alpha)}{\Gamma(\tau+1-\alpha)}\frac{\Gamma(q+1/\alpha-1)}{\Gamma(1/\alpha-1)}\Phi_\alpha(n,q) \tag{13}$$

where $\Phi_\alpha(n,q)=\sum_{k=0}^{q}\frac{(-1)^k}{k!(q-k)!}(-\alpha k)_n$ and $(x)_n \equiv \Gamma(n+x)/\Gamma(x)$ denotes Pochhammer's symbol. Note that $\mathbb{P}_\tau(n,q|\tau_e)\neq 0$ iff $0\leqslant q\leqslant n\leqslant \tau-\tau_e$. The conditions $0\leqslant q$ and $n\leqslant\tau-\tau_e$ are obvious, since $1/\Gamma(k)=0$ by definition if k is a negative integer or zero. Furthermore, the condition $q<n$ can be easily seen if $\Phi_\alpha(n,q)$ is transformed into the following equivalent form: $\Phi_\alpha(n,q)=\frac{1}{q!}\frac{d^n}{dz^n}z^{n-1}(1-z^{-\alpha})^q|_{z=1}$. This result coincides with the fact that the size of a cluster n cannot be less than the corresponding number of in-degrees q.

V. JOINT DISTRIBUTION OF CLUSTER SIZE AND IN-DEGREE

Equation (13) provides the conditional probability that a particular edge which was connected to the network at τ_e is in state (n,q) at $\tau>\tau_e$. In a fully developed network, however, the time when a particular edge is connected to the network is usually not known. Moreover, the development of an individual link is usually not as important as the properties of the finally developed link ensemble. Therefore, we are more interested in the total probability $\mathbb{P}_\tau(n,q)$, that is the probability that a randomly selected edge is in state (n,q) at τ, than the conditional probability (13). The total probability can be calculated with the help of the total probability theorem

$$\mathbb{P}_\tau(n,q) = \sum_{\tau_e=1}^{\tau}\mathbb{P}_\tau(n,q|\tau_e)\mathbb{P}_\tau(\tau_e), \tag{14}$$

where $\mathbb{P}_\tau(\tau_e)$ is the probability that a randomly selected edge was included into the network at τ_e. According to the construction of the network one edge is added to the network at every time step, therefore, $\mathbb{P}_\tau(\tau_e)=1/\tau$. The following formula can be obtained after the above summation has been carried out

$$\mathbb{P}_\tau(n,q) = \frac{\tau+1-\alpha}{\tau}\frac{(1/\alpha-1)_q}{(2-\alpha)_{n+1}}\Phi_\alpha(n,q), \tag{15}$$

where $0<\alpha\leqslant 1$. In star topology, that is when $\alpha=1$, the joint distribution $\mathbb{P}_\tau(n,q)$ evidently degenerates to $\mathbb{P}_\tau(n,q)=\delta_{n,0}\delta_{q,0}$.

The ER limit of joint distribution can be obtained via the $\alpha\to 0$ limit of (15) (see Appendix B for details):

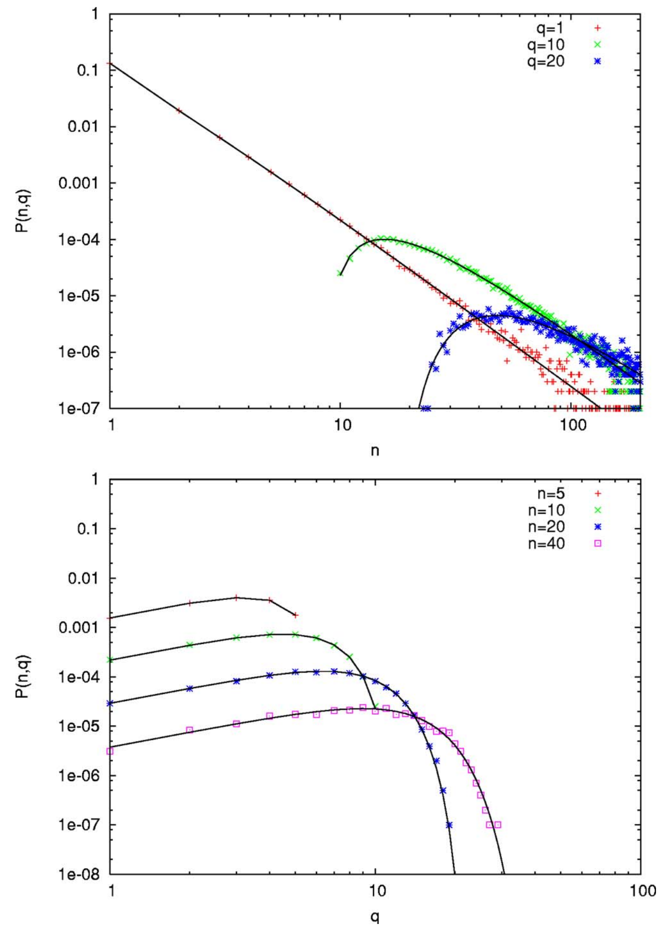

FIG. 2. (Color online) Joint empirical distribution of cluster size and in-degree at $\alpha=1/2$ (symbols), and analytic formula (15) (solid lines) are compared on double-logarithmic plot. Simulation results have been obtained from 100 realizations of 10^5 size networks.

$$\lim_{\alpha\to 0}\mathbb{P}_\tau(n,q) = \frac{\tau+1}{\tau}\sum_{k=q-1}^{n-1}(-1)^{k+n-1}\frac{\binom{k}{q-1}S_{n-1}^{(k)}}{\Gamma(n+3)} \tag{16}$$

where $0<q\leqslant n<\tau$ and $S_n^{(m)}$ denote the Stirling numbers of the first kind. Note, that for the special case $n=q=0$ the ER limit is $\lim_{\alpha\to 0}\mathbb{P}_\tau(0,0)=\frac{\tau+1}{2\tau}$.

The above formulas have been verified by extensive numerical simulations. The joint empirical cluster size and in-degree distribution has been compared with the analytic formula (15) for $\alpha=1/2$ in Fig 2. Figures 2(a) and 2(b) represent intersections of the joint distribution with cutting planes of fixed in-degrees and cluster sizes, respectively. The figures confirm that the empirical distributions, obtained as relative frequencies of links with cluster size n and in-degree q in 100 network realizations, are in complete agreement with the derived analytic results.

Equation (15) is the fundamental result of this section. The derived distribution is exact for any finite value of τ, that is for any finite BA trees. This result is precious for modeling a number of real networks, where the size of the network is small, compared to the relevant range of cluster size or in-degree. If the size of the network is much larger than the relevant range of cluster size or in-degree, then it is practical to consider the network as infinitely large, that is to take the $\tau\to\infty$ limit. For the above joint distributions (15) and (16) the $\tau\to\infty$ limit is evident, since the τ dependent prefactors obviously tend to 1 if the size of the networks grow beyond every limit.

VI. MARGINAL DISTRIBUTIONS OF CLUSTER SIZE AND IN-DEGREE

We have derived the joint probability distribution of the cluster size and the in-degree in the previous section. In many cases it is sufficient to know the probability distribution of only one random variable, since the information on the other variable is either unavailable or not needed. It is also possible that the one dimensional distribution is needed especially, for example, for the calculation of a conditional distribution in Sec. VII.

The one-dimensional (marginal) distributions $\mathbb{P}_\tau(n)$ and $\mathbb{P}_\tau(q)$ can be obtained from joint distribution $\mathbb{P}_\tau(n,q)$ as follows

$$\mathbb{P}_\tau(n)=\sum_{q=0}^{n}\mathbb{P}_\tau(n,q),\quad \mathbb{P}_\tau(q)=\sum_{n=q}^{\tau-1}\mathbb{P}_\tau(n,q).$$

After substituting (15) into the above formulas the following expressions are obtained

$$\mathbb{P}_\tau(n)=\frac{\tau+1-\alpha}{\tau}\frac{1-\alpha}{(n+1-\alpha)(n+2-\alpha)}. \tag{17}$$

if $0\leq n<\tau$ and $\mathbb{P}_\tau(n)=0$ if $n\geq\tau$. Furthermore,

$$\mathbb{P}_\tau(q)=\frac{\tau+1-\alpha}{\tau}\frac{1}{\alpha}\frac{(1/\alpha-1)_{1/\alpha}}{(q+1/\alpha-1)_{1/\alpha+1}} - \frac{\tau+1-\alpha}{\tau}\frac{(1/\alpha-1)_q}{(2-\alpha)_\tau}\sum_{k=0}^{q}\frac{(-1)^k}{k!(q-k)!}\frac{(-\alpha k)_\tau}{\alpha k+2-\alpha}. \tag{18}$$

if $0\leq q<\tau$ and $\mathbb{P}_\tau(q)=0$ otherwise. Rice's method [20] has been applied to evaluate the first term of $\mathbb{P}_\tau(q)$ in closed form.

The ER limit of the marginal cluster size distribution can obviously be obtained from (17) at $\alpha=0$. Furthermore, the ER limit of the marginal in-degree distribution can be derived analogously to the limit of the joint distribution, shown in Appendix B

$$\lim_{\alpha\to0}\mathbb{P}_\tau(q)=\frac{\tau+1}{\tau}\frac{1}{2^{q+1}} + \frac{\tau+1}{\tau}\frac{1}{\Gamma(\tau+2)\Gamma(q)}\frac{d^{q-1}}{d\alpha^{q-1}}\left.\frac{(1+\alpha)_{\tau-1}}{2-\alpha}\right|_{\alpha=0}. \tag{19}$$

If the size of the network grows beyond every limit, that is if $\tau\to\infty$, then the marginal distributions become much simpler

$$\mathbb{P}_\infty(n)=\frac{1-\alpha}{(n+1-\alpha)(n+2-\alpha)} \tag{20}$$

$$\mathbb{P}_\infty(q)=\frac{1}{\alpha}\frac{(1/\alpha-1)_{1/\alpha}}{(q+1/\alpha-1)_{1/\alpha+1}} \tag{21}$$

$$\lim_{\alpha\to0}\mathbb{P}_\infty(q)=2^{-q-1}. \tag{22}$$

The asymptotic behavior of the cluster size and in-degree distributions differ significantly. The tail of the cluster size distribution follows power law with exponent 2 either in BA or ER network, independently of α. However, we learned that the tail of the in-degree distribution follows power law with exponent $1/\alpha+1=2+a$ in BA networks, and it falls exponentially in ER topology, which agree with the well known results of previous works [9].

It is worth noting that the mean cluster size diverges logarithmically as the size of the network tends to infinity: $\mathbb{E}_\tau\{n\}=\sum_{n=0}^{\tau-1}n\mathbb{P}_\tau(n)=(1-\alpha)\ln\tau+O(1)$. The expectation value of the in-degree, however, obviously remains finite: $\mathbb{E}_\tau\{q\}=\frac{\tau}{\tau+1}<1$, and $\mathbb{E}_\infty\{q\}=1$ if the size of the network is infinite. Moreover, the standard error of the in-degree can be also given exactly when the size of the network grows beyond every limit

$$\mathbb{E}_\infty\{(q-1)^2\}=\frac{2}{|1-2\alpha|}. \tag{23}$$

This result implies that the fluctuations of the in-degree diverge in a boundless network, if $\alpha=1/2$, that is in the classical BA model.

Our analytic results have been verified with computer simulations. Since cumulative distributions are more suitable to be compared with simulations than ordinary distributions, we matched the corresponding complementary cumulative distribution functions (CCDF) against simulation data. The CCDF of cluster size, $F_\tau^c(n)=\sum_{n'=n}^{\tau-1}\mathbb{P}_\tau(n')$ can be calculated straightforwardly

$$F_\tau^c(n)=\frac{\tau+1-\alpha}{\tau}\frac{1-\alpha}{n+1-\alpha}-\frac{1-\alpha}{\tau}, \tag{24}$$

where $0\leq n<\tau$ and $0\leq\alpha\leq1$. The CCDF of in-degree, $F_\tau^c(q)=\sum_{q'=q}^{\tau-1}\mathbb{P}_\tau(q')$ is more complex, however

$$F_\tau^c(q)=\frac{\tau+1-\alpha}{\tau}\frac{(1/\alpha-1)_{1/\alpha}}{(q+1/\alpha-1)_{1/\alpha}}-\frac{1-\alpha}{\tau} + \frac{\tau+1-\alpha}{\tau}\frac{(1/\alpha-1)_q}{(2-\alpha)_\tau} \times\sum_{k=0}^{q-2}\frac{(-1)^k}{k!(q-2-k)!}\frac{(1-\alpha-\alpha k)_{\tau-1}}{(k+1/\alpha)(k+2/\alpha)} \tag{25}$$

where $0\leq q<\tau$ and $0<\alpha\leq1$. If the size of the network grows beyond every limit, then the CCDFs are the following:

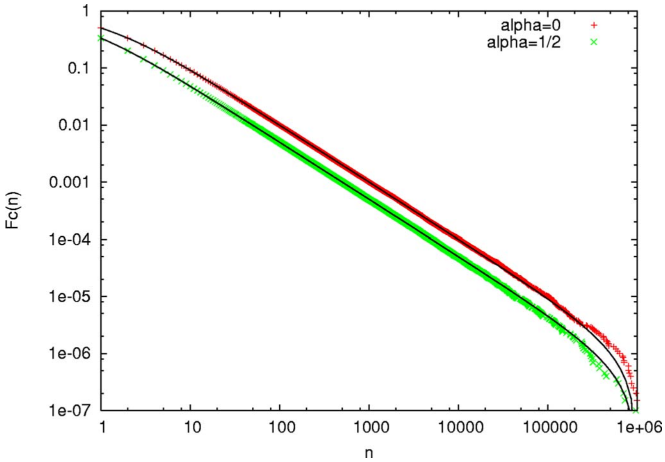

FIG. 3. (Color online) Figure 3 shows comparison of empirical CCDFs of cluster size distributions (points) with analytic formula (24) (lines) on logarithmic plots, at $\alpha=0$, and 1/2. Empirical distributions have been obtained from 10 realizations of $N=10^6$ size networks.

$$F_\infty^c(n) = \frac{1-\alpha}{n+1-\alpha}, \quad F_\infty^c(q) = \frac{(1/\alpha - 1)_{1/\alpha}}{(q+1/\alpha-1)_{1/\alpha}}, \qquad (26)$$

where $0 \leqslant n$, $0 \leqslant q$ and $0 < \alpha < 1$.

Comparison of analytic CCDF of cluster size (24) and empirical distributions are shown in Fig. 3 for $\alpha=0$, 1/3, 1/2, and 2/3. Experimental data has been collected from 10 realizations of 10^6 node networks. Figure 3 shows that simulations fully confirm our analytic result.

On Fig. 4 analytic formula (25) and the empirical CCDFs of in-degree, obtained from the same 10^6 node realizations, are compared. Note the precise match of the simulation and the theoretical distribution on almost the whole data range. Some small discrepancy can be observed around the low probability events. This deviation is caused by the aggregation of errors on the cumulative distribution when some rare event occurs in a finite network.

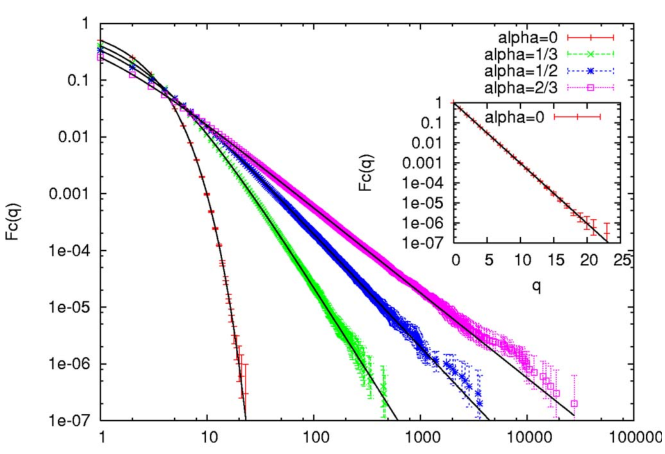

FIG. 4. (Color online) Figure 4 shows comparison of empirical CCDFs of in-degree distributions (points) with analytic formula (25) (lines) on logarithmic plots, at $\alpha=0$, 1/3, 1/2, and 2/3. Empirical distributions have been obtained from 10 realizations of $N=10^6$ size networks. Inset: Comparison at $\alpha=0$ on semi-logarithmic plot.

VII. CONDITIONAL PROBABILITIES AND EXPECTATION VALUES

In the previous sections exact joint and marginal distributions of cluster size and in-degree have been analyzed for both finite and infinite networks. All these distributions provide general statistics of the network. In this section we proceed further, and we investigate the scenario when the younger in-degree of a randomly selected link is known. We ask the cluster size distribution under this condition, that is the conditional distribution $\mathbb{P}_\tau(n|q)$. The results of the previous sections are referred to below to obtain the conditional probability distribution, and eventually the conditional expectation of cluster size. For the sake of completeness, the conditional distribution and expectation of in-degree are given as well at the end of this section.

The conditional cluster size distribution can be given by the quotient of the joint and the marginal in-degree distributions by definition

$$\mathbb{P}_\tau(n|q) = \frac{\mathbb{P}_\tau(n,q)}{\mathbb{P}_\tau(q)}. \qquad (27)$$

The exact conditional distribution for any finite network can be obtained after substituting (15) and (18) into the above expression. For a boundless network the conditional distribution takes the simpler form

$$\mathbb{P}_\infty(n|q) = \alpha \frac{(2/\alpha - 1)_{q+1}}{(2-\alpha)_{n+1}} \Phi_\alpha(n,q), \qquad (28)$$

where $0 \leqslant q \leqslant n$. If $n \gg 1$, then $\mathbb{P}_\infty(n|q) \sim \alpha(2/\alpha - 1)_{q+1}/n^3 + O(1/n^4)$, that is the conditional cluster size distribution falls faster than the ordinary cluster size distribution. It follows that the mean of the conditional cluster size distribution will not diverge like the mean of the ordinary distribution.

What is the expected size of a cluster under the condition that the in-degree of its root is known? For practical reasons, we do not calculate $\mathbb{E}_\tau\{n|q\}$ directly, but we calculate $\mathbb{E}_\tau\{n+2-\alpha|q\} = \mathbb{E}_\tau\{n|q\} + 2 - \alpha$ instead

$$\mathbb{E}_\tau\{n+2-\alpha|q\} = \frac{1}{\mathbb{P}_\tau(q)} \sum_{n=q}^{\tau-1} (n+2-\alpha)\mathbb{P}_\tau(n,q). \qquad (29)$$

Since $(n+2-\alpha)\mathbb{P}_\tau(n,q) = \frac{\tau+1-\alpha}{\tau} \frac{(1/\alpha-1)_q}{(2-\alpha)_n} \Phi_\alpha(n,q)$, the above summation can be given similarly to the marginal distribution $\mathbb{P}_\tau(q)$ in (18)

$$\sum_{n=q}^{\tau-1} (n+2-\alpha)\mathbb{P}_\tau(n,q) = \frac{\tau+1-\alpha}{\tau} \frac{1/\alpha - 1}{q+1/\alpha-1} - \frac{\tau+1-\alpha}{\tau} \frac{(1/\alpha-1)_q}{(2-\alpha)_{\tau-1}} \times \sum_{k=0}^{q} \frac{(-1)^k}{k!(q-k)!} \frac{(-\alpha k)_\tau}{\alpha k + 1 - \alpha}.$$

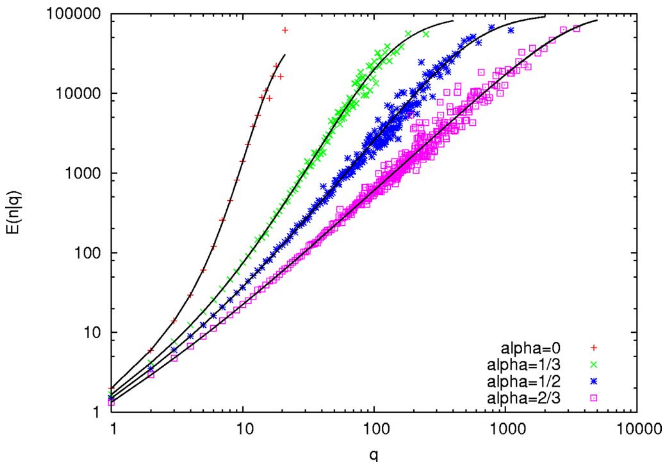

FIG. 5. (Color online) Figure 5 shows the average cluster size as the function of the in-degree q, obtained from 100 realizations of 10^5 size networks. Simulation data has been collected at $\alpha=0$, 1/3, 1/2, and 2/3 parameter values. Analytical result (30) of conditional expectation $\mathbb{E}_\tau\{n|q\}$ is shown with continuous lines.

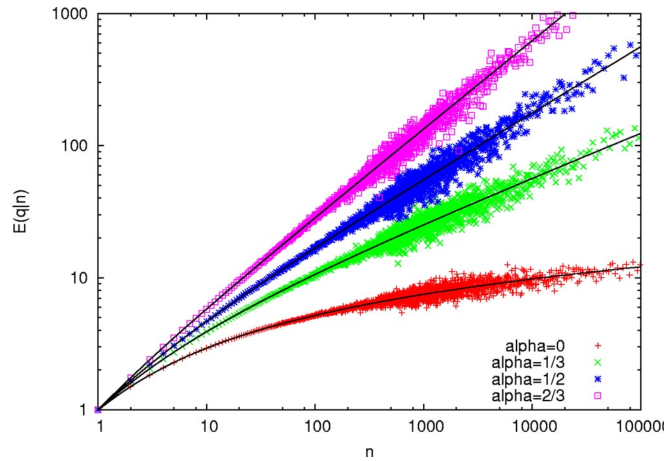

FIG. 6. (Color online) Figure 6 shows the average in-degree as the function of the cluster size n, obtained from 100 realizations of 10^5 size networks. Simulation data has been collected at $\alpha=0$, 1/3, 1/2, and 2/3 parameter values. Analytical result (35) of conditional expectation $\mathbb{E}_\tau\{q|n\}$ is shown with continuous lines.

After replacing the above sum in $\mathbb{E}_\tau\{n|q\}$, the following equation can be obtained

$$\mathbb{E}_\tau\{n+2-\alpha|q\} = (1-\alpha)\frac{(q+1/\alpha)_{1/\alpha}}{(1/\alpha-1)_{1/\alpha}}G_\tau(q), \quad (30)$$

where

$$G_\tau(q) = \frac{1-\dfrac{(1/\alpha-1)_{q+1}}{(1-\alpha)_\tau}\sum_{k=0}^{q}\dfrac{(-1)^k}{k!(q-k)!}\dfrac{(-\alpha k)_\tau}{k+1/\alpha-1}}{1-\dfrac{(2/\alpha-1)_{q+1}}{(2-\alpha)_\tau}\sum_{k=0}^{q}\dfrac{(-1)^k}{k!(q-k)!}\dfrac{(-\alpha k)_\tau}{k+2/\alpha-1}}. \quad (31)$$

The identity $\lim_{\tau\to\infty} G_\tau(q)\equiv 1$ implies that $G_\tau(q)$ involves the finite scale effects, and the factors preceding $G_\tau(q)$ give the asymptotic form of $\mathbb{E}_\tau\{n+2-\alpha|q\}$

$$\mathbb{E}_\infty\{n+2-\alpha|q\} = (1-\alpha)\frac{(q+1/\alpha)_{1/\alpha}}{(1/\alpha-1)_{1/\alpha}}. \quad (32)$$

It can be seen that the expectation of cluster size, under the condition that the in-degree is known, is finite in an unbounded network. It stands in contrast to the unconditional cluster size, discussed in the previous section, which diverges logarithmically as the size of the network grows beyond every limit.

In the ER limit, the expected conditional cluster size becomes

$$\lim_{\alpha\to 0}\mathbb{E}_\infty\{n+2|q\} = 2^{q+1}. \quad (33)$$

The fundamental difference between the scale-free and non-scale-free networks can be observed again. In the scale-free case the expected conditional cluster size asymptotically grows with the in-degree to the power of $1/\alpha$, while in the later case it grows exponentially. On Fig. 5 the exact analytic formula (30) is compared with simulation results at $\alpha=0$, 1/3, 1/2, and 2/3. The simulations clearly justify our analytic solution.

Let us investigate shortly the opposite scenario, that is when the cluster size is known and the statistics of the in-degree under this condition is sought. The conditional distribution can be obtained from the combination of Eqs. (15) and (17), and the definition

$$\mathbb{P}_\tau(q|n) = \frac{\mathbb{P}_\tau(n,q)}{\mathbb{P}_\tau(n)}. \quad (34)$$

The conditional expectation of in-degree can be acquired by the same technique as the conditional expectation of cluster size. Let us calculate $\mathbb{E}_\tau\{q+1/\alpha-1|n\}=\mathbb{E}_\tau\{q|n\}+1/\alpha-1$ instead of $\mathbb{E}_\tau\{q|n\}$ directly

$$\mathbb{E}_\tau\{q+1/\alpha-1|n\} = \frac{1}{\mathbb{P}_\tau(n)}\sum_{q=0}^{n}(q+1/\alpha-1)\mathbb{P}_\tau(n,q) = \frac{\Gamma(2-\alpha)}{\alpha}(n+1-\alpha)_\alpha, \quad (35)$$

where $0\leqslant n<\tau$. Note, that the conditional expectation of in-degree is independent of τ, that is of the size of the network. In the ER limit the expectation of the in-degree becomes

$$\lim_{\alpha\to 0}\mathbb{E}_\tau\{q|n\} = \Psi(n+1)+\gamma, \quad (36)$$

where $\Psi(x)=\frac{d}{dx}\ln\Gamma(x)$ denotes the digamma function, and $\gamma=-\Psi(1)\approx 0.5772$ is the Euler-Mascheroni constant. Asymptotically the expectation of the in-degree in a scale-free tree grows with the cluster size to the power of α, while in a ER tree it grows only logarithmically, since $\Psi(n+1)=\log n + O(1/n)$. Therefore, conditional in-degree and conditional cluster size are mutually inverses *asymptotically*. Figure 6 shows the analytic solution (35) and simulation data at $\alpha=0$, 1/3, 1/2, and 2/3 parameter values. Simulation data has been collected from 100 realizations of 10^5 size networks.

VIII. CONDITIONAL DISTRIBUTION OF EDGE BETWEENNESS

Using the results of the previous sections, we are finally ready to answer the problem which motivated our work, that is the distribution of the edge betweenness under the condition that the in-degree of the younger node of the link is known. It has been noted at the beginning of Sec. III that the edge betweenness can be expressed with cluster size

$$L = (n+1)(\tau - n). \tag{37}$$

Therefore, conditional edge betweenness can be given formally by the following transformation of random variable n

$$\mathbb{P}_\tau(L|q) = \sum_{n=0}^{\tau-1} \delta_{L,(n+1)(\tau-n)} \mathbb{P}_\tau(n|q). \tag{38}$$

Obviously, $\mathbb{P}_\tau(L|q)$ is nonzero only at those values of L, where (37) has integer solution for n. If

$$n_L = \frac{\tau-1}{2} - \sqrt{\frac{(\tau+1)^2}{4} - L} \tag{39}$$

is such an integer solution of the quadratic equation (37), and $L \neq (\tau+1)^2/4$, then

$$\mathbb{P}_\tau(L|q) = \mathbb{P}_\tau(n_L|q) + \mathbb{P}_\tau(\tau - 1 - n_L|q). \tag{40}$$

If $L=(\tau+1)^2/4$ is integer, then $\mathbb{P}_\tau(L|q)=\mathbb{P}_\tau(n_L|q)$.

The conditional expectation of edge betweenness can be obtained from (37)

$$\mathbb{E}_\tau\{L|q\} = \tau \mathbb{E}_\tau\{n+1|q\} - \mathbb{E}_\tau\{(n+1)n|q\}. \tag{41}$$

Therefore, for the exact calculation of $\mathbb{E}_\tau\{L|q\}$ the first and the second moment of the conditional cluster size distribution are required. The first moment, that is the mean, has been derived in the previous section. In order to calculate the second moment let us use the technique we have developed in the previous sections. Let us consider

$$\mathbb{E}_\tau\{(n+2-\alpha)(n+1-\alpha)|q\} = \frac{\tau+1-\alpha}{\tau} \frac{(1/\alpha-1)_q}{\mathbb{P}_\tau(q)} \sum_{n=q}^{\tau-1} \frac{\Phi_\alpha(n,q)}{(2-\alpha)_{n-1}}. \tag{42}$$

We shall be cautious when the summation for n is evaluated. The $k=1$ term in $\Phi_\alpha(n,q)=\sum_{k=0}^{q}\frac{(-1)^k}{k!(q-k)!}(-\alpha k)_n$ must be treated separately to avoid a divergent term

$$\sum_{n=q}^{\tau-1} \frac{\Phi_\alpha(n,q)}{(2-\alpha)_{n-1}} = \frac{1-\alpha}{(q-1)!}[\alpha\Psi(\tau-\alpha) - \alpha\Psi(1-\alpha) - \Psi(q) - \gamma] - \frac{1}{\alpha}\frac{1}{(2-\alpha)_{\tau-2}} \sum_{k=2}^{q} \frac{(-1)^k}{k!(q-k)!} \frac{(-\alpha k)_\tau}{k-1}$$

The exact formula for $\mathbb{E}_\tau\{L|q\}$ can be obtained straightforwardly, after (30) and the above expressions have been substituted into (41).

Let us consider the scenario when the size of the network tends to infinity. Equation (37) implies that edge betweenness diverges as $\tau \to \infty$, therefore, L should be rescaled for an infinite network. From the asymptotics of the digamma function $\Psi(\tau-\alpha)=\ln \tau + O(1/\tau)$ it follows that $\mathbb{E}_\tau\{(n+2-\alpha)(n+1-\alpha)|q\}$ grows only logarithmically, slower than the linear growth of $\tau\mathbb{E}_\tau\{n+2-\alpha|q\}$. Therefore, edge betweenness asymptotically grows linearly as the size of the network grows beyond every limit. Let us rescale edge betweenness

$$\Lambda_\tau = \frac{L(\tau)}{\tau+1} \tag{43}$$

and let us consider the limit $\Lambda=\lim_{\tau\to\infty}\Lambda_\tau=n_\Lambda+1$. The CCDF of the rescaled edge betweenness can be given by

$$F^c_\infty(\Lambda|q) = \lim_{\tau\to\infty} \sum_{n=n_{\Lambda_\tau}}^{\tau-1-n_{\Lambda_\tau}} \mathbb{P}_\tau(n|q) = \frac{1}{\mathbb{P}_\infty(q)} \sum_{n=\Lambda-1}^{\infty} \mathbb{P}_\infty(n,q).$$

When the summation has been carried out, the following equation is obtained

$$F^c_\infty(\Lambda|q) = \frac{(2/\alpha-1)_{q+1}}{(2-\alpha)_{\Lambda-1}} \sum_{k=0}^{q} \frac{(-1)^k}{k!(q-k)!} \frac{(-\alpha k)_{\Lambda-1}}{k+2/\alpha-1}, \tag{44}$$

where $q+1 \leqslant \Lambda$. If $1 < q \ll \Lambda$, then only the first term of the sum should be taken into account, and it is easy to see that

$$F^c_\infty(\Lambda|q) = \frac{\alpha^2(1-\alpha)}{2\Gamma(2/\alpha-1)} \frac{q^{2/\alpha}}{\Lambda^2} + O(1/\Lambda^{2+\alpha}). \tag{45}$$

It can be seen that the scaling exponent -2 is independent of α. The above asymptotic formula has been obtained for infinite networks. The same power law scaling can be observed in finite size networks as (45) if $\Lambda_\tau \ll \tau$. However, $F^c_\tau(\Lambda_\tau|q)\equiv 0$ if $\Lambda_\tau > \tau$ in finite networks, therefore asymptotic formula (45) evidently becomes invalid if $\Lambda_\tau \approx \tau$.

It is obvious that as the size of the network grows larger and larger, asymptotic formula (44) becomes more and more accurate. One can ask how fast the convergence is. From elementary estimations of $F^c_\tau(\Lambda_\tau|q)$ one can show that for fixed Λ_τ

$$F^c_\tau(\Lambda_\tau|q) = F^c_\infty(\Lambda_\tau|q) - (1 - F^c_\infty(\Lambda_\tau|q)) \frac{\alpha^2(1-\alpha)}{2} \frac{1}{\tau^2} + O(1/\tau^{2+\alpha}), \tag{46}$$

that is corrections to the asymptotic formula decrease with τ^{-2} for large τ.

On Fig. 7 comparison of analytic formula (44) with simulation results is presented for $q=1$ and $q=2$. The empirical CCDF of rescaled edge betweenness, under the condition that in-degree q is known, is shown for 10^4, 10^5, and 10^6 size networks, at $\alpha=1/2$ parameter value. The empirical CCDFs of rescaled edge betweenness evidently collapse to the same curve for different size networks, and they coincide precisely with our analytic result.

The expectation of the rescaled edge betweenness under the condition that in-degree q is known can be given by $\mathbb{E}_\infty\{\Lambda|q\}=\mathbb{E}_\infty\{n_\Lambda+1|q\}$. Using (32) and (33) we get

$$\mathbb{E}_\infty\{\Lambda|q\} = (1-\alpha)\frac{(q+1/\alpha)_{1/\alpha}}{(1/\alpha-1)_{1/\alpha}} - 1 + \alpha, \tag{47}$$

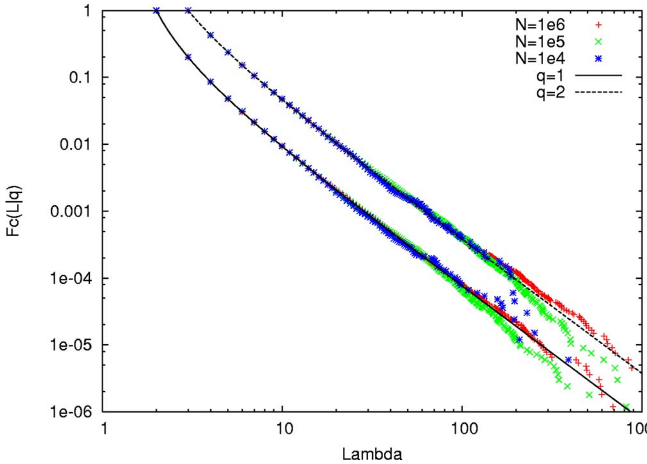

FIG. 7. (Color online) Figure 7 shows CCDF of edge betweenness under the condition that the in-degree q is known. Empirical CCDF has been obtained from 100 realizations of $N=10^4$ and $N=10^5$, and 10 realizations of $N=10^6$ size networks at $\alpha=1/2$ parameter value. Continuous lines show analytic result (44).

$$\lim_{\alpha \to 0} \mathbb{E}_\infty\{\Lambda | q\} = 2^{q+1} - 1. \qquad (48)$$

One can see that $\mathbb{E}_\infty\{\Lambda|q\} \sim q^{1/\alpha}$ for $q \gg 1$ if $\alpha > 0$ and $\mathbb{E}_\infty\{\Lambda|q\} \sim e^q$ for $q \gg 1$ if $\alpha \to 0$.

Analytic results (47) and (48), and simulation data are shown in Fig. 8 at $\alpha=1/2$ and $\alpha=0$ parameter values. Numerical data has been collected from the same 10^4, 10^5, and 10^6 size networks as above. As the size of the network grows, a larger and larger range of the rescaled empirical data collapses to the same analytic curve. On the high degree region some discrepancy can be observed due to the finite scale effects.

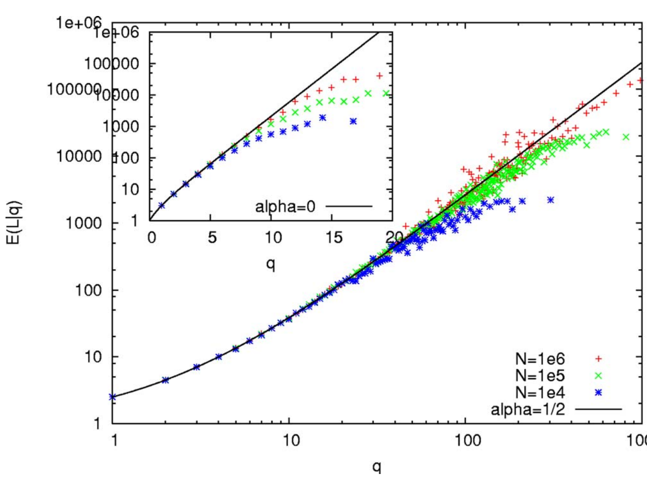

FIG. 8. (Color online) Figure 8 shows average edge betweenness under the condition that the in-degree q is known as the function of q on log-log plot. Numerical data has been collected from 100 realizations of $N=10^4$ and $N=10^5$, and 10 realizations of $N=10^6$ size networks at $\alpha=1/2$ parameter value. Inset shows the same scenario at $\alpha=0$ parameter value on semi-logarithmic plot. Continuous lines show analytic results (47) and (48).

Finally, let us note that the precise unconditional distribution of edge betweenness $\mathbb{P}_\tau(L) = \sum_{n=0}^{\tau-1} \delta_{L,(n+1)(\tau-n)} \mathbb{P}_\tau(n)$ can be obtained from (17) as well. Furthermore, CCDF of the unconditional betweenness $F^c_\tau(L) = \sum_{n=n_L}^{\tau-n_L-1} \mathbb{P}_\tau(n)$ can be derived in closed form

$$F^c_\tau(L) = \frac{\tau+1-\alpha}{\tau} \frac{(1-\alpha)(\tau-2n_L)}{(n_L+1-\alpha)(\tau-n_L+1-\alpha)}. \qquad (49)$$

For the sake of simplicity we have assumed during our calculations that in-degrees of the younger nodes are provided. However, it is possible that even though both two in-degrees of every link are known, we cannot distinguish them from each other, that is we cannot tell which is the younger node. How could we extend our results to this scenario? Let us consider a new edge when it is connected to the network. The in-degree of the new node is obviously 0. The in-degree of the other node, which the new node is connected to, is equal to or larger than one. Due to preferential attachment the larger the in-degree is the faster it grows. Even if preferential attachment is absent, the growth rate of every in-degree is the same. Therefore, it is expected that the initial deficit in the in-degree of the younger node grows or remains at the same level during the evolution of the network. It follows that it is a reasonable approximation to substitute the in-degree of the younger node q with $q_{\min} = \min(q_1, q_2)$ in our formulas.

IX. CONCLUSIONS

A typical network construction problem is to design network infrastructure without wasting precious resources at places where not needed. An appropriate design strategy is if network resources are allocated proportionally to the expected traffic. In a mean field approximation the expected traffic is proportional to the number of shortest paths going through a certain network element, that is the betweenness.

The precise calculation of all the betweenness require complete information on the network structure. In real life, however, the number of shortest paths is often impossible to tell because the structure of the network is not fully known. One of the practical results of this paper is that the expectation of edge betweenness can be estimated precisely when a limited local information on network structure—the in-degree of the younger node—is available.

Another difficulty of network design is that the size of real networks is finite. Moreover, the size of real networks is often so small that asymptotic formulas can be applied only with unacceptable error. The other important subject of our results is that the derived formulas are exact even for finite networks, which allows better design of realistic finite size networks.

Various statistical properties of evolving random trees have been investigated in this paper. We have focused on the cluster size, the in-degree and the edge betweenness. We have considered the $m=1$ case of the BA model extended with initial attractiveness for modeling random trees. Initial attractiveness allows fine tuning of the scaling parameter. Moreover, in the limit of the tuning parameter $\alpha \to 0$ the applied model tends to a non-scale-free structure, which is in many aspects similar to the classical ER model. Therefore, we were able to investigate both the scale-free and the non-scale-free scenario within the same framework.

First, the evolution of cluster size and in-degree of a specific edge have been modeled as a bivariate Markov process. The master equation, associated with the Markov process, has led us to a linear partial difference equation. An exact analytic solution of the master equation, which satisfies the initial conditions as well, has been found. The solution provides the joint probability distribution of cluster size and in-degree for a specific edge.

Using the above results we have derived the joint probability distribution of cluster size and in-degree for a randomly selected edge. It is of more practical importance than the joint distribution for a specific edge because, in contrast to the former distribution, it provides the statistical description of the whole network. We also derived the joint distribution in the ER limit. Note that the obtained formulas are exact for even finite size networks. In addition, the formulas for unbounded networks have been presented as well.

We have continued our analysis with the one-dimensional marginal distributions. We have shown some fundamental differences in the scaling properties of the marginal cluster size and in-degree distributions. Our results here, compared to previous results in the literature, is that we have found exact analytic formulas not only for the large, but also for the small cluster size and in-degree region.

Although the marginal distributions have their own importance, we have derived them in order to obtain conditional probability distributions. From the combination of the joint and the marginal distributions we have given the conditional distributions of cluster size and in-degree. We have also presented conditional expectations of cluster size and in-degree for both finite and unbounded networks. We have found that asymptotically the conditional cluster size grows with in-degree to the power of $1/\alpha$ and the conditional in-degree grows with cluster size to the power of α, respectively. The ER limit has been discussed as well. We have shown that the conditional cluster size grows exponentially and the conditional in-degree grows logarithmically when $\alpha \to 0$.

Finally, by applying the transformation of random variables we have derived the distribution of edge betweenness under the condition that the corresponding in-degree is known. We have found that the conditional expectation of edge betweenness grows linearly with the size of the network. For the analysis of unbounded networks we have defined the rescaled edge betweenness Λ, and derived its distribution and expectation under the condition that in-degree q is provided. Our analytic results have been verified at different network sizes and parameter values by extensive numerical simulations. We have demonstrated that numerical simulations fully confirm our analytic results.

For the future, we hope that the methods we have developed in this paper allow us to describe cluster size and edge betweenness in more general scenarios. For example, when not only the younger, but both two in-degrees of links are considered.

ACKNOWLEDGMENTS

The authors thank the partial support of the National Science Foundation (OTKA T37903), the National Office for Research and Technology (NKFP 02/032/2004 and NAP 2005/ KCKHA005) and the EU IST FET Complexity EVERGROW Integrated Project. The authors also thank Máté Maródi for the fruitful discussions and his helpful comments.

APPENDIX A: EXPANSION OF THE KRONECKER-DELTA FUNCTION

We have seen that the general solution of Eq. (7) is $\mathbb{P}_\tau(n,q|\tau_e)=\Sigma_{\lambda_1,\lambda_2}C_{\lambda_1,\lambda_2}f(\tau)g(n)h(q)$, and the initial condition is $\mathbb{P}_{\tau_e}(n,q|\tau_e)=\delta_{n,0}\delta_{q,0}$, where

$$\delta_{n,m} = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m \end{cases} \tag{A1}$$

is the Kronecker-delta function, and n and m are integers. Coefficients C_{λ_1,λ_2} are calculated in this section. First we show that

$$\delta_{n,0} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{1}{\Gamma(n-k+1)}. \tag{A2}$$

Note that we can consider $m=0$ without any loss of generality, since $\delta_{n,m} \equiv \delta_{n-m,0}$.

If $n<0$, then the summand in (A2) is zero by definition, indeed. If $n>0$, then

$$\sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{1}{\Gamma(n-k+1)} = \frac{1}{n!}\sum_{k=0}^{n} \binom{n}{k}(-1)^k = 0 \tag{A3}$$

follows from the binomial theorem. Finally, for $n=0$,

$$\sum_{k=0}^{0} \frac{(-1)^k}{k!} \frac{1}{\Gamma(-k+1)} = \frac{(-1)^0}{0!}\frac{1}{\Gamma(1)} = 1. \tag{A4}$$

Coefficients C_{λ_1,λ_2} can be obtained from the term by term comparison of $\mathbb{P}_{\tau_e}(n,q|\tau_e)=\Sigma_{\lambda_1,\lambda_2}C_{\lambda_1,\lambda_2}f(\tau_e)g(n)h(q)$ with the expansion of the initial condition $\delta_{n,0}\delta_{q,0}$, shown above. One can easily confirm with the help of identity $f(n)\delta_{n,0} \equiv f(0)\delta_{n,0}$ that the same terms appear on both sides, if $\lambda_1 = -k_1$, and $\lambda_2 = -\alpha k_2$, and coefficients C_{k_1,k_2} are the following:

$$C_{k_1,k_2} = \frac{(-1)^{k_1+k_2}}{k_1!k_2!}\frac{\Gamma(\tau_e+1-\alpha)}{\Gamma(\tau_e-k_1)}\frac{1}{\Gamma(-\alpha k_2)}\frac{1}{\Gamma(1/\alpha-1)}. \tag{A5}$$

Finally, to obtain (15) the summation for k_1 can be carried out explicitly

$$\sum_{k_1=0}^{n} \frac{(-1)^{k_1}}{k_1!\Gamma(n-k_1+1)}\frac{\Gamma(\tau-k_1)}{\Gamma(\tau_e-k_1)} = \frac{\Gamma(\tau-\tau_e+1)}{\Gamma(n+1)\Gamma(\tau_e)}\frac{\Gamma(\tau-n)}{\Gamma(\tau-\tau_e-n+1)}.$$

APPENDIX B: THE $\alpha \to 0$ LIMIT OF JOINT DISTRIBUTION $\mathbb{P}_\tau(n,q)$

In this section we prove that the ER limit of the joint probability $\mathbb{P}_\tau(n,q)$ is (16).

Theorem *Let us consider $\mathbb{P}_\tau(n,q)$ as defined in (15), where $0<q<n<\tau$ are integers. Then the following limit holds*

$$\lim_{\alpha\to 0}\mathbb{P}_\tau(n,q)=\frac{\tau+1}{\tau\Gamma(n+3)}\sum_{k=q-1}^{n-1}(-1)^{n-1-k}S_{n-1}^{(k)}\binom{k}{q-1}. \tag{B1}$$

Proof. First, let us note that $\Phi_\alpha(n,q)$ in (15) can be rewritten in the following equivalent form: $\Phi_\alpha(n,q)=\alpha\sum_{k=0}^{q-1}\frac{(-1)^k(1-\alpha-\alpha k)_{n-1}}{k!(q-1-k)!}$. Next, Pochhammer's symbol $(1/\alpha-1)_q$ is replaced with its asymptotic form: $(1/\alpha-1)_q=1/\alpha^q(1+O(\alpha))$. After the obvious limits have been evaluated the following equation is obtained

$$\lim_{\alpha\to 0}\mathbb{P}_\tau(n,q)=\frac{\tau+1}{\tau\Gamma(n+3)}\lim_{\alpha\to 0}\frac{\sum_{k=0}^{q-1}\frac{(-1)^k(1-\alpha-\alpha k)_{n-1}}{k!(q-1-k)!}}{\alpha^{q-1}}. \tag{B2}$$

The above limit, by definition, can be substituted with $q-1$ order differential at $\alpha=0$, if all the lower order derivates of the sum are zero at $\alpha=0$. Indeed,

$$\lim_{\alpha\to 0}\frac{\sum_{k=0}^{q-1}\frac{(-1)^k(1-\alpha-\alpha k)_{n-1}}{k!(q-1-k)!}}{\alpha^{q-1}} = \frac{1}{m!}\frac{d^m}{d\alpha^m}\sum_{k=0}^{q-1}\frac{(-1)^k(1-\alpha-\alpha k)_{n-1}}{k!(q-1-k)!}\Bigg|_{\alpha=0} = \frac{1}{m!}\frac{d^m(1+\alpha)_{n-1}}{d\alpha^m}\Bigg|_{\alpha=0}\sum_{k=0}^{q-1}\frac{(-1)^k(-k-1)^m}{k!(q-1-k)!},$$

where the sum is 0 if $m<q-1$ and 1 if $m=q-1$. Therefore, the limit can be transformed to

$$\lim_{\alpha\to 0}\mathbb{P}_\tau(n,q)=\frac{\tau+1}{\tau\Gamma(n+3)}\frac{1}{(q-1)!}\frac{d^{q-1}(1+\alpha)_{n-1}}{d\alpha^{q-1}}\Bigg|_{\alpha=0}. \tag{B3}$$

Finally, let us consider the power expansion of Pochhammer's symbol: $(x)_m=\sum_{k=0}^{m}(-1)^{n-k}S_m^{(k)}x^k$, where $S_m^{(k)}$ are the Stirling numbers of the first kind. The expansion formula has been applied at $x=1+\alpha$ and $m=n-1$, which implies

$$\lim_{\alpha\to 0}\mathbb{P}_\tau(n,q) = \frac{\tau+1}{\tau\Gamma(n+3)}\sum_{k=q-1}^{n-1}\frac{(-1)^{n-1-k}S_m^{(k)}}{(q-1)!}\frac{d^{q-1}(1+\alpha)^k}{d\alpha^{q-1}}\Bigg|_{\alpha=0} = \frac{\tau+1}{\tau\Gamma(n+3)}\sum_{k=q-1}^{n-1}(-1)^{n-1-k}S_{n-1}^{(k)}\binom{k}{q-1}.$$

□

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[6] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **281**, 69 (2000).
[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, SIGCOMM Comput. Commun. Rev. **29**, 251 (1999).
[8] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[9] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[10] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[11] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[12] B. Bollobás, Discrete Math. **33**, 1 (1981).
[13] G. Szabó, M. Alava, and J. Kertész, Phys. Rev. E **66**, 026101 (2002).
[14] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[15] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[16] B. Bollobás and O. Riordan, Phys. Rev. E **69**, 036114 (2004).
[17] J. Szymański, in *Random Graphs '85*, edited by M. Karonski and Z. Palka (North-Holland, Amsterdam, 1987), Vol. 144 of North-Holland Math. Stud., pp. 297–305.
[18] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[19] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[20] A. M. Odlyzko, in *Handbook of Combinatorics*, edited by R. L. Graham, M. Götschel, and L. Lovász (Elsevier, Amsterdam, 1995), Vol. 2, pp. 1063–1229.