Effect of random edge failure on the average path length

This article has been downloaded from IOPscience. Please scroll down to see the full text article.
(http://iopscience.iop.org/1751-8121/44/41/415002)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 202.195.129.244
The article was downloaded on 22/09/2011 at 03:10

Please note that terms and conditions apply.
Effect of random edge failure on the average path length

Dongchao Guo, Mangui Liang, Dandan Li and Zhongyuan Jiang

Institute of Information Science, Beijing Jiaotong University, 100044, Beijing,
People’s Republic of China

E-mail: mgliang58@gmail.com and 08112070@bjtu.edu.cn

Received 27 June 2011, in final form 29 August 2011
Published 20 September 2011
Online at stacks.iop.org/JPhysA/44/415002

Abstract
We study the effect of random removal of edges on the average path length (APL) in a large class of uncorrelated random networks in which vertices are characterized by hidden variables controlling the attachment of edges between pairs of vertices. A formula for approximating the APL of networks suffering random edge removal is derived first. Then, the formula is confirmed by simulations for classical ER (Erdős and Rényi) random graphs, BA (Barabási and Albert) networks, networks with exponential degree distributions as well as random networks with asymptotic power-law degree distributions with exponent $\alpha > 2$. 

PACS numbers: 89.75.Hc, 02.50.−r

1. Introduction

Complex networks which are ubiquitous in our world usually display robustness in different levels, e.g., scale-free (SF) networks exhibit strong resistance to random failure of vertices [1] but weakness in suffering intentional attack on most connected vertices [2]. It is usually to ascribe the property of strong to random removal but weak to attack of the SF network to its power-law degree distribution [3, 4]. Thus, the robustness of a complex network can be defined as the capability of maintaining its original structural (static) or dynamic properties. It is of interest to investigate the influence of random failure of vertices on the statistical properties of complex networks, such as the average path length (APL) [5], which is computed by averaging the values of the length of shortest paths between all pairs of vertices in a specific class of networks. Although several formulas have been proposed to estimate the APL [6–8], they are either no more than scaling relations (i.e. rough estimations rather than a formula) of the APL or conceptually a little complex. Based on the so-called hidden variable network model generalized in [12], Fronczak et al [11] derived a formula for the average distance between each pair of vertices characterized by given values of hidden variables $h_i$ and $h_j$ and attained good agreement only for dense networks (i.e. networks with sizable quantity of edges) between
theoretical predictions and numerical results. Holme et al [5] also proposed numerical studies for the change of the APL after complex networks suffer random removal. There is still lack of a satisfactory exact analytical method to estimate the change of the APL after random failure takes place.

On the other hand, relatively rare attention has been paid to measure the robustness in terms of edge. He et al [9] studied the robustness of several network models subjected to edge removal. They proposed an exact formula for the critical fraction of removed edges causing a network to break into a giant component and a finite component and pointed out that the finite component is mostly of size 1. Guillaume et al [10] argued that the robustness of a network in terms of edge is vastly different from that of the network in the form of vertex removal. A distinguishing feature of random edge removal is that the removal process almost does not get the number of vertices changed after a small fraction of edges fail. This fact discriminates edge removal from vertex removal.

In this paper, based on the hidden variable network model [12] and the APL estimation formula derived in [11], we derive an approximate formula characterizing the change of the APL of uncorrelated random networks after suffering random failure of edges. Our work provides related researchers a comparatively precise approximate formula to estimate the extent of the damage (caused by random edge removal) a network undergoes in terms of the change of the APL. The formula derived performs better in a dense network rather than in a sparse network. The significance of our work is elaborated as follows. The edge (or link) in a network usually demonstrates for the information or energy transmitting tunnel. Thus, plenty of accidents happened in real world (e.g. a wire of a power grid is broken, an optical fiber in an optical network fails, a road is jammed) can be abstracted as the random edge removal phenomena. The formula derived in this paper can, at least for dense networks, predict the possible consequences caused by the above-mentioned accidents and may further direct us to avoid the adverse effect in advance (e.g. our work may help to design a robust traffic network and a novel traffic control strategy avoiding that vast plenty of jams emerge simultaneously which incurs dramatic increases in the APL). To those who concern percolation phenomena and virus spreading, our work also makes sense. By generalizing percolation theory, Cohen et al analyzed the stability of networks under random failures [2]. One can naturally relate our work to the study of percolation phenomena. Moreover, random edge removal will also make virus spreading slower as the APL increases. The rest of this paper is organized as follows. In section 2, we propose the formula. In section 3, simulations for different sorts of complex networks will be performed to confirm the formula. Finally, we conclude this paper in section 4.

2. Theoretical solution measuring effect of random edge failure on the APL

Before elaborating the network model on which our study is based, some other models should first be introduced for the sake of comparison. A network where edges are placed randomly and uniformly between all vertex pairs owns the Poisson degree distribution. The classical Erdös and Rényi (ER) model of such networks was proposed in [15] and has been deeply studied, although it is too random to describe real-world networks. Evolving networks based on preferential attachment has been introduced by Barabási and Albert [13] to interpret the origin of SF degree distribution. Without using the preferential attachment scheme, Caldarelli et al [14] proposed another static network generating model which can account for the emergence of the SF degree distribution. In this model, one first assigns each vertex a tag $h_i$ (hidden variable, fitness) which is randomly obtained from a probability distribution $\rho(h)$. Then, the probability of assigning an edge between a pair of vertices, $i$ and $j$, depends on the function $2$
\( f(h_i, h_j) \), where \( h_i \) and \( h_j \) are the tags assigned to vertices \( i \) and \( j \), and \( f(h_i, h_j) \) is the function of these two tags. Boguña and Pastor-Satorras [12] generalized the above network models and proposed a common formalism that can describe the classical ER graph [15], best-known Barabási and Albert (BA) model [13] and the model proposed by Caldarelli et al [14]. In what follows, this common formalism will be called the hidden variable model which is also the basis of our work.

We briefly rehearse the main characteristics of the hidden variable model [12]. In order to generate a network of vertices \( N \) with degree distribution \( P(k) \) as expected, each vertex of the network should first be assigned a hidden variable \( h_i \) randomly drawn from the hidden variable distribution \( R(h) \). Denote by \( p_{ij} \) the establishment probability of an edge between vertices \( i \) and \( j \).

The relation between the degree distribution \( P(k) \) of the uncorrelated network constructed in this manner and the hidden variable \( R(h) \) is

\[
P(k) = \int_0^\infty \frac{e^{-h}h^k}{k!} R(h) \, dh,
\]

or in the discrete form

\[
P(k) = \sum_{h=0}^\infty \frac{e^{-h}h^k}{k!} R(h).
\]

This relation is usually called the Poisson transform. Here, we only focus on the continuous form of the Poisson transform (3). More corollaries can be obtained and listed as follows:

\[
\langle h^n \rangle = \langle k(k-1)\cdots(k-n+1) \rangle.
\]

\[
\langle h \rangle = \langle k \rangle, \quad \langle h^2 \rangle = \langle k^2 \rangle - \langle k \rangle.
\]

Usually, as researchers are more familiar with the degree distribution \( P(k) \) than with the notation of a hidden variable \( R(h) \), the inverse Poisson transform is intriguing to us. Based on the method of the inverse Poisson transform proposed by Wolf and Mehta [17], Fronczak et al [16] took networks with the Poisson degree distribution, exponential degree distribution and power-law degree distribution as examples to demonstrate that one can obtain the concrete form of \( R(h) \) provided one knows the form of corresponding \( P(k) \). For the sake of simplicity, we will repeat the method in appendix A and here we only list some crucial conclusions. By (3), we obtain the relation

\[
\int_0^\infty e^{-h}R(h) e^{i\chi} \, dh = \int_0^\infty e^{-h}R(h) \sum_{k=0}^\infty \frac{(i\chi h)^k}{k!} \, dh
\]

\[
= \sum_{k=0}^\infty (i\chi)^k \int_0^\infty \frac{h^k}{k!} \cdot R(h) \cdot e^{-h} \, dh
\]

\[
= \sum_{k=0}^\infty (i\chi)^k P(k)
\]

\[
= F(i\chi)
\]

\[
= [F(z)]_{z=i\chi},
\]

(7)
where \( F(i\lambda) \) is the generating function \( F(z) \) of \( P(k) \) with \( z = i\lambda \). It is obvious that the generating function \( F(z) \) of \( P(k) \) with \( z = i\lambda \) is equal to the Fourier transform of \( e^{-hR(h)} \) and that

\[
R(h) = \frac{e^h}{2\pi} \int_{-\infty}^{\infty} F(i\lambda) e^{-i\lambda h} d\lambda. \tag{8}
\]

Given \( P(k) \), by (7) and (8), one can obtain the corresponding \( R(h) \) and this procedure is called the inverse Poisson transform. By the way, Fronczak et al [11], based on the hidden variable model and the above-mentioned inverse Poisson transform method, proposed an approximate formula for estimating the APL of random uncorrelated networks given that \( R(h) \) of the network model is known, only suit well for dense networks with plenty of edges. If one obtains the \( P(k) \) of a network, the related \( R(h) \) at the hidden variable level will be derived, followed by the APL by using the formula proposed in [11].

At the moment, we start to concentrate on the issue we care: the change of the APL after networks suffer random edge removal. After randomly deleting a fraction \( f \) of all edges, the new degree distribution is [4]

\[
P_f(k) = \sum_{k_0=k}^{\infty} P(k_0) \binom{k_0}{k} (1-f)^k i^{k_0-k}, \tag{9}
\]

where \( P_f(k) \) denotes the degree distribution after a fraction \( f \) of edges are removed given that the original distribution is \( P(k) \). Because \( P_f(k) \) follows a more complex form rather than the power-law distribution given that \( P(k) \) is of the power-law distribution [18], at least for the case of the power-law distribution it seems hardly to obtain the corresponding \( R(h) \) by (7) and (8) just by \( P_f(k) \). Fortunately, there exists an alternative way to cope with this inverse transform problem. The generating function of \( P_f(k) \), denoted by \( F_f(i\lambda) \), can be rewritten as

\[
F_f(i\lambda) = \sum_{k=0}^{\infty} (i\lambda)^k P_f(k)
\]

\[
= \sum_{k=0}^{\infty} (i\lambda)^k \sum_{k_0=k} P(k_0) \binom{k_0}{k} (1-f)^k i^{k_0-k}
\]

\[
= \sum_{k_0=0}^{\infty} P(k_0) f^{k_0} \sum_{k=0}^{k_0} \binom{k_0}{k} \left( 1 - \frac{f}{i\lambda} \right)^k
\]

\[
= \sum_{k_0=0}^{\infty} P(k_0) f^{k_0} \left( 1 + \frac{1-f}{i\lambda} \right)^{k_0}
\]

\[
= \sum_{k_0=0}^{\infty} P(k_0) (f + (1-f)i\lambda)^{k_0}
\]

\[
= F \left( f + (1-f)i\lambda \right), \tag{10}
\]

where \( F \left( f + (1-f)i\lambda \right) \) is \( F(z) \) in (7) with \( z = f + (1-f)i\lambda \) and \( f \) is the fraction of edges deleted. Equation (10) is a key relationship demonstrating that the generating function of \( P(k) \) changes from \( F(i\lambda) \) to \( F \left( f + (1-f)i\lambda \right) \) after a fraction \( f \) of all edges are removed. For the sake of simplicity, set \( p = f \) and \( q = 1-f \) with \( 0 \leq f \leq 1 \). By (7), \( F \left( f + (1-f)i\lambda \right) \) can be newly expressed as
\[
F(p + qix) = \sum_{k=0}^{\infty} (p + qix)^k \mathcal{P}(k)
= \sum_{k=0}^{\infty} (p + qix)^k \int_0^\infty \frac{h^k}{k!} R(h) e^{-h} dh
= \int_0^\infty \left[ \sum_{k=0}^{\infty} \frac{(ph + qix)^k}{k!} \right] \cdot R(h) e^{-h} dh
= \int_0^\infty e^{-qixh} R(h) e^{-qh} dh.
\] (11)

Let \( h' = qh \) and, from (11), we further obtain
\[
\int_0^\infty e^{-qixh} R(h) e^{-qh} dh = q^{-1} \int_0^\infty e^{ih} R(h q^{-1}) e^{-h'} dh'
\Rightarrow q^{-1} \int_0^\infty e^{ih} R(h q^{-1}) e^{-h} dh.
\] (12)

According to (12), it is obvious that while \( F(ix) \) changes to \( F(f + (1 - f)ix) \), the conversion of \( R(h) \) is
\[
R(h) \rightarrow R_f(h) = (1 - f)^{-1} R(h(1 - f)^{-1}).
\] (13)

Knowing the hidden variable distribution, the formula of the APL proposed in [11] is
\[
l = -2 \langle \ln h \rangle + \ln N + \ln \langle h^2 \rangle - \gamma + \frac{1}{2}.
\] (14)

where \( \gamma \approx 0.5772 \) is Euler’s constant, \( \langle h^2 \rangle \) is the moment of order 2 of \( h \) and \( \beta \) is defined the same as (2). By simply substituting \( R_f(h) \) into (14), one will obtain the formula for estimating the APL of networks with a fraction \( f \) of all edges removed. In what follows, we will apply the formula to different sorts of networks.

Randomly and uniformly drawing edges between all pairs of vertices in a network will yield a network with the Poisson degree distribution following the form
\[
P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad k \geq 0.
\] (15)

The corresponding \( R(h) \) of (15) is (as derived in [16])
\[
R(h) = \delta(h - \langle k \rangle),
\] (16)

where \( \delta(\cdot) \) is the Dirac delta function and \( \langle k \rangle = \langle h \rangle \) is the average degree. Thus, after the breakdown of a fraction \( f \) of all edges, \( R(h) \) turns out to be
\[
R_f(h) = (1 - f)^{-1} R \left( \frac{h}{1 - f} \right)
= (1 - f)^{-1} \delta \left( \frac{h}{1 - f} - \langle k \rangle \right)
= \delta \left( h - (1 - f)\langle k \rangle \right),
\] (17)

where the last step is derived by applying the property of the delta function, \( \delta(ax) = |a| \delta(x) \). Substituting (17) into (14), we obtain
\[
l = \frac{\ln N - \gamma}{\ln ((1 - f)\langle k \rangle)} + \frac{1}{2}.
\] (18)

5
The following is the exponential degree distribution and its related hidden variable distribution [16]:

\[ P(k) = \frac{(\langle k \rangle)^k}{(1 + \langle k \rangle)^{k+1}}, \quad k \geq 0, \]  

\[ R(h) = \frac{e^{-h/\langle k \rangle}}{\langle k \rangle}, \quad h \geq 0. \]  

In the same manner as the Poisson case, we obtain

\[ R_f(h) = \frac{e^{-h/\langle (1-f)\langle k \rangle \rangle}}{(1-f)\langle k \rangle}, \quad h \geq 0. \]  

Theoretically, the corresponding APL formula may not be deduced because \( \langle \ln h \rangle = \int_0^\infty R_f(h) \ln h \, dh \) in (14) cannot be calculated. Instead, networks we actually generate follow the exponential degree distribution with domain \( k \geq 1 \), which ensures that \( \langle \ln h \rangle \) could be done.

The next example is the SF network the degree distribution of which follows the form \( P(k) \sim k^{-\alpha} \). We adopt the model proposed by Fronczak et al [11], possessing the asymptotic SF degree distribution for \( k \gg 1 \) originating from the power-law distribution at the hidden variable level,

\[ R(h) = (\alpha - 1) h^{\alpha - 1} h^{-\alpha}, \quad h = m, \ldots, h_{\text{max}}, \]  

where \( h_{\text{max}} \approx m N^{1/(\alpha - 1)} \). For large \( k \), \( P(k) \) of networks generated by following the above \( R(h) \) follow the power-law form, namely \( P(k) \sim k^{-\alpha} \). We can rewrite \( R(h) \) as in the following form:

\[ R(h) = C h^{-\alpha} [u(h - m) - u(h - h_{\text{max}})], \]  

where \( C = (\alpha - 1) h^{-\alpha} \) is an appropriate normalization constant, and \( u(h) \) is the unit step function following

\[ u(h) = \begin{cases} 
1 & \text{for } h \geq 0 \\
0 & \text{for } h < 0. 
\end{cases} \]  

By (13), we obtain

\[ R_f(h) = C (1 - f)^{-1} [h (1 - f)^{-1}]^{-\alpha} [u(h (1 - f)^{-1} - m) - u(h (1 - f)^{-1} - h_{\text{max}})] 
= C (1 - f)^{-\alpha} h^{-\alpha} [u(h (1 - f) - m) - u(h (1 - f) h_{\text{max}})]. \]  

Substituting (24) into (14) (see appendix B), we obtain

\[ l = \frac{-2 \ln h}{\ln h^2 - \ln (h^2) + \ln (1 - f) + \frac{1}{2}}. \]  

It is obvious that the discrepancy between (14) and (25) is nothing more than the term \( \ln (1 - f) \) in the denominator of (25). Thus, we can obtain formulas for different scale exponents via only a bit modification respectively to the three formulas, proposed in [16], for the SF distribution with scale exponent \( \alpha > 3, \alpha = 3 \) and \( 2 < \alpha < 3 \), respectively. Thus for \( \alpha > 3 \),

\[ p^{\alpha>3} = \frac{\ln N + \ln \left( \frac{e-1}{e-2} \right) - \frac{2}{\alpha-1} - \gamma}{\ln \left( \frac{e-2}{e-3} \right) + \ln m + \ln (1 - f) + \frac{1}{2}}. \]  

For \( \alpha = 3 \),

\[ p^{\alpha=3} = \frac{\ln N + \ln \ln N - 1 - \gamma}{\ln \ln N + \ln \left( \frac{e}{e-2} \right) + \ln (1 - f) + \frac{1}{2}}. \]
For $2 < \alpha < 3$, 
\[
P^{\alpha<3} = \left(\frac{2}{\alpha-1}\right) \ln N + \ln \left(\frac{\alpha-1}{\alpha-2}\right) - \left(\frac{2}{\alpha-1}\right) - \gamma \ln \ln N + \ln m + \ln (1 - f) + \frac{1}{2}. 
\]
(28)

The final example is the BA SF network [13], the evolving model associated with the preferential attachment scheme. With an initial network of $m_0$ vertices, each time step a new vertex $v_i$ is added to the existing network and $m$ new edges are introduced to link between $v_i$ and other vertices. The probability $p_{ij}$ of establishing an edge between $v_i$ and any other vertex $v_j$ in the existing network is $p_{ij} = m k_j / \sum k_i$. This procedure of constructing a network leads to [11, 13]
\[
P(k) \sim k^{-3}, \quad k \geq m 
\]
(29)
\[
R(h) = \frac{2}{N} h^{-3}, \quad h = 1/\sqrt{N}, \ldots, 1, 
\]
(30)
\[
l = \frac{\ln N + \ln \ln N - 1 - \gamma}{\ln \ln N + \ln \left(\frac{\alpha}{2}\right)} + 1/2, 
\]
(31)
where in (31) $\gamma = 0.5772$. In the same manner, we obtain the APL formula
\[
l = \frac{\ln N + \ln \ln N - 1 - \gamma}{\ln \ln N + \ln \left(\frac{\alpha}{2}\right) + \ln (1 - f)} + \frac{1}{2} 
\]
(32)
for BA networks of which a fraction $f$ of all edges are removed.

3. Simulation results

In order to prove the above-listed formulas, simulations for networks with different hidden variable distributions (i.e. with various degree distributions) have been performed. By following the construction method of networks with hidden variables, we first generate plenty of networks following a specific hidden variable distribution. Then we select a network, randomly remove a fraction $f$ of all its edges and calculate its average path length. We repeat this procedure and finally average the values of the APL of the generated networks suffering random edge removal at failure fraction $f$. According to the experimental study in [9], the connectivity of a network will be definitely destroyed after a fraction $f_c$ of edges are removed randomly. Taking ER networks for example, $f_c = (\langle k \rangle - \ln N) / \langle k \rangle$, where $\langle k \rangle$ is the average degree and $N$ is the original size of the network. A randomly generated ER network with average degree $\langle k \rangle = 10$ and size $N = 2000$ definitely breaks into a giant component and some small components if a fraction $f_c \approx 0.23$ of all edges are removed. Even when only a small proportion edge removal occurs, the network can also get fragmented. While using the formula derived we thus assume that after suffering random edge removal, the network still maintains its connectivity or at least keeps the size of the giant component in the same order as $N$ (i.e. that a sizeable proportion of vertices are disconnected is not allowed). If the connectivity of the network is destroyed, we calculate only the APL of its giant component as long as the size of the giant component is approximate to $N$. Also, as the removal fraction increases, the size of the giant component decreases. Therefore, we test only the network with just a small proportion removal and obtain samples in the range of $0 \leq f \leq 0.1$, where $f$ is the removal fraction.

Figure 1 illustrates the predictions of (18) in comparison with numerical results of simulations in ER graphs. There exists well coincidence between analytical predictions and
Figure 1. APL $l$ versus the fraction of removed edges $f$ in ER random graphs with $\langle k \rangle = 10$. The scatter points stand for simulation results averaged over $10^3$ realizations. Solid curves with open circles, open squares, open diamonds and open triangles represent theoretical predictions for networks with sizes $N = 200$, $N = 500$, $N = 10^3$ and $N = 2 \cdot 10^3$ respectively.

Figure 2. APL $l$ as a function of the fraction of removed edges $f$ for networks with asymptotic exponential degree distributions. Curves with open graphics represent the values of the numerical calculation of the analytic formula, while scatter points express simulation results.

Simulation results, which can be ascribed to the fact that networks with Poisson degree distributions always maintain the Poisson form even after random edge (vertex) removal occurs. However, this elegant coincidence only takes place in networks with degree distributions owning the above-mentioned feature. By the way, it is of interest to distinguish diameter, defined as the maximal shortest distance between any pair of vertices, from the APL we discuss in this paper. According to the argument $d_{ER} = \ln N / \ln \langle k \rangle$ in [19], we may intuitively argue that after a fraction $f$ of all edges are removed $d_{ER} = \ln N / [(1 - f) \ln \langle k \rangle]$ though more simulations need to be performed to validate this statement.

Figure 2 shows the case of an exponential degree distribution. We generate networks with a hidden variable $h \geq 1$, randomly remove edges and calculate the APL. Analytical results are obtained by the numerical method depicted as follows: we numerically calculate the values of (21) in discrete indices of $h$, take them into (14) and numerically compute the results of (14).
Figure 3. APL $l$ as a function of the fraction of removed edges $f$ for asymptotic SF networks with SF exponents of (a) $\alpha = 4$, (b) $\alpha = 3$ and (c) $\alpha = 2.5$, respectively. Formulas (26), (27) and (28) are thus validated by the above three cases, respectively. For instance, in (a) for SF networks with $\alpha = 4$, the scatter full squares stand for the numerical calculations for networks of size $N = 200$. The curve with open squares represents the analytic results of (26) in $N = 200$ and $\alpha = 4$.

Unfortunately, perfect correspondence between theoretical predictions and experimental results does not show again in the SF case as illustrated in figure 3. The reasons for the discrepancy between analytical predictions and experimental results can be listed as follows. Note that Fronczak et al argued that (14) proposed in [11] only fits well for the case of a dense network. However, if we seriously research the study in [11], we will be aware that even for the dense network the analytical results provided by (14) also exhibit deviation from the experimental results, especially for the networks of relatively small size and also for a few large networks. Consequently, the prediction accuracy of the formulas for the APL proposed in this paper is severely affected by that of (14) proposed in [11]. Because the mean degree $\langle k \rangle$ decreases as the fraction of removed edges $f$ rise, the prediction accuracy of the proposed formulas will decay. Thus, we predict only the change of the APL of the networks with only a small portion of edges breaking down, i.e. $0 \leq f \leq 0.1$. Moreover, the proposed formulas are derived based on the continuous forms of $R(h)$ while in computer experiments only discrete forms can be adopted. In order to avoid this problem, the method that we adopt in the exponential networks needs to be applied.

More or less, the APL formula acts better for the BA case, as shown in figure 4, than for the SF case with $\alpha = 3$ although (27) and (32) apparently enjoy the same form. Although the
Figure 4. APL $l$ as a function of the fraction of removed edges $f$ for BA networks with average degree $\langle k \rangle = 2m = 10$, where $m$ is the number of edges introduced in each time step of the construction procedure of the BA model. Scatter data represent results of numerical simulations averaged over $10^3$ realizations, while the curves with open graphics express theoretical results of (32) in $N = 200$, $N = 500$, $N = 1000$ and $N = 2000$, respectively.

reason is still unknown, one may ascribe it to the pure SF behavior of BA in comparison with the asymptotic SF behavior of the SF network generated by the hidden variable model.

In fact, the APL of many other real networks may not be estimated in the same way as we did in this paper. We need first to estimate the analytical forms of degree distribution $P(k)$ of these networks and then do inverse Poisson transform. However, we inevitably cannot avoid experimental noise on $P(k)$ estimated from real networks and this consequently causes fluctuations in $R(h)$. To solve this problem, Earnshaw and Haughey [20] used the combination of cubic splines $\tilde{R}(h) = \sum_i c_i B_i(h)$ to approximate $R(h)$ and then obtain $\tilde{P}(k) = \sum_i c_i A_i(k)$, where $A_i(k)$ is the Poisson transform of $B_i(h)$ and $B_i(h)$ is a cubic B spline function. Then, the problem reduces to obtain the least-squares fitting of $\tilde{P}(k)$ to the observed $P(k)$ of real-world networks.

4. Conclusions

In this paper, we proposed an approximate formula to predict the average path length (APL) of uncorrelated random networks suffering random edge removal. The network generating model on which our work is based is the hidden variable model which generates a network with the expected degree distribution by the so-called hidden variable distribution. Most researchers are only familiar with the degree distribution rather than the hidden variable distribution. We thus first rehearsed the method that helps to obtain the hidden variable distribution from a given degree distribution, namely the inverse Poisson transform. Then, we investigated the new degree distribution after networks suffer random edge removal and obtained the corresponding hidden variable distribution via the above-mentioned inverse Poisson transform method. Combining formula (14) for estimating the APL which only requires the hidden variable distribution, we finally obtained the prediction formula and applied it to ER networks, exponential networks, asymptotic scale-free networks and BA networks respectively in order to validate it. The results we obtained hold well in dense networks and provide a generalized formalism to investigate the robustness of various sorts of networks. In this paper, we focused only on the effect of random breakdown rather than malicious attack.
which is another method to test the topological and the functional (or called the structural and the dynamic) robustness (stability) of networks though distinct attack strategies make different results. Those researchers who concern the study regarding percolation and stability of networks may benefit from our work. Moreover, the conclusion will definitely help to build up a robust network which always holds its original features while confronted with failures or attacks. There are also extensions for further studies. Note that the sum of betweenness of all edges is equivalent to the sum of the lengths of all shortest paths given a shortest path algorithm, where the betweenness of an edge is defined as the number of the shortest paths traversing the edge. Thus, the findings of this paper may be useful to develop a dynamic model for the change of edge betweenness after the breakdown of edges occurs. Inspired by the work about the effect of random removal of nodes on the maximum node betweenness [21], we argue that this methodology, combined with our findings in this paper, should be applied to study the effect of random edge breakdown on the maximum edge betweenness.

Acknowledgments

We would like to thank the anonymous reviewers for their comments and suggestions. This work has been partly supported by China’s 863 program under grant no 2007AA01Z203, by the Foundation of Beijing Jiaotong University under grant no 2006XZ002, by China’s 973 program under grant no 2007CB307101 and by the National Natural Science Foundation of China under grant no 60772039.

Appendix A. Inverse Poisson transform

In this appendix, we briefly rehearse the method for the inverse Poisson transform proposed by Wolf and Mehta [17]. Knowing (3), $P(k)$ is the Poisson transform of $R(h)$. Our purpose, for a given $P(k)$, is to find a uniquely corresponding $R(h)$ satisfying (3). Let us first consider the following formula:

$$F(ix) = \int_{0}^{\infty} e^{ixh} R(h) e^{-h} dh$$

$$= \int_{0}^{\infty} \sum_{k=0}^{\infty} \frac{(ixh)^k}{k!} R(h) e^{-h} dh$$

$$= \sum_{k=0}^{\infty} (ix)^k \int_{0}^{\infty} \frac{h^k}{k!} R(h) e^{-h} dh$$

$$= \sum_{k=0}^{\infty} (ix)^k P(k),$$

(A.1)

where the last step of is derived by applying the definition of $P(k)$ in (3) and $F(ix)$ is the generating function for the degree distribution $P(k)$. It is obvious that $F(ix)$ is the Fourier transform of $R(h)e^{-h}$. Thus, one can further obtain the relation as shown in (8),

$$R(h) = \frac{e^{h}}{2\pi} \int_{-\infty}^{\infty} F(ix) e^{-ixh} dx.$$  

(A.2)

The discrete counterpart of the above formula with continuous form is illustrated in [16] and we will not consider it because in this paper we focus only on the continuous form.

Also notable is that only a few forms of $P(k)$ can do inverse Poisson transform analytically. Other cases are, not unlikely, need to accomplish inverse transform by the
numerical method or approximation. Let us take the case \( P(k) = Ck^{-\alpha} \) with \( k \geq 1 \) (also shown in [16]), for example. We first obtain its generating function which takes the form 
\[
F(ix) = C \sum_{k=1}^{\infty} (ix)^{k-\alpha} = C \operatorname{Li}_\alpha(ix),
\]
where \( \operatorname{Li}_\alpha(ix) \) is the polylogarithm function. However, it is impossible to acquire the inverse Fourier transform of the polylogarithm function and consequently we cannot obtain the corresponding \( R(h) \) by applying (A.2).

For inverting the Poisson transform for the experimentally obtained \( P(k) \), the method proposed in [22] should be applied. For the case where \( P(k) \) is not known exactly or the experimentally obtained \( P(k) \) accompanies with sampling noise, Earnshaw and Haughey [20] proposed a method to approximate the inverse Poisson transform by using cubic splines.

### Appendix B. Statistical characteristics of SF networks

In what follows, we elaborate on how to calculate some statistical characteristics of the SF case. First, recall the conclusion for the SF case (i.e. (24))
\[
R_f(h) = (1 - f)^{-1} R((1 - f)^{-1} h) = C (1 - f)^{a-1} h^{-\alpha},
\]
where \( m \leq h \leq h_{\text{max}} \) and \( h_{\text{max}} \approx m N^{1/\alpha-1} \). By the above definition, we then obtain the following relations:

\[
\langle h \rangle_f = \int_{(1-f)m}^{(1-f)h_{\text{max}}} (1 - f)^{-1} R((1 - f)^{-1} h) \cdot h \, dh = \int_{(1-f)m}^{(1-f)h_{\text{max}}} (1 - f) R((1 - f)^{-1} h) \cdot \frac{h}{1 - f} \, dh = \int_{m}^{h_{\text{max}}} (1 - f) R(h) \cdot h \, dh = (1 - f) \langle h \rangle,
\]
\[
\langle h^2 \rangle_f = \int_{(1-f)m}^{(1-f)h_{\text{max}}} (1 - f)^2 R((1 - f)^{-1} h) \cdot \left( \frac{h}{1 - f} \right)^2 \, dh = (1 - f)^2 \langle h^2 \rangle,
\]
\[
\langle \ln h \rangle_f = \int_{(1-f)m}^{(1-f)h_{\text{max}}} (1 - f)^{-1} R((1 - f)^{-1} h) \cdot \ln h \, dh = \int_{(1-f)m}^{(1-f)h_{\text{max}}} R((1 - f)^{-1} h) \cdot \left( \ln \frac{h}{1 - f} + \ln (1 - f) \right) \, dh = \int_{m}^{h_{\text{max}}} R(h) \cdot (\ln h + \ln (1 - f)) \, dh = \langle \ln h \rangle + \ln (1 - f).
\]

Substituting the above formulas into (14), we obtain the APL formula for SF networks with a fraction \( f \) of all edges removed:
\[
l_f = \frac{-2\langle \ln h \rangle_f + \ln N + \ln \langle h^2 \rangle_f - \gamma}{\ln N + \ln \langle h^2 \rangle_f - \ln \beta_f} + \frac{1}{2}
\]
\[
= \frac{-2[\langle \ln h \rangle + \ln (1 - f)] + \ln N + \ln [(1 - f)^2 \langle h^2 \rangle] - \gamma}{\ln [(1 - f)^2 \langle h^2 \rangle] - \ln [(1 - f) \langle h \rangle]} + \frac{1}{2}
\]
\[
= \frac{-2\langle \ln h \rangle_f + \ln N + \ln \langle h^2 \rangle_f - \gamma}{\ln \langle h^2 \rangle_f - \ln \langle h \rangle_f + \ln (1 - f)} + \frac{1}{2}.
\]
The above conclusions about expectation and second-order moment are not only valid for the SF case, but also for other cases as shown below. After a fraction $f$ of all edges being removed, there exist the following relations [4]:

$$\langle k \rangle_f = \langle k \rangle \cdot (1 - f), \quad (B.6)$$

$$\langle k^2 \rangle_f = \langle k^2 \rangle (1 - f)^2 + \langle k \rangle f (1 - f), \quad (B.7)$$

where $\langle k \rangle$ is the original expectation before random edge failure occurs. Thus, by applying (5), we obtain

$$\langle h \rangle_f = \langle k \rangle_f
= \langle k \rangle (1 - f)
= \langle h \rangle (1 - f), \quad (B.8)$$

$$\langle h^2 \rangle_f = \langle k(k - 1) \rangle_f
= \langle k^2 \rangle_f - \langle k \rangle_f
= \langle k^2 \rangle (1 - f)^2 + \langle k \rangle f (1 - f) - \langle k \rangle (1 - f)
= \langle k^2 \rangle (1 - f)^2 - \langle k \rangle (1 - f)^2
= \langle k(k - 1) \rangle (1 - f)^2
= \langle h^2 \rangle (1 - f)^2. \quad (B.9)$$

References