Probability in the Engineering and Informational Sciences, **15**, 2001, 535–555. Printed in the U.S.A.

PATHS IN THE SIMPLE RANDOM GRAPH AND THE WAXMAN GRAPH

Piet Van Mieghem
Information Technology and Systems
Delft University of Technology
Delft, The Netherlands
E-mail: P.VanMieghem@its.tudelft.nl

The Waxman graphs are frequently chosen in simulations as topologies resembling communications networks. For the Waxman graphs, we present analytic, exact expressions for the link density (average number of links) and the average number of paths between two nodes. These results show the similarity of Waxman graphs to the simpler class $G_p(N)$. The first result enables one to compare simulations performed on the Waxman graph with those on other graphs with same link density. The average number of paths in Waxman graphs can be useful to dimension (or estimate) routing paths in networks. Although higher-order moments of the number of paths in $G_p(N)$ are difficult to compute analytically, the probability distribution of the hopcount of a path between two arbitrary nodes seems well approximated by a Poisson law.

1. INTRODUCTION

The current prominent position of the Internet has fueled network topological studies. Whereas a couple of years ago the design and performance evaluation of ATM switch fabrics spurred queuing analysis, the Internet has shifted the focal domain somewhat more toward the network topology. There are at least two reasons why graph theory seems increasingly useful.

First, the Internet topology itself justifies investigations in its own right. The Internet is a complex system that is growing and changing over time, similar to a living organism. Increasing numbers of studies are being published; some refer-

 0269-9648/01 $12.50

ences can be found in [6,9–11]. Second, many of the properties of network protocol behavior require a graph on which to perform actions. Usually, only simulations are feasible and the graph-theoretical aspects reduce to the choice of a class of easily generated topologies. In many cases, very specific network topologies are considered and, clearly, the conclusions apply (in most cases) only to the considered topology, although often more general statements are possible. Especially in routing algorithm studies, classes of random graphs (r.g.'s), as explained in Section 2, are employed. In general, the set of topologies of a certain class of r.g.'s is complete in the sense that any graph can be represented by that class. This property is interesting for determining the actual or average complexity of routing algorithms (and not only the worst case, as illustrated in [4,8]). Indeed, by simulating a large number of r.g.'s and by analyzing the behavior of interest in every r.g., a probability density function of the behavior of interest is obtained from which any other information (such as the average, variance, maximum, etc.) can be deduced. The completeness of the class of r.g.'s ensures that all possible modes of the routing algorithm are excited, similar to impulse responses in linear system theory.

In practice, simulations can never exhaust the class of r.g.'s with N nodes and, on average, $E[L]$ links because the number of different topologies, on average, equals $\binom{N(N-1)/2}{E[L]}$. Hence, depending on the specific properties of the class of r.g.'s, a particular topology structure features a higher probability of occurrence than in another class of r.g.'s. Deciding which class of r.g.'s for simulations is most suited is often intuitively justified.

In communications network simulations, Waxman graphs are frequently used. Waxman graphs are named after Bernard M. Waxman, who introduced them in [12]. Although the Waxman graphs belong to a broader class of random graphs, as shown in Section 2, the relation and some basic properties of Waxman graphs have, as far as we know, not been previously published. Here, we present analytic results which closely show the similarity in link density (average number of links) and the average number of paths between two nodes with a simpler class of r.g.'s. Although, at first glance, the r.g. classes differ substantially, these results show that average behavior seems not so different. The first result allows one to compare simulations performed on the Waxman graph with those computed on other graphs with the same link density. The average number of paths in Waxman graphs can be useful to dimension (or estimate) routing paths in networks.

In addition to the average, the explicit computation of the complete distribution of the number of paths between two arbitrary nodes in random graphs is shown to be a hard problem. However, the related probability of the hopcount of a path between two arbitrary nodes in $G_p(N)$ seems well approximated by a relatively simple Poisson distribution.

2. RANDOM GRAPHS

There exists an astonishingly large amount of properties of random graphs (r.g.'s) [2]. We refer to the book of Bollobas [2] for an excellent discussion, the recent update of Bollobas's book by Janson et al. [3], and the survey article on recursive trees by Smythe and Mahmoud [5]. The two most frequently occurring models for r.g.'s are $G(N, E)$ and $G_p(N)$. The class $G(N, E)$ constitutes the set of graphs with N nodes and E edges. The class of r.g.'s denoted by $G_p(N)$ consists of all graphs with N nodes in which the edges (or links) are chosen independently and with probability p. A natural refinement of $G_p(N)$ is the model $G_{\{p_{ij}\}}(N)$, where the edges are still chosen independently but where the probability of $i \to j$ being an edge is exactly p_{ij}. The Waxman graph is an example of $G_{\{p_{ij}\}}(N)$. In the class $G_p(N)$, the number of links is not deterministic, but is known, on average, as $pE_{\max}$, where the maximum number of links $E_{\max}$ in a (bidirectional) topology with N nodes is $E_{\max} = N(N-1)/2 \equiv \binom{N}{2}$. This situation is coined a *full mesh* and $G_1(N) = G(N, E_{\max})$ is called the complete graph K_N.

From the point of view of telecommunication networks, by far the most interesting graphs are those with connected topology. This limitation restricts the value of p from below by a critical threshold (i.e., $p > p_c$), where, for large N, $p_c \sim (\ln N)/N$ corresponds to the link density leading to disconnectivity in the r.g.'s. Connectedness of r.g.'s has received considerable attention in the past [2, Chap. 7].

The Waxman graphs are believed to be better representatives of telecommunication networks than r.g.'s of the class $G_p(N)$. The Waxman graph belongs to the family $G_{p_{ij}}(N)$ with $p_{ij} = f(\vec{r}_i - \vec{r}_j)$, where the vector $\vec{r}_i$ represents the position of a node i and all nodes are uniformly distributed in a hypercube of size Z in the m-dimensional space. The dependence on distance is reflected by $f(\vec{r})$, which is a positive, real function of the m coordinates of the vector $\vec{r}$. For example, for the Waxman graph, the distance function is $f(\vec{r}) = e^{-\alpha|\vec{r}|}$, where $|\vec{r}|$ is a norm, denoting a distance from the origin. The idea of relating the probability of a link between nodes i and j to some function of the distance between those nodes stems from the correspondence with realistic telecommunications networks. The farther two nodes lie separated, the less the need for a direct link between them. The example in Figure 1 illustrates the topology of a random graph of $G_p(N)$ and a Waxman graph with the same identifiers p and N.

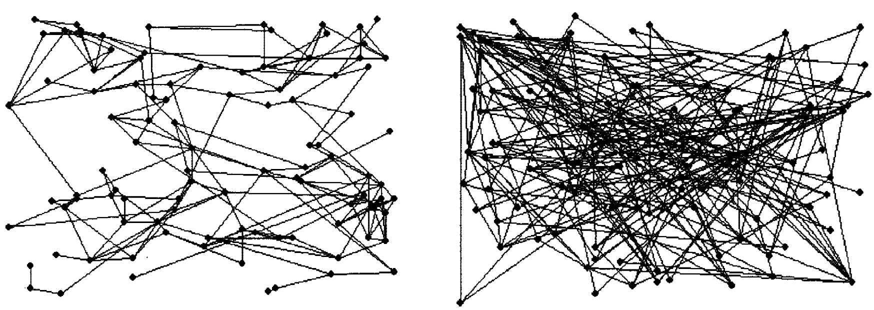

FIGURE 1. Waxman and random graphs. (a) A Waxman graph ($N = 100$, $a = 11$, $p = 0.04$); (b) the graph $G_{0.04}(100)$.

3. THE LINK DENSITY p IN $G_{p_{ij}}(N)$

The number of links $L[\{\vec{r}\}]$ in a particular Waxman graph specified by the nodal positions $\{\vec{r}\} = \{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N\}$ is

$$L[\{\vec{r}\}] = \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(\vec{r}_i - \vec{r}_j) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} e^{-\alpha|\vec{r}_i - \vec{r}_j|}.$$

The average over all possible configurations of nodes in a chosen finite volume V of the m-dimensional space (or the average over all possible Waxman topologies with N nodes and generated by $f(\cdot)$) reads

$$E[L] = \int P[\{\vec{r}\}]L[\{\vec{r}\}]d[\{\vec{r}\}].$$

Since the position of every node in the volume-restricted m-dimensional space is equally likely, the probability distribution is simply uniform or

$$P[\{\vec{r}\}]d[\{\vec{r}\}] = \prod_{k=1}^{N} \frac{d\vec{r}_k}{V}.$$

Hence,

$$\begin{aligned} E[L] &= \prod_{k=1}^{N} \int_V \frac{d\vec{r}_k}{V} \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(\vec{r}_i - \vec{r}_j) \\ &= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int_V \frac{d\vec{r}_i}{V} \int_V \frac{d\vec{r}_j}{V} f(\vec{r}_i - \vec{r}_j) \\ &= E_{\max} \int_V \frac{d\vec{r}}{V} \int_V \frac{d\vec{s}}{V} f(\vec{r} - \vec{s}). \end{aligned} \quad \textbf{(1)}$$

Immediately, the link density for r.g.'s in $G_{p_{ij}}(N)$ follows as

$$p = \frac{E[L]}{E_{\max}} = \int_V \frac{d\vec{r}}{V} \int_V \frac{d\vec{s}}{V} f(\vec{r} - \vec{s}). \quad \textbf{(2)}$$

Unfortunately, in most cases, the integral in the last equation cannot be executed explicitly. However, for the Waxman graph in a square ($m = 2$) with size Z and where $f(\vec{r}) = e^{-\alpha|\vec{r}|}$, explicit computation is possible as illustrated in Appendix A. The decay rate $a = \alpha Z$ of the existence of the link is expressed uniquely in terms of the link density p as, for $a > 0$,

$$\begin{aligned} p(a) = \frac{2}{a^4} [6(1 - 2e^{-a} + e^{-\sqrt{2}a}) + 2a(-4 - 2e^{-a} + 3\sqrt{2}e^{-\sqrt{2}a}) \\ + a^2(4e^{-\sqrt{2}a} + \pi)] + \frac{8g_1(a)}{a} + \frac{8g_2(a)}{a^2} \end{aligned} \quad \textbf{(3)}$$

with, of course, $p(0) = 1$ and where

$$g_1(y) = \frac{dg_2(y)}{dy} = \int_1^{\sqrt{2}} e^{-yx}\sqrt{x^2 - 1}\, dx, \tag{4}$$

$$g_2(y) = \int_1^{\sqrt{2}} e^{-yx}\sqrt{1 - 1/x^2}\, dx. \tag{5}$$

Relation (3) shows that the link density $p(a)$ is only a function of the decay rate a (and not of other parameters as Z). Zegura et al. [13] have considered $p_{ij} = \alpha \exp(-|\vec{r}_i - \vec{r}_j|/\beta L)$ and were led to the same conclusion concerning $p(a)$ via extensive simulations.

In [9], we have shown that when both the $G_p(N)$ and the Waxman graph possess exponentially or uniformly distributed link weights, the probability distribution of the hopcount of the shortest path between two arbitrary nodes is almost identical in both graphs for the same link density p, related via (3), even for a relatively small number of nodes N.

4. THE AVERAGE NUMBER OF PATHS

Paths from a source node, say A, to a destination node, say B, can be categorized according to the number of hops (or the hopcount) of that path, which equals 1 plus the number of different intermediate nodes along the path from A to B. A path with hopcount j is completely characterized by a list of $j + 1$ different nodes, $\mathcal{P}_{A\to B} = [n_1, n_2, \ldots, n_{j+1}]$ with $n_1 = A$, $n_{j+1} = B$, and $n_k \neq n_m$ for all $k, m \in [1, j + 1]$. Sometimes, a more illustrative representation is given, such as $\mathcal{P}_{A\to B} = (n_1 \to n_2)(n_2 \to n_3) \cdots (n_j \to n_{j+1})$. The maximum number of hops is clearly $N - 1$, otherwise a node will appear twice in the path list, indicating that there is a loop.

We first give the general definition of the number of paths with j hops, from which we compute the average number of paths for both the class $G_p(N)$ and $G_{p_{ij}}(N)$ in the next sections.

4.1. The Number of Paths with *j* Hops

Let $X_j(A \to B; N)$ denote the random variable (r.v.) of the number of paths with j hops between a source node A and a destination node B in $G_p(N)$. The most general expression for the number of paths with j hops between node A and node B is

$$X_j(A \to B; N) = \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} 1_{A\to k_1} \cdot 1_{k_1 \to k_2} \cdot \cdots \cdot 1_{k_{j-1}\to B}, \tag{6}$$

where 1_x is the indicator function which equals 1 if the condition x is true, else it is 0. Clearly, the number of paths with one hop equals $X_1(A \to B; N) = 1_{A\to B}$. The maximum number of j hop paths is attained in the complete graph and equals

$$\max(X_j(A \to B; N)) = \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} 1 = \frac{(N-2)!}{(N-j-1)!}. \tag{7}$$

In an earlier paper [7], we demonstrated that, for $N \geq 3$, the total number of paths between two nodes in the complete graph is precisely $[e(N-2)!]$, where $e = 2.71828\ldots$ and $[x]$ denotes the largest integer smaller than or equal to x. Since any graph is a subgraph of the complete graph, this implies that the maximum total number of paths between two nodes in any graph is upper-bounded by $[e(N-2)!]$. In the sequel, we will simplify the notation $X_j(A \to B; N)$ to X_j, because we are not interested in a specific "source A–destination B" pair.

4.2. The Class $G_p(N)$

THEOREM 1: *For the class $G_p(N)$, it holds that*

$$E[X_j] = \frac{(N-2)!}{(N-j-1)!}\, p^j, \qquad 1 \leq j \leq N-1. \tag{8}$$

PROOF: We give two different proofs.

A. The maximum number of different paths with precisely j hops is given by (7). Since each individual path with j hops has probability p^j, we obtain (8) immediately.

B. From (6), we have

$$E[X_j] = E\left[\sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} 1_{A\to k_1} \cdot 1_{k_1 \to k_2} \cdot \cdots \cdot 1_{k_{j-1}\to B}\right]$$

$$= \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} E[1_{A\to k_1} \cdot 1_{k_1 \to k_2} \cdot \cdots \cdot 1_{k_{j-1}\to B}].$$

Since all links $k_m \to k_{m+1}$ for all $0 \leq m \leq j$ are different and independent, each having equal probability p, we have

$$E[1_{A\to k_1} \cdot 1_{k_1\to k_2} \cdot \cdots \cdot 1_{k_{j-1}\to B}] = E[1_{A\to k_1}] \cdot E[1_{k_1\to k_2}] \cdot \cdots \cdot E[1_{k_{j-1}\to B}] = p^j.$$

Thus,

$$E[X_j] = \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} p^j$$

$$= p^j \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} 1 = p^j \max(X_j)$$

from which (8) follows. ■

The average total number of paths between A and B follows from (8), as

$$E\left[\sum_{j=1}^{N-1} X_j\right] = (N-2)!\,p^{N-1}\sum_{l=0}^{N-2}\frac{p^{-l}}{l!}.$$

Since $\sum_{l=0}^{N-2}(p^{-l}/l!) = e^{1/p} - \sum_{l=N-1}^{\infty}(p^{-l}/l!)$, we denote $Q = \sum_{l=N-1}^{\infty}(p^{-l}/l!)$. An upper bound

$$Q < \frac{p^{N-1}}{(N-1)!}\sum_{k=0}^{\infty}\left(\frac{1}{pN}\right)^k = \frac{p^{N-1}}{(N-1)!}\,\frac{pN}{pN-1}$$

for $p > 1/N$ is readily obtained. A close lower bound for Q is derived invoking the beta function [1, Sect. 6.2.1]. Since

$$Q = \left(\frac{1}{p}\right)^{N-1}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{p}\right)^k}{(N-1+k)!}$$

and

$$\frac{1}{(N-1+k)!} = \frac{1}{\Gamma(N+k)} = \frac{B(N,k)}{\Gamma(N)\Gamma(k)},$$

we have

$$\begin{aligned}
Q &= \left(\frac{1}{p}\right)^{N-1}\left(\frac{1}{(N-1)!} + \frac{1}{(N-1)!}\sum_{k=1}^{\infty}\frac{\left(\frac{1}{p}\right)^k}{(k-1)!}\,B(N,k)\right)\\
&= \frac{\left(\frac{1}{p}\right)^{N-1}}{(N-1)!}\left(1 + \frac{1}{p}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{p}\right)^k}{k!}\int_0^1 t^{N-1}(1-t)^k\,dt\right)\\
&= \frac{\left(\frac{1}{p}\right)^{N-1}}{(N-1)!}\left(1 + \frac{1}{p}\int_0^1 t^{N-1}e^{(1-t)/p}\,dt\right) = \frac{\left(\frac{1}{p}\right)^{N-1}}{(N-1)!}\left(1 + \frac{e^{1/p}}{p}\,I\right),
\end{aligned}$$

where

$$I = \int_0^1 t^{N-1}e^{-t/p}\,dt > \frac{e^{-1/p}}{N}.$$

Combining lower and upper bounds yields

$$\frac{\left(\frac{1}{p}\right)^{N-1}}{(N-1)!}\left(1 + \frac{1}{pN}\right) < Q < \frac{\left(\frac{1}{p}\right)^{N-1}}{(N-1)!}\left(1 + \frac{1}{pN-1}\right)$$

and, finally, for $p > 1/N$,

$$(N-2)!\,p^{N-1}e^{1/p} - \frac{pN}{(pN-1)(N-1)} < E\left[\sum_{j=1}^{N-1} X_j\right] < (N-2)!\,p^{N-1}e^{1/p} - \frac{pN+1}{pN(N-1)}. \tag{9}$$

In particular, for the complete graph ($p = 1$), randomness disappears and the total number of paths must be an integer. Since $0 < N/(N-1)(N-1) < 1$, for $N \geq 3$, the above bounds lead to the exact result $[e(N-2)!]$, mentioned earlier.

Using Stirling's [1, formula 6.1.38] approximation $(N-2)! = \sqrt{2\pi}(N-2)^{N-3/2}e^{-(N-2)}e^{\theta/12(N-2)}$ in (9) and $p = 1/(N-2)$ demonstrates that $E[\sum_{j=1}^{N-1} X_j] = O(1/\sqrt{N})$ for large N. Since $1/(N-2) < p_c$, the absence of paths between two arbitrary nodes for large N is expected. The value of $p = p_D$ for which $E[\sum_{j=1}^{N-1} X_j] < 1$ may be regarded as a total disconnectivity threshold and is computed accurately as (26) in Appendix C or, less precise, $p_D = 1/N + O\left(\sqrt{(\log N)/N^3}\right) < p_c$. With this link density p_D, a node in $G_p(N)$ is connected, on average, to only one other node. Just below or around the disconnectivity threshold p_c, a sufficiently large cluster may exist in which communication among a majority of (Internet) users is still possible. Around p_D (only a logarithmic factor in N smaller than p_c), communication is not possible anymore.

4.3. The Class $G_{p_{ij}}(N)$

THEOREM 2: *For the class $G_{p_{ij}}(N)$, it holds that*

$$E[X_j] = \frac{(N-2)!}{(N-j-1)!}\,F_j, \qquad 1 \leq j \leq N-1, \tag{10}$$

where

$$F_j = \int_V \frac{d\vec{r}_1}{V}\int_V \frac{d\vec{r}_2}{V} \cdots \int_V \frac{d\vec{r}_j}{V}\, f(\vec{r}_1 - \vec{r}_2)f(\vec{r}_2 - \vec{r}_3)\cdots f(\vec{r}_j - \vec{r}_{j+1}). \tag{11}$$

PROOF: With (6), where $E[1_{i\to j}] = p_{ij}$, we have

$$\begin{aligned}
E[X_j(\{\vec{r}\})] &= E\left[\sum_{k_1\neq\{A,B\}}\sum_{k_2\neq\{A,k_1,B\}}\cdots\sum_{k_{j-1}\neq\{A,k_1,\ldots,k_{j-2},B\}} 1_{A\to k_i}\cdot 1_{k_1\to k_2}\cdot\ \cdots\ \cdot 1_{k_{j-1}\to B}\right]\\
&= \sum_{k_1\neq\{A,B\}}\sum_{k_2\neq\{A,k_1,B\}}\cdots\sum_{k_{j-1}\neq\{A,k_1,\ldots,k_{j-2},B\}} E[1_{A\to k_1}]\\
&\qquad \cdot E[1_{k_1\to k_2}]\cdot\ \cdots\ \cdot E[1_{k_{j-1}\to B}] \quad \text{(by independence)}\\
&= \sum_{k_1\neq\{A,B\}}\sum_{k_2\neq\{A,k_1,B\}}\cdots\sum_{k_{j-1}\neq\{A,k_1,\ldots,k_{j-2},B\}} f(\vec{r}_A - \vec{r}_{k_1})f(\vec{r}_{k_1} - \vec{r}_{k_2})\\
&\qquad \cdots f(\vec{r}_{k_j} - \vec{r}_B),
\end{aligned}$$

where $\{\vec{r}\}$ refers to a set of N nodal position vectors defining the Waxman graph. Averaging over all possible Waxman topologies yields

$$E[X_j] = \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} \cdots \sum_{k_{j-1} \neq \{A,k_1,\ldots,k_{j-2},B\}} F(\{\vec{r}\}), \tag{12}$$

where

$$F(\{\vec{r}\}) = \prod_{k=1}^{N} \int_V \frac{d\vec{r}_k}{V} f(\vec{r}_A - \vec{r}_{k_1}) f(\vec{r}_{k_1} - \vec{r}_{k_2}) \cdots f(\vec{r}_{k_j} - \vec{r}_B)$$
$$= \int_V \frac{d\vec{r}_A}{V} \int_V \frac{d\vec{r}_{j_1}}{V} \cdots \int_V \frac{d\vec{r}_B}{V} f(\vec{r}_A - \vec{r}_{k_1}) f(\vec{r}_{k_1} - \vec{r}_{k_2}) \cdots f(\vec{r}_{k_j} - \vec{r}_B),$$

where the N-fold m-dimensional integral reduces to a j-fold one because if $l \neq k_i$ with k_i, one of the summation indices in the j-fold summation, $\int_V d\vec{r}_l/V = 1$. This multidimensional integral can be evaluated step by step. Indeed, we can first perform the average over $\vec{r}_A$ that only appears in the first factor of the product of $f(\cdot)$'s. The result is clearly dependent on $\vec{r}_{k_1}$, say $g(\vec{r}_{k_1})$. Next, the average over $\vec{r}_{k_1}$ can be computed; thus, $\int_V (d\vec{r}_{k_1}/V) g(r_{k_1}) f(\vec{r}_{k_1} - \vec{r}_{k_2})$, which only depends on r_{k_2}. Proceeding with this reasoning shows that $F(\{\vec{r}\}) = F_j$ is independent of the summation indices k_i in (12) but dependent on j, the number of m-dimensional integrations. A more direct argument to see this follows from the fact that the positions $\vec{r}_A, \vec{r}_{k_1}, \ldots, \vec{r}_{k_{j-1}}, \vec{r}_B$ are equal in distribution to $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_{j+1}$. Hence, using (7) leads to (10). ■

In the case of the Waxman graph, where $f(\vec{r}) = e^{-\alpha|\vec{r}|}$, we can prove somewhat more. We can write (11) as

$$F_j = E\left[\exp\left(-\alpha \sum_{k=1}^{j} |\vec{r}_{k+1} - \vec{r}_k|\right)\right].$$

The function $x \mapsto e^{-x}$ is a convex function; so by Jensen's inequality, for any random variable X,

$$E[e^{-X}] \geq e^{-E[X]}.$$

Applying this inequality with $X = -\alpha \sum_{k=1}^{j} |\vec{r}_{k+1} - \vec{r}_k|$ with $E[X] = -\alpha \sum_{k=1}^{j} E[|\vec{r}_{k+1} - \vec{r}_k|] = -\alpha j E[|\vec{r} - \vec{s}|]$ yields

$$F_j \geq F^j,$$

where

$$F = e^{-\alpha E[|\vec{r}-\vec{s}|]} = \int_V \frac{d\vec{r}}{V} f(\vec{r} - \vec{s}) \tag{13}$$

and $E[|\vec{r} - \vec{s}|]$ is the average distance between two arbitrary points $\vec{r}$ and $\vec{s}$ in the m-dimensional volume V. The distribution function of the latter, denoted by $g(r)$, is also computed in Appendix A for a square in two dimensions. Hence, for the Waxman graph, it holds that

$$E[X_j] \geq \frac{(N-2)!}{(N-j-1)!} F^j. \tag{14}$$

If F is associated to a link density p, the right-hand side in upper bound in (14) equals the expected number of paths (8) in the simple class $G_p(N)$. When $V \to \infty$, $\int_V d\vec{r} f(\vec{r}-\vec{s}) = \int_V d\vec{r} f(\vec{r})$ and $V^j F_j = (VF)^j$ holds. Since, for the Waxman graph, $\int_V d\vec{r} f(\vec{r})$ is finite for finite dimensions m, and for $\alpha > 0$, $\lim_{V\to\infty} F_j = F^j = 0$. Hence, only in the limit $V \to \infty$, which is equivalent to $p = 0$, the equality sign holds in (14).

5. ON THE GENERATING FUNCTION $\varphi_{X_j}(z) = E[z^{X_j}]$ IN $G_p(N)$

The analytic computation of higher-order moments, $E[X_j^k]$, becomes exceedingly difficult due to the high correlation structure (overlap) of the paths between A and B. Hence, the computation of the probability distribution of the number of paths with j hops between two arbitrary nodes in $G_p(N)$ is, to the best of our knowledge, still an open problem. At least, the variance $\text{var}[X_j]$ seems desirable to estimate, from $G_p(N)$, more closely the number of paths with j hops in a network with N nodes and link density p. In this last section, we motivate the difficulty of the problem, partly by computations, partly via simulations.

The probability generating function (p.g.f.) of the number of paths with j hops is denoted by $\varphi_{X_j}(z) = E[z^{X_j}] = \sum_{k=0}^{\infty} P[X_j = k] z^k$.

In the case $j = 1$, we have that $X_1 = 1_{1\to N}$. Thus, $X_1 = 0$ if there is no link between node 1 and N, which occurs with probability $1-p$, or $X_1 = 1$ if there is a link between node 1 and N, an event that has probability p. The p.g.f. $\varphi_{X_1}(z) = P[X_1 = 0] + P[X_1 = 1]z = (1-p) + pz$.

In the case $j = 2$, initially, all paths with two hops start at 1 and visit an intermediate node i different from 1 and N, from which they depart to the final destination N. Thus, $X_2 = \sum_{i=2}^{N-1} 1_{1\to i} 1_{i\to N}$ and $P[X_2 = k] = \binom{N-2}{k}(p^2)^k(1-p^2)^{N-2-k}$ because $\sum_{i=2}^{N-1} 1_{1\to i} 1_{i\to N}$ can only attain the value k if there are precisely k nonzero terms. The latter event has probability p^2 because both the link $(1 \to i)$ and the link $(i \to N)$ must exist. In all the remaining $N-2-k$ terms, there must be at least one link $(1 \to i)$ or $(i \to N)$ absent, an event with probability $1-p^2$. Finally, the binomial coefficient appears since we can choose these k nonzero terms out of $N-2$ possible precisely in $\binom{N-2}{k}$ ways. So we arrive at

$$\varphi_{X_2}(z) = \sum_{k=0}^{N-2} \binom{N-2}{k} (p^2 z)^k (1-p^2)^{N-2-k} = (1-p^2+p^2 z)^{N-2}.$$

The case $j = 3$ is illustrated in Figure 2. For $j > 2$, we observe that it is possible that paths with j hops overlap partly, which implies that there is a dependence between certain paths. This dependence seriously complicates a probabilis-

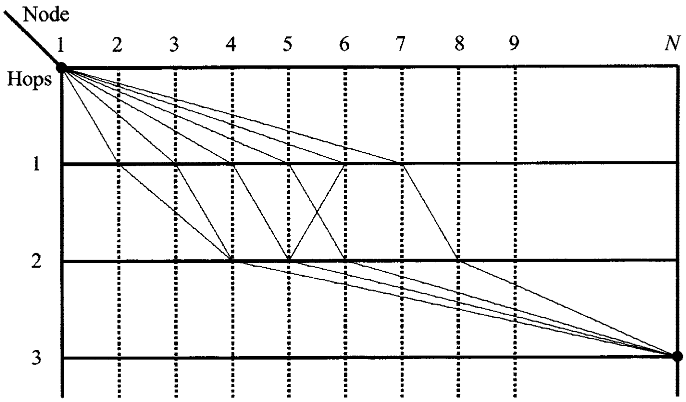

FIGURE 2. A sketch of a counting method of paths with three hops.

tic analysis as in the case with one and two hops. The maximum overlap between paths with j jops consists of $j - 2$ shared links. Hence, each path with j hops has at least two links different from other paths in the set of j-hop paths. This property suggests that the dependence is rather weak for $j = 3$, as also confirmed by simulations plotted in Figure 3. Moreover, for $j = 3$, the variance can be computed. As shown in Appendix B, the result suggests that the case $j = 3$ can be approximated by a Gaussian (for large N).

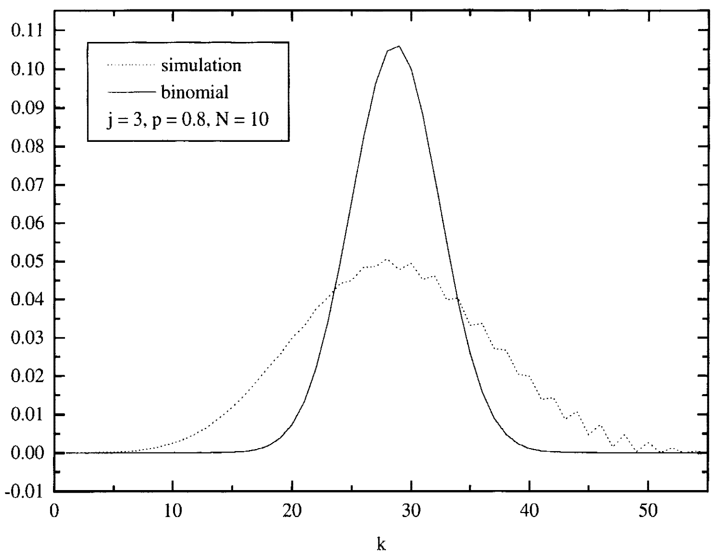

FIGURE 3. For $j = 3$ hops, the width of the Gaussian fit for the simulated results is 16.182 and is 7.5 for the binomial. The centers are 28.42 and 28.76, respectively.

Ignoring the possible overlap or dependence, the r.v. for the number of paths with three hops is $X_3 = \sum_{i\neq(1,N)} \sum_{j\neq(1,i,N)} 1_{1\to i} 1_{i\to j} 1_{j\to N}$ and $P[X_3 = k] = \binom{(N-2)(N-3)}{k}(p^3)^k(1-p^3)^{(N-2)(N-3)-k}$ because, now, the maximum number of three hop paths is $(N-2)(N-3)$, and a nonzero term in the double summation requires that all three factors be unity. This event has probability p^3. Thus, we observe that the double summation of X_3 is separated in "contributing" terms and "noncontributing" terms resulting in $\varphi_{X_3}(z) = (1-p^3+p^3z)^{(N-2)(N-3)}$. Arguments similar to those above, assuming negligible dependence, lead to

$$P[X_j = k] = \binom{\max(X_j)}{k}(p^j)^k(1-p^j)^{\max(X_j)-k} \tag{15}$$

and

$$\varphi_{X_j}(z) = (1-p^j+p^jz)^{(N-2)!/(N-j-1)!}. \tag{16}$$

The simulations below show that formula (15), referred to as "binomial" in the figures, is seriously deficient for $j > 2$ and that the correlation structure in the overlap is very dominant. Also, a Gaussian approximation still adequate in the case $j = 3$, as shown in Appendix B, seems not possible. The p.d.f.'s shown in Figures 5 and 6 indicate that analytic computation (a combinatorial analysis) seems hardly tractable.

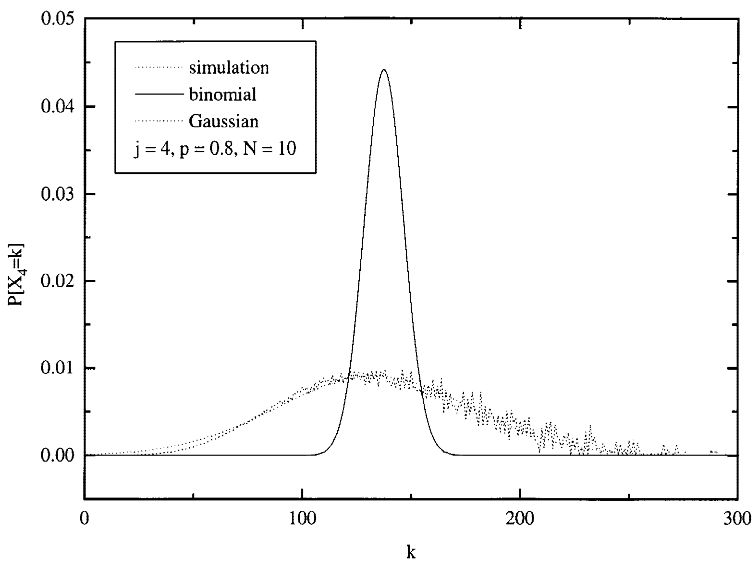

FIGURE 4. For $j = 4$ hops, the width of the Gaussian fit for the simulated results is 88.21 and is 18 for the binomial. The centers are 132.8 and 137.58, respectively.

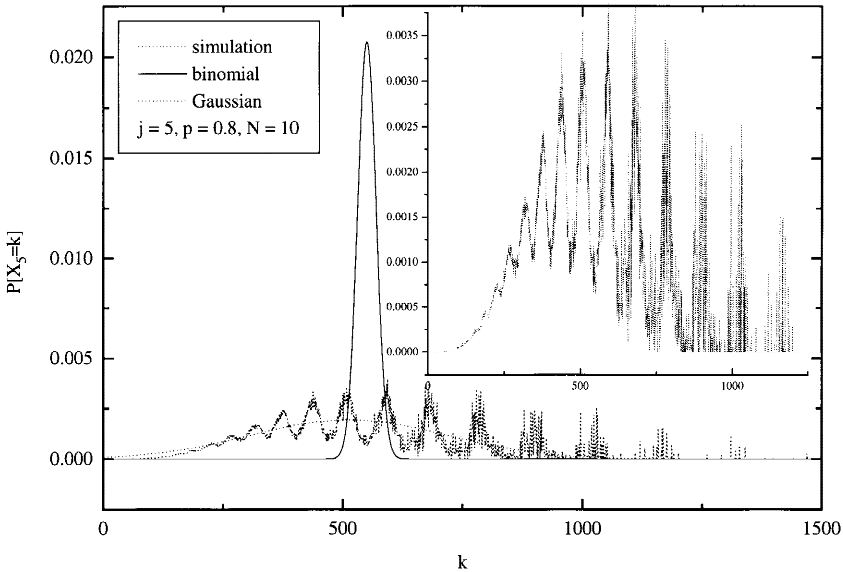

FIGURE 5. For $j = 5$ hops, the width of the Gaussian fit for the simulated results is 410 and is 38.48 for the binomial. The centers are 512.8 and 550.4, respectively.

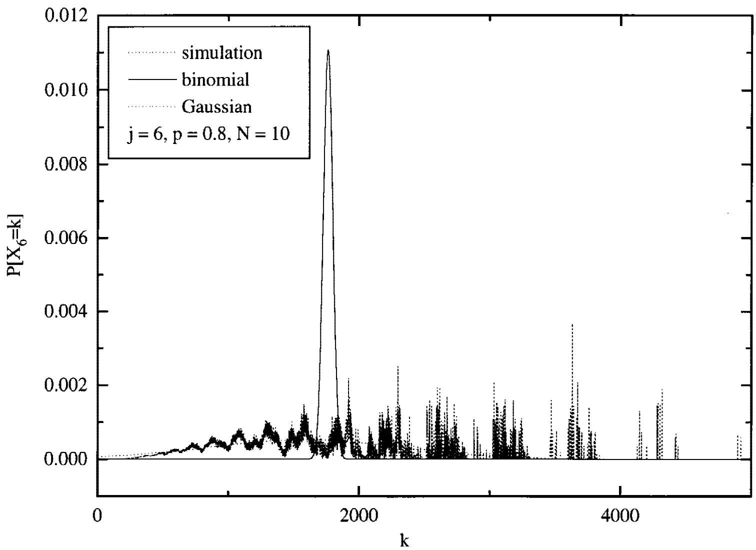

FIGURE 6. For $j = 6$ hops, the width of the Gaussian fit for the simulated results is 1504.7 and is 72.1 for the binomial. The centers are 1584 and 1761, respectively.

In spite of the difficulty in determining the p.d.f. for the number of paths with j hops, simulations[1] indicate that the probability that an arbitrary path between node A and node B consists of j hops, denoted by $P[\text{path} = j \text{ hops}]$, is well approximated by a Poisson distribution. In a particular r.g. Q of $G_p(N)$, the probability of a j-hop path equals $X_j(Q)/\sum_{k-1}^{N-1} X_k(Q)$, where $X_k(Q)$ denotes the number of paths in the r.g. Q with k hops. Hence, averaging over all $Q \in G_p(N)$ yields

$$P[\text{path} = i \text{ hops}] = E_Q\left[\frac{X_j(Q)}{\sum_{k=1}^{N-1} X_k(Q)}\right].$$

Since the r.v. X_j, the number of paths with j hops, is also averaged over all $Q \in G_p(N)$, we can write

$$P[\text{path} = i \text{ hops}] = E\left[\frac{X_j}{\sum_{k=1}^{N-1} X_k}\right]. \tag{17}$$

Unfortunately, (17) is intractable to compute for large N. However, simulations (Fig. 7) show that the p.d.f. of the hopcount in a connected r.g. of $G_p(N)$ is, for $p > p_c$, well approximated by

$$P[\text{path} = i \text{ hops}] \simeq \frac{\dfrac{\left(\frac{1}{p}\right)^{N-i-1}}{(N-i-1)!}}{\sum_{k=0}^{N-2} \dfrac{\left(\frac{1}{p}\right)^{k}}{k!}}. \tag{18}$$

Indeed, assuming that, in (17), $\sum_{k=1}^{N-1} X_k \simeq c$, where c is a constant, the probability that a path between two arbitrary nodes in a random graph of $G_p(N)$ consists of j hops becomes proportional to the expected number of paths with j hops $E[X_j]$. Hence,

$$P[\text{path} = i \text{ hops}] = c^{-1}\,\frac{(N-2)!}{(N-i-1)!}\,p^i,$$

where the proportionality factor c^{-1} follows from the probability normalization condition $\sum_{i=1}^{N-1} P[\text{path} = i \text{ hops}] = 1$. Since $\sum_{k=0}^{N-2}(1/p)^k/k! < e^{1/p}$, we obtain a Poisson lower bound

[1] Only relatively small N can be simulated since the total number of paths grows proportional to $(N-2)!$, as follows from (9).

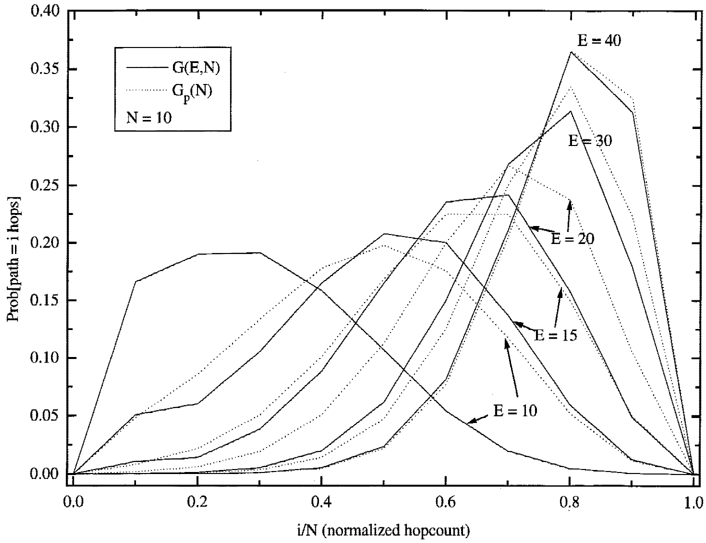

FIGURE 7. A comparison of the p.d.f. of the hopcount for $G(N,E)$ and $G_p(N)$ computed respectively via the simulations and approximated by the Poisson lower bound. The difference between the Poisson lower bound and the simulation results cannot be distinguished on this graph for $E \geq 15$.

$$P[\text{path} = i \text{ hops}] \approx \frac{\left(\frac{1}{p}\right)^{N-i-1} e^{-1/p}}{(N-i-1)!},$$

which agrees remarkably well with simulation results, as shown in Figure 7. Hence, the assumption $\sum_{k=1}^{N-1} X_k \simeq c$ seems a good approximation, which indicates that summing over all possible hops considerably smooths the peculiar correlation structures for the larger hops k.

Acknowledgment

We would like to thank Gerard Hooghiemstra from Delft University of Technology for pointing to the Jensen's inequality.

References

1. Abramowitz, M. & Stegun, I.A. (1985). *Handbook of mathematical functions.* New York: Dover Publications.
2. Bollobas, B. (1985). *Random graphs.* London: Academic Press.
3. Janson, S., Luczak, T., & Rucinski, A. (2000). *Random graphs.* New York: Wiley.
4. Neve, H. & Van Mieghem, P. (2000). TAMCRA: A tunable accuracy multiple constraints routing algorithm. *Computer Communications* 23: 667–679.

5. Smythe, R.T. & Mahmoud, H.M. (1995). A survey of recursive trees. *Theoretical Probability and Mathematical Statistics* 51: 1–27.
6. van der Hofstad, R., Hooghiemstra, G., & Van Mieghem, P. (2001). First passage percolation on the random graph. *Probability in the Engineering and Informational Sciences* 15: 225–237.
7. Van Mieghem, P. (1998). A lower bound for the end-to-end delay in networks: Application to Voice over IP. In *IEEE Globecom 1998*, Vol. 4, pp. 2508-2513.
8. Van Mieghem, P., De Neve, H., & Kuipers, F. (2001). Hop-by-hop quality of service routing. *Computer Networks* (to appear).
9. Van Mieghem, P., Hooghiemstra, G., & van der Hofstad, R. (2000). A scaling law for the hopcount in internet. Report 2000125. Delft University of Technology, Delft.
10. Van Mieghem, P., Hooghiemstra, G., & van der Hofstad, R. (2001). On the efficiency of multicast. *IEEE Transactions on Networking* (to appear).
11. Van Mieghem, P., Hooghiemstra, G., & van der Hofstad, R.W. (2001). Stochastic model for the number of traversed routers in internet. In *Proceedings of Passive and Active Measurement: PAM-2001.*
12. Waxman, B.M. (1998). Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications* 6(9): 1617–1622.
13. Zegura, E.W., Calvert, K.L., & Donahoo, M.J. (1997). A quantitative comparison of graph-based models for internet topology. *IEEE/ACM Transactions on Networking* 5(6): 770–783.

APPENDIX A
Expression for the Link Density in Two-Dimensional Waxman Graphs

The expression for the link density given by (2) is explicitly computed here assuming that all nodes lie within a square with size Z. We use Cartesian coordinates. Thus we have the integral

$$p = \frac{1}{Z^4}\int_0^Z dx_1 \int_0^Z dx_2 \int_0^Z dy_1 \int_0^Z dy_2\, f\left(\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}\right). \tag{19}$$

In the first stage, we use symmetry to reduce the fourfold integral to a double integral. Substitute $u = x_1 - x_2$ with x_2 as constant. Also, define $w = y_1 - y_2$. Then

$$\int_0^Z dx_1 \int_0^Z dx_2\, f(\sqrt{(x_1-x_2)^2+w^2}) = \int_0^Z dx_2 \int_{-x_2}^{Z-x_2} du\, f(\sqrt{u^2+w^2}),$$

and after partial integration we obtain

$$\int_0^Z dx_1 \int_0^Z dx_2\, f\left(\sqrt{(x_1-x_2)^2+w^2}\right) = 2\int_0^Z (Z-u)\, du\, f\left(\sqrt{u^2+w^2}\right).$$

A similar treatment on the y coordinate leads us to

$$p = \frac{4}{Z^4}\int_0^Z du \int_0^Z dw (Z-u)(Z-w) f\left(\sqrt{u^2+w^2}\right). \tag{20}$$

In the second stage, the integral is transformed from Cartesian coordinates to polar coordinates as

$$p = \frac{4}{Z^4}\left[\int_0^Z r\,dr \int_0^{\pi/2} d\phi (Z - r\cos\phi)(Z - r\sin\phi) f(r) + \int_Z^{Z\sqrt{2}} r\,dr \int_{\arccos(Z/r)}^{\pi/2-\arccos(Z/r)} d\phi (Z - r\cos\phi)(Z - r\sin\phi) f(r)\right]. \tag{21}$$

Clearly, the ϕ integral is elementary,

$$\int_a^b (Z - r\cos\phi)(Z - r\sin\phi)\, d\phi = Z^2(b-a) - rZ(\sin\phi - \cos\phi)|_a^b - \frac{r^2}{4}\cos 2\phi\Big|_a^b.$$

Applied to the first integral in (21), this yields

$$\int_0^{\pi/2} (Z - r\cos\phi)(Z - r\sin\phi)\, d\phi = \frac{\pi}{2}Z^2 - 2rZ + \frac{r^2}{2}.$$

Analogously, the second integral becomes

$$\int_{\arccos(Z/r)}^{\pi/2-\arccos(Z/r)} (Z - r\cos\phi)(Z - r\sin\phi)\, d\phi = Z^2\left(\frac{\pi}{2} - 1\right) + 2rZ\sqrt{1 - \frac{Z^2}{r^2}} - 2Z^2\arccos\left(\frac{Z}{r}\right) - \frac{r^2}{2}.$$

Using these results in (21) gives

$$p = \frac{4}{Z^4}\left[\int_0^Z rf(r)\left(\frac{\pi}{2}Z^2 - 2rZ + \frac{r^2}{2}\right) dr + \int_Z^{Z\sqrt{2}} f(r)r\left(Z^2\left(\frac{\pi}{2} - 1\right) + 2rZ\sqrt{1 - \frac{Z^2}{r^2}} - 2Z^2\arccos\left(\frac{Z}{r}\right) - \frac{r^2}{2}\right) dr\right]. \quad \textbf{(22)}$$

This is about as far as we can go without specifying $f(r)$. In passing, we note that (22) immediately gives the probability distribution function $g(r)$ of the distance r between two arbitrary points in the square with size Z. For, the average number of links can also be written as $p = \int_0^{Z\sqrt{2}} g(r)f(r)\, dr$ and from (22), it follows that

$$g(r) = \frac{1}{Z^4}(2\pi r Z^2 - 8r^2 Z + 2r^3) \qquad (0 \le r \le Z)$$
$$= \frac{1}{Z^4}\left(Z^2(2\pi - 4)r + 8Zr\sqrt{r^2 - Z^2} - 8Z^2 r\arccos\left(\frac{Z}{r}\right) - 2r^3\right) \qquad (Z \le r \le \sqrt{2}Z). \quad \textbf{(23)}$$

Here, we choose $f(r) = e^{-\alpha r}$. In this case, (22) can be further simplified. In particular, the first integral is

$$\int_0^Z re^{-\alpha r}\left(\frac{\pi}{2}Z^2 - 2rZ + \frac{r^2}{2}\right) dr = \frac{1}{2\alpha^4}\left[(6 - 8\alpha Z + \alpha^2 Z^2 \pi) - e^{-\alpha Z}(6 - 2\alpha Z + \alpha^2 Z^2(\pi - 5) + \alpha^3 Z^3(\pi - 3))\right].$$

The second integral in (22) is separated into two parts. The first part is

$$\int_Z^{Z\sqrt{2}} re^{-\alpha r}\left(Z^2\left(\frac{\pi}{2}-1\right)-\frac{r^2}{2}\right)dr$$
$$= \frac{e^{-\alpha Z}}{2\alpha^4}\left(-6-6\alpha Z+\alpha^2 Z^2(\pi-5)+\alpha^3 Z^3(\pi-3)\right)$$
$$- \frac{e^{-\sqrt{2}\alpha Z}}{2\alpha^4}\left(-6-6\sqrt{2}\alpha Z+\alpha^2 Z^2(\pi-8)+\sqrt{2}\alpha^3 Z^3(\pi-4)\right).$$

The second part becomes, after a partial integration,

$$\int_Z^{Z\sqrt{2}} re^{-\alpha r}\left(2rZ\sqrt{1-\frac{Z^2}{r^2}}-2Z^2\arccos\left(\frac{Z}{r}\right)\right)dr$$
$$= \frac{e^{-\sqrt{2}\alpha Z}}{2\alpha^4}\left(Z^2(1+\sqrt{2}\alpha Z)(\pi-4)\right)$$
$$+ \frac{2Z}{\alpha}\int_Z^{\sqrt{2}Z} e^{-\alpha r}\left(r+\frac{1}{\alpha}\right)\sqrt{1-\frac{Z^2}{r^2}}\,dr.$$

Unfortunately, the last integral cannot be evaluated analytically. Series expansion is possible but does not lead to attractive results. Therefore, we rewrite that integral into a suitable form for numerical integration as

$$\int_Z^{\sqrt{2}Z} e^{-\alpha r}\left(r+\frac{1}{\alpha}\right)\sqrt{1-\frac{Z^2}{r^2}}\,dr = Z^2 g_1(Z\alpha)+\frac{Z}{\alpha}g_2(Z\alpha),$$

where $g_1(y)$ and $g_2(y)$ are given in (4) and (5), respectively. Putting all pieces together yields the final result (3), where $a = \alpha Z$. The latter demonstrates that the link density is only a function of one parameter, a. At last, we list some numerical values, apart from the trivial $p(0) = 1$ and $p(\infty) = 0$:

$$p(0.2) = 0.902077 \qquad p(0.4) = 0.815725$$
$$p(0.6) = 0.739417 \qquad p(0.8) = 0.671840$$
$$p(1.0) = 0.611868 \qquad p(1.2) = 0.558533$$
$$p(1.4) = 0.511000 \qquad p(1.6) = 0.468548$$
$$p(1.8) = 0.430557 \qquad p(2.0) = 0.396486$$
$$p(2.2) = 0.365868 \qquad p(2.4) = 0.338297$$
$$p(2.6) = 0.313420 \qquad p(2.8) = 0.290930$$
$$p(3.0) = 0.270557 \qquad p(3.2) = 0.252066$$
$$p(3.4) = 0.235251 \qquad p(3.6) = 0.219933$$
$$p(3.8) = 0.205951 \qquad p(4.0) = 0.193166.$$

APPENDIX B
Computation of $E[X_3^2]$

From definition (6), we immediately have that

$$E[X_3^2] = \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_1 \neq \{A,B\}} \sum_{k_2 \neq \{A,k_1,B\}} E[1_{A\to j_1} \cdot 1_{j_1 \to j_2} \cdot 1_{j_2 \to B} \cdot 1_{A \to k_1} \cdot 1_{k_1 \to k_2} \cdot 1_{k_2 \to B}]. \tag{24}$$

The computation of $E[1_{A\to j_1} \cdot 1_{j_1 \to j_2} \cdot 1_{j_2 \to B} \cdot 1_{A \to k_1} \cdot 1_{k_1 \to k_2} \cdot 1_{k_2 \to B}]$ is a combinatorial exercise. The (in)equalities between the set $\{j_1, j_2\}$ and $\{k_1, k_2\}$ must be investigated, with a total of 2^4 possibilities. In general, for j hops, this amount increases as $2^{(j-1)^2}$ and rapidly leads to infeasible analytic treatment. The path restriction amounts to $j_1 \neq j_2$ and, similarly, $k_1 \neq k_2$. Table B1 contains all possibilities where $j \overset{?}{=} k$ is coded by 1 if true and by 0 if not true. Each of the seven nonzero cases contributes to (24) as shown in Table B2. Summing all contributions finally leads to

$$E[X_3^2] = \frac{(N-2)!}{(N-6)!} p^6 + 2\frac{(N-2)!}{(N-5)!}(p^5 + p^6) + \frac{(N-2)!}{(N-4)!}(p^3 + p^5). \tag{25}$$

On the other hand, approximation (16) gives

$$E[X_3^2] \approx \varphi''_{X_3}(1) = \left[\frac{(N-2)!}{(N-4)!} p^3\right]^2 - \frac{(N-2)!}{(N-4)!} p^6.$$

TABLE B1. All Possibilities

No.	$j_1 \overset{?}{=} k_1$	$j_2 \overset{?}{=} k_2$	$j_1 \overset{?}{=} k_2$	$j_2 \overset{?}{=} k_1$	$E[1_{A\to j_1} \cdot 1_{j_1 \to j_2} \cdot 1_{j_2 \to B} \cdot 1_{A \to k_1} \cdot 1_{k_1 \to k_2} \cdot 1_{k_2 \to B}]$
0	0	0	0	0	p^6
1	0	0	0	1	p^6
2	0	0	1	0	p^6
3	0	0	1	1	p^5
4	0	1	0	0	p^5
5	0	1	0	1	0
6	0	1	1	0	0
7	0	1	1	1	0
8	1	0	0	0	p^5
9	1	0	0	1	0
10	1	0	1	0	0
11	1	0	1	1	0
12	1	1	0	0	p^3
13	1	1	0	1	0
14	1	1	1	0	0
15	1	1	1	1	0

TABLE B2. Nonzero Cases

No.	Contribution
0	$p^6 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_1 \neq \{A,j_1,j_2,B\}} \sum_{k_2 \neq \{A,j_1,j_2,k_1,B\}} 1 = [(N-2)!/(N-6)!]\, p^6$
1	$p^6 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_2 \neq \{A,j_1 j_2,B\}} 1 = [(N-2)!/(N-5)!]\, p^6$
2	$p^6 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_1 \neq \{A,j_1 j_2,B\}} 1 = [(N-2)!/(N-5)!]\, p^6$
3	$p^5 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} 1 = [(N-2)!/(N-4)!]\, p^5$
4	$p^5 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_1 \neq \{A,j_1 j_2,B\}} 1 = [(N-2)!/(N-5)!]\, p^5$
8	$p^5 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} \sum_{k_1 \neq \{A,j_1 j_2,B\}} 1 = [(N-2)!/(N-5)!]\, p^5$
12	$p^3 \sum_{j_1 \neq \{A,B\}} \sum_{j_2 \neq \{A,j_1,B\}} 1 = [(N-2)!/(N-4)!]\, p^3$

Hence, for large N and fixed p, we observe that the exact and approximate results yield $E[X_3^2] \sim N^4 p^6 + 2N^3(p^5 + p^6) + O(N^2)$ and $E[X_3^2] \sim N^4 p^6 + O(N^2)$, respectively, hence agreeing to first order in N. With $E[X_3] \sim N^2 p^3$, the exact variance is $\text{var}[X_3] \sim 2N^3(p^5 + p^6) + O(N^2)$ and

$$\frac{X_3 - E[X_3]}{\sqrt{\text{var}[X_3]}} \sim \frac{\sqrt{N}(X_3/N^2 - p^3)}{\sqrt{2(p^5 + p^6)}}\left(1 + O\left(\frac{1}{N}\right)\right),$$

which suggests that the random variable $\sqrt{N}(X_3/N^2 - p^3)$ tends to a Gaussian with mean 0 and variance $2\,(p^5 + p^6)$ for large N.

APPENDIX C
The Total Disconnectivity Threshold p_D

Using the asymptotic formula [1, formula 6.1.41] for $\log \Gamma(N)$ for large N in (9) yields

$$\log\left(E\left[\sum_{j=1}^{N-1} X_j\right]\right) < \log \Gamma(N-1) + (N-1)\log p + \frac{1}{p}$$
$$= (N-1)\log(p(N-1)) - \frac{1}{2}\log\frac{N-1}{2\pi} - (N-1) + \frac{1}{p} + O\left(\frac{1}{N}\right).$$

Let $p = (N-1)^\alpha f(N-1)$ with $f(\cdot) > 0$ and $f(x) = o(x)$ for large x. Set $x = N-1$. Then

$$\log\left(E\left[\sum_{j=1}^{x} X_j\right]\right) < x\log(x^{\alpha+1} f(x)) - \frac{1}{2}\log\frac{x}{2\pi} - x + \frac{x^{-\alpha}}{f(x)} + O\left(\frac{1}{x}\right).$$

Only if $\alpha = -1$ does the right-hand side tend to a finite limit, provided $f(x)$ is suitably chosen. If we choose $g(x) = O(1/x^\beta)$ with $0 < \beta < 1$, the requirement for $f(x)$ to achieve that $\log(E[\sum_{j=1}^{x} X_j]) \to 0$ for large x, is

$$\log(f(x)) + \frac{1}{f(x)} - 1 = \frac{g(x)}{x} + \frac{1}{2x}\log\frac{x}{2\pi}.$$

Since the right-hand side can be made arbitrarily small for large x, a Taylor expansion of the left-hand side around $z = f(x) = 1$ is sufficiently accurate. We have

$$\log(z) + \frac{1}{z} - 1 = \frac{(z-1)^2}{2} - \frac{2}{3}(z-1)^3 + O((z-1)^4).$$

Confining to first order yields

$$f(x) \simeq 1 + 2\sqrt{\frac{g(x)}{x} + \frac{1}{2x}\log\frac{x}{2\pi}}.$$

Hence, we arrive at

$$p_D \sim \frac{1 + \sqrt{\frac{2}{N}\log\frac{N}{2\pi}}\left[1 + O\left(\frac{1}{N^\beta}\right)\right]}{N} < p_c \sim \frac{\log N}{N}. \tag{26}$$