The flooding time in random graphs

Remco van der Hofstad* Gerard Hooghiemstra† Piet Van Mieghem‡

Delft University of Technology
Information Technology and Systems
P.O. Box 5031, 2600 GA Delft, The Netherlands
December 19, 2001

Abstract

Based on our analysis of the hopcount of the shortest path between two arbitrary nodes in the class $G_p(N)$ of random graphs, the corresponding flooding time is investigated. The flooding time $T_N(p)$ is the minimum time needed to reach all other nodes from one node.

We show that, after scaling, the flooding time $T_N(p)$ converges in distribution to the two-fold convolution $\Lambda^{(2*)}$ of the Gumbel distribution function $\Lambda(z) = \exp(-e^{-z})$, when the link density p_N satisfies $\frac{Np_N}{(\log N)^3} \to \infty$ if $N \to \infty$.

Key words and phrases: flooding time, Internet, random graphs.

AMS 1991 *Subject classifications: Primary 60K25, Secondary 60J25.*

1 Introduction

In order to offer high quality multimedia services as telephony, real-time video, file transfer, etc. over an Internet-like future network, all routers in a subnet or an autonomous domain must have a same, consistent view of the network topology. If, at a certain router, significant changes in the available resources occur or a periodic timer triggers for an update of the routing tables, this router may decide to inform all other peers in the subnet about these changes. The most commonly used process that informs each node (router) about changes in the network topology is called *flooding*: the source node initiates the flooding process by sending the packet with topology information to all adjacent neighbors and every router forwards the packet on all interfaces except for the incoming one and duplicate packets are discarded. Flooding is particularly simple and robust since it progresses, in fact, along all possible paths from the emitting node to the receiving node. Hence, a flooded packet reaches a node in the network in the shortest possible time (if overheads in routers are ignored). Therefore, an interesting problem lies in the determination of the flooding time T_N, which is the minimum time needed to inform all nodes in a network with N nodes. Only after a time T_N, all topology databases at each

*email: R.W.vanderHofstad@its.tudelft.nl
†email: G.Hooghiemstra@its.tudelft.nl
‡email: P.VanMieghem@its.tudelft.nl

router in the network are again synchronized, i.e. all routers possess the same topology information. The flooding time T_N is defined as the minimum time needed to reach all $N-1$ remaining nodes from a source node over their respective shortest paths. In a network, each link is specified by a link weight, a positive real number. The shortest path between two given nodes is the path between these nodes, that minimizes the sum of the link weights along that path.

In this paper we focus on the flooding time $T_N(p)$ in the random graph $G_p(N)$, where the link weights are exponentially distributed positive numbers with mean 1. In [5], a first-passage percolation result on the random graph $G_p(N)$ was obtained for the case that the probability $p = p_N$ of an edge(link) being present tends to zero as $N \to \infty$, in such a way that $Np_N \to \infty$. The number of edges(hops) H_N of the shortest path between two arbitrary nodes was shown to satisfy

$$E\left[x^{H_N}\right] = \kappa_N(x)(1+o(1)), \quad N \to \infty, \tag{1}$$

where

$$\kappa_N(x) = \frac{\Gamma(x+N)}{\Gamma(N+1)\Gamma(x+1)}, \tag{2}$$

and $\Gamma(x)$ is the Gamma function.

Although the random graph is not a good model for the *graph* of the Internet, the asymptotic distribution of H_N given by (1) has a remarkable fit with measured Internet data as shown in [8]. It is reasonable to assign random weights to the edges, because operators are suggested by Cisco's OSPF implementation to use link weights that are inverse proportional to the bandwidth of the links. Moreover, we showed in [8] that, within the class of polynomial weights with distribution function (d.f.) $F(x) = x^\alpha$, $x \downarrow 0$; $\alpha > 0$, only uniform weights, $\alpha = 1$, or equivalently exponential weights, gave a proper fit. Finally, measurements of the number of links in a multicast shortest path tree rooted at a source node also justifies the remarkable accuracy of the shortest path tree derived from $G_p(N)$ with exponential weights [9]. Hence, shortest path trees computed from the random graph with exponential weights seem, on empirical grounds, a good model for *shortest path trees* in the Internet.

Our main result is that for the random graph $G_p(N)$, where $p = p_N$, satisfies $Np_N/(\log N)^3 \to \infty$, the distribution of the fooding time $T_N(p)$, properly normalized, converges to the two-fold convolution of the Gumbel distribution (see Theorem 3.1). It is interesting to compare this result with the following different result by Bollobás [3, Theorem 10, p. 233] which holds under identical technical conditions for p_N. If $Np_N/(\log N)^3 \to \infty$, then the limit distribution of the diameter of the random graph $G_p(N)$, denoted by $\text{diam}(G_p(N))$ is concentrated on two consecutive integers $d = d(N) \geq 2$ and $d+1$. More precisely, for each $c > 0$ and with $p_N \in (0,1)$, defined by $p^d N^{d-1} = \log(N/c)$ and if $Np_N/(\log N)^3 \to \infty$, then

$$\lim_{N\to\infty} \Pr[\text{diam}(G_p(N)) = d] = 1 - \lim_{N\to\infty} \Pr[\text{diam}(G_p(N)) = d+1] = e^{-c/2}.$$

The flooding time in $G_p(N)$ with all link weights equal to 1 is bounded from above by the diameter of $G_p(N)$.

The paper is outlined as follows. Exact and asymptotic results for the complete graph, with exponential link weights, are derived in Section 2. Using a probabilistic setting, Section 3 extends the results for the complete graph to all connected graphs in $G_p(N)$, satisfying $Np_N/(\log N)^3 \to \infty$.

2 The complete graph

2.1 Derivation of the flooding transform

We first show that for the complete graph containing N nodes and with independent, exponentially distributed link weights (each with mean 1), the flooding time T_N has Laplace transform

$$\varphi_N(x) = E[e^{-xT_N}] = \int_0^\infty e^{-xt} f_{T_N}(t)\,dt = \prod_{n=1}^{N-1} \frac{n(N-n)}{n(N-n)+x}. \tag{3}$$

Indeed, as demonstrated in [8], the shortest path problem can be exactly described in terms of the Markov chain with transition rates

$$\lambda_{n,n+1} = n(N-n), \qquad\qquad n = 1, 2, \cdots, N-1, \tag{4}$$

where the state n corresponds to the number of discovered nodes. The flooding time T_N equals the absorption time, starting from state $n = 1$ of the birth-process with rates (4). By definition of a continuous time Markov chain, the time $T_N = \sum_{n=1}^{N-1} \tau_n$, where $\tau_1, \tau_2, \cdots, \tau_{N-1}$ are independent, exponentially distributed random variables with parameter $\lambda_{n,n+1}$. Hence, $E[e^{-xT_N}] = E[\exp(-x \sum_{n=1}^{N-1} \tau_n)] = \prod_{n=1}^{N-1} E[e^{-x\tau_n}]$, from which (3) follows.

The average flooding time equals

$$E[T_N] = \sum_{n=1}^{N-1} E[\tau_n] = \sum_{n=1}^{N-1} \frac{1}{n(N-n)} = \frac{2}{N} \sum_{n=1}^{N-1} \frac{1}{n} = \frac{2}{N}(\psi(N) + \gamma), \tag{5}$$

where ψ is the digamma function [1, sec. 6.3] and γ is Euler's constant [1, 6.1.3]. From (5) we conclude that

$$E\left[\frac{NT_N}{2}\right] \sim \log N,$$

which demonstrates that the average flooding time in the complete graph decreases to zero when $N \to \infty$.

The variance of T_N equals

$$\text{var}[T_N] = \sum_{n=1}^{N-1} \text{var}(\tau_n) = \sum_{n=1}^{N-1} \frac{1}{n^2(N-n)^2} = \frac{2}{N^2} \sum_{n=1}^{N-1} \frac{1}{n^2} + \frac{4}{N^3} \sum_{n=1}^{N-1} \frac{1}{n}. \tag{6}$$

For large N, we have that $\text{var}[T_N] = \frac{\pi^2}{3N^2} + O\left(\frac{\log N}{N^3}\right)$.

2.2 The asymptotic law for T_N

The exact expression (23) for the law of T_N (with $N = 2M$), derived in the appendix, does not provide much insight. Because we are interested in the flooding time in *large* networks, we investigate the asymptotic distribution of T_N, for N large. In fact, we present in this article three different methods to compute the asymptotic law for T_N, each with its own merit. First, in the appendix A.3, we compute the limit for $M \to \infty$ of the exact expression (25) after proper scaling and we find with this method the convergence rate of (25) towards the limit (12). Second, in this section, that same limit (12) is found faster and more elegantly. In addition, the scaling law for T_N specified by (10) naturally pops up. Finally, the third way (Theorem 3.1) uses probabilistic arguments and extends the asymptotic law of T_N derived for the complete graph to the class of random graphs $G_p(N)$ that are connected almost surely.

Rewrite (3) as

$$\varphi_N(x) = \prod_{n=1}^{N-1} \frac{n(N-n)}{n(N-n)+x} = \frac{[(N-1)!]^2}{\prod_{n=1}^{N-1}\left[x + \frac{N^2}{4} - \left(n - \frac{N}{2}\right)^2\right]}. \tag{7}$$

For $N = 2M$, using $\frac{\Gamma(z+m)}{\Gamma(z+1)} = \prod_{n=1}^{m-1}(n+z)$, we deduce

$$\varphi_{2M}(x) = \left(\frac{\Gamma(2M)\Gamma(1+\sqrt{x+M^2}-M)}{\Gamma(M+\sqrt{x+M^2})}\right)^2. \tag{8}$$

For large M, there holds $\sqrt{x+M^2} \sim M + \frac{x}{2M}$, provided $|x| < 2M$. After substitution of $x = 2My$ in (8), with $|y| < 1$, we obtain

$$\varphi_{2M}(2My) \sim \Gamma^2(1+y)\frac{\Gamma^2(2M)}{\Gamma^2(2M+y)} \sim \Gamma^2(1+y)(2M)^{-2y},$$

from which the asymptotic relation

$$\lim_{N\to\infty} N^{2y}\varphi_N(Ny) = \Gamma^2(1+y), \qquad |y| < 1, \tag{9}$$

follows. Equivalently, we have for $|y| < 1$,

$$\lim_{N\to\infty} E[e^{-y(NT_N - 2\log N)}] = \lim_{N\to\infty} \frac{1}{N}\int_{-\infty}^{\infty} e^{-yt} f_{T_N}\left(\frac{t+2\log N}{N}\right) dt = \Gamma^2(1+y).$$

This limit demonstrates that the distribution function of $NT_N - 2\log N$ converges to a probability distribution with Laplace transform $\Gamma^2(1+y)$. Moreover, we can prove convergence of *densities*. Let us define the normalized density function

$$g_N(t) = \frac{1}{N} f_{T_N}\left(\frac{t+2\log N}{N}\right). \tag{10}$$

By the inversion theorem for Laplace transforms we obtain for $t \in \mathsf{R}$,

$$\lim_{N\to\infty} g_N(t) = \lim_{N\to\infty} \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{yt} N^{2y}\varphi_N(Ny)dy,$$

where $0 < c < 1$. Since $\Gamma(z)$ is analytic over the entire complex plane except for simple poles at the points $z = -n$ for $n = 0, 1, 2, ...$, we find that $N^{2y}\varphi_N(Ny)$ is analytic whenever the real part of y is nonnegative. Evaluation along the line $\mathrm{Re}(y) = c = 0$ then gives

$$\lim_{N\to\infty} g_N(t) = \lim_{N\to\infty} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{itu} N^{2iu}\varphi_N(iNu)du.$$

As dominating function we take

$$|e^{itu}N^{2iu}\varphi_N(iNu)| = |\varphi_N(iNu)| \le \frac{1+u^2}{u^4},$$

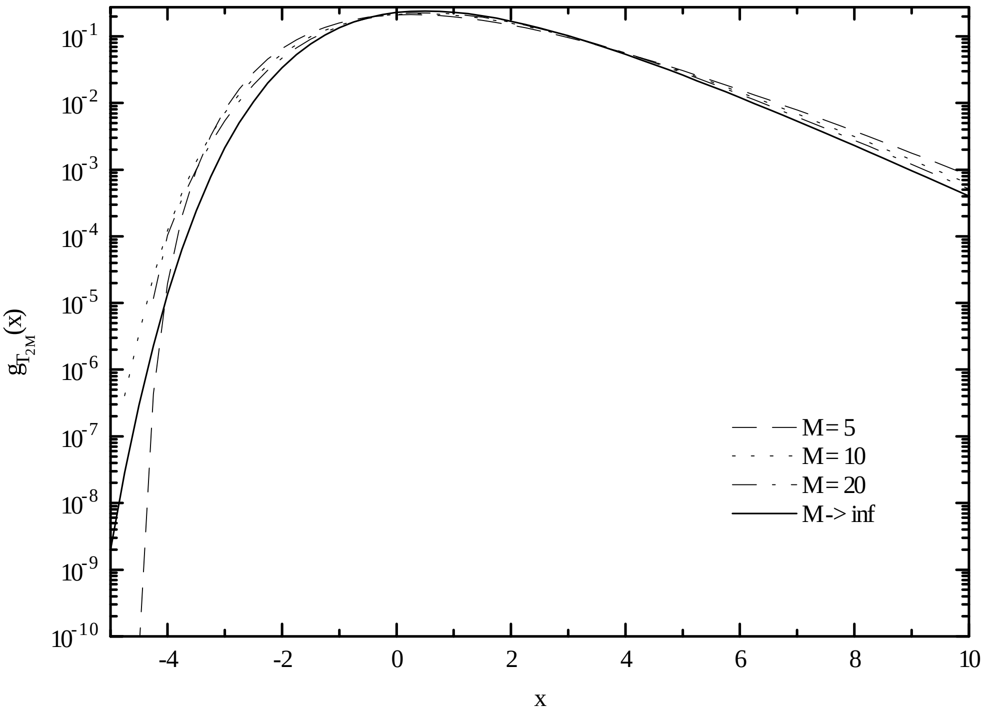

Figure 1: The scaled density $g_N(t)$ for three values of $N = 2M$ (dotted lines) and the asymptotic result (full line).

when $|u| > 1$, and $|\varphi_N(iNu)| \le 1$, for $|u| \le 1$. This follows from the first equality in (7), using only the factors in the product with $n = 1$ and $n = N - 1$, and bounding the other factors using

$$\frac{n(N-n)}{|n(N-n) + iNu|} \le 1.$$

Hence, from Lebesgue's dominated convergence theorem we obtain

$$\begin{aligned} \lim_{N\to\infty} g_N(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itu} \lim_{N\to\infty} N^{2iu} \varphi_N(iNu) du = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{ty} \lim_{N\to\infty} N^{2y} \varphi_N(Ny) dy \\ &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{ty} \Gamma^2(1+y) dy. \end{aligned} \tag{11}$$

The right-hand side of (9) is a perfect square, which indicates that the limit distribution is a two-fold convolution. Indeed, the Mellin transform of the exponential function is

$$e^{-t} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} t^{-y} \Gamma(y)\, dy, \qquad c > 0,$$

and thus with $t = e^{-u}$,

$$\frac{d}{du} \left(e^{-e^{-u}} \right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{yu} \Gamma(y+1)\, dy,$$

which shows that (11) is the two-fold convolution of the probability density $\frac{d}{dt}\Lambda(t)$, where $\Lambda(t) = e^{-e^{-t}}$ is known as the Gumbel distribution. Further, the two-fold convolution is given by

$$\frac{d}{dt} \left(\Lambda^{(2*)}(t) \right) = e^{-t} \int_{-\infty}^{\infty} e^{-e^{-u}} e^{-e^{-(t-u)}} du$$

$$
\begin{aligned}
&= e^{-t}\int_{-\infty}^{\infty} \exp\left[-2e^{-t/2}\cosh\left(\frac{t}{2}-u\right)\right] du \\
&= 2e^{-t}\int_{0}^{\infty} \exp\left[-2e^{-t/2}\cosh(u)\right] du = 2e^{-t}K_0\left(2e^{-t/2}\right),
\end{aligned}
$$

where K_0 denotes the modified Bessel function of order zero [1, 9.6]. Hence,

$$
\lim_{N\to\infty} g_N(t) = g(t) = \frac{d}{dt}\left(\Lambda^{(2*)}(t)\right) = 2e^{-t}K_0\left(2e^{-t/2}\right), \tag{12}
$$

and the corresponding distribution function is

$$
\lim_{N\to\infty} \Pr[NT_N - 2\log N \le z] = 2\int_{-\infty}^{z} e^{-t}K_0(2e^{-t/2})dt = 2e^{-z/2}K_1\left(2e^{-z/2}\right). \tag{13}
$$

Observe that the right-hand side of (12) is maximal for $t = 0.506357$, which is slightly smaller than $\gamma = 0.577261$, but still in accordance with $E[T_N]$ given by (5). The asymmetry shows that $\{NT_N \ge 2\log N + z\}$ is much more likely than the event $\{NT_N \le 2\log N - z\}$, which confirms the intuition that the flooding time can be much longer than the average, but not so much shorter than $E[T_N]$. Figure 1 illustrates the convergence of $g_N(t)$ to the limit in (12). As shown in the Appendix A.3, the rate of convergence for $N \to \infty$ is of order $O\left(\frac{1}{N^{1-\delta}}\right)$ for any fixed x and for arbitrarily small $\delta > 0$.

3 The flooding time in $G_p(N)$

We now proceed with the more interesting case where $p \neq 1$. In $G_p(N)$, for $p < 1$, the number of neighbors of a given node has a binomial distribution with average $(N-1)p$. This leads to double stochasticity: both the value of the link weight (exponentially distributed) and the presence of a link (Bernoulli with parameter p).

Let us first consider the absorption time $A_N(p)$ in the Markov process (birth-process) with birth-rate $\lambda_{n,n+1} = np(N-n)$, where $p = p_N$. For the complete graph where $p = 1$, the absorption time $A_N(p)$ is equal to the flooding time $T_N(p)$. For $p < 1$, this is no longer true because of the double stochasticity. However, as shown below, for $p < 1$, the flooding time $T_N(p)$ can be sandwiched between two absorption times $A_N^{\pm}$ of two associated Markov chains $\{X_N^{\pm}(t)\}$. Therefore, we concentrate first on the absorption time $A_N(p)$ satifying

$$
A_N(p) = \sum_{n=1}^{N-1} \tau_n(p),
$$

where $\tau_n(p)$, $1 \le n \le N-1$, are independent and $\tau_n(p)$ has an exponential distribution with parameter $np(N-n)$. Then,

$$
\varphi_N(x;p) = E[e^{-xA_N(p)}] = \prod_{n=1}^{N-1} \frac{np(N-n)}{np(N-n)+x} \equiv \varphi_N(x/p),
$$

which implies that $A_N(p)$ is distributed as $\frac{T_N}{p}$. The analysis in the preceding section yields

$$
\lim_{N\to\infty} \Pr[Np_N A_N - 2\log N \le z] = \Lambda^{(2*)}(z). \tag{14}
$$

One wonders why a result like (13) or (14) holds for a *sum* of independent random variables, whereas it is well known that the Gumbel distribution Λ is one of the extreme value distributions. The norming constants involved in (13) or (14) are in fact extreme value norming constants. Although T_N or A_N is the sum of independent r.v.'s, it is *not* asymptotically normal because the individual summands $\tau_1, \tau_2, \cdots, \tau_{N-1}$ do not satisfy the Lindeberg condition [4, p. 262]. This phenomenon can be partly understood from the following example, which is close to the results (13) and (14) .

Example: We denote by $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, the order statistics of a sample $X_1, X_2, \ldots, X_n$. Let $Y_1 = X_{(1)}$ and $Y_j = X_{(j)} - X_{(j-1)}$, $2 \leq j \leq n$, denote the spacings of an i.i.d. sequence $X_1, X_2, \ldots, X_n$ of exponentially distributed random variables with parameter 1. Then [4, p. 19], $Y_1, Y_2, \ldots, Y_n$ are independent and $Y_j, 1 \leq j \leq n$, has an exponential distribution with parameter $(n+1-j)$. Moreover

$$\Pr\left[(Y_1 + \ldots + Y_n) - \log n \leq z\right] = \Pr\left[X_{(n)} - \log n \leq z\right] \to \Lambda(z). \qquad ¤$$

A reasoning similar to this example explains, in the next theorem, why the flooding time in the random graph $G_p(N)$ satisfies (14), with A_N replaced by $T_N(p)$.

Theorem 3.1 *If $Np_N/(\log N)^3 \to \infty$ as $N \to \infty$, then the flooding time $T_N(p)$ in $G_p(N)$, with $p = p_N$, satisfies*

$$\lim_{N\to\infty} \Pr\left[Np_N T_N(p_N) - 2\log N \leq z\right] = \Lambda^{(2*)}(z), \tag{15}$$

where $\Lambda^{(2*)}$ denotes the two-fold convolution of the Gumbel d.f. Λ.

PROOF: We may assume that the random graph $G_p(N)$ is connected [3]. The flooding time $T_N(p_N)$ is then the minimum time to reach all $N-1$ nodes from source node 1 over the links with exponential weights. For $A > 0$ define

$$N^{\pm} = \lfloor N(1 \pm A(1-p_N)\log N/(Np_N))\rfloor,$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. In [5], a construction has been given where the discovery process of nodes in $G_p(N)$ starting from the source node 1 was sandwiched between two birth processes $\{X_N^{\pm}(t),\, t \geq 0\}$, with state space $\{1, 2, \ldots, N^{\pm}\}$, initial value $X_N^{\pm}(0) = 1$ and birth rate

$$\lambda^{\pm}_{n,n+1} = n\left\{(N-n)p_N \pm \sqrt{A(N-n)p_N(1-p_N)\log N}\right\}. \tag{16}$$

As a consequence of that construction and the bounds in the proof of Lemma 2.5 of the aforementioned paper the random variable $T_N(p)$ is in between the absorption times $A_N^{\pm}$ of the birth processes $\{X_N^{\pm}(t)\}$, with probability exceeding $1 - 2N^{2-A}$, which tends to 1, if $A > 2$. Hence, if we show that for some $A > 2$ in each of the birth processes $X_N^{\pm}(t), t \geq 0$, the asymptotic relation (15) is satisfied with $T_N(p_N)$ replaced by $A_N^{\pm}$, then we are done. We will give the proof for A_N^{+} (the proof for A_N^{-} is similar).

Observe from the Markov property of $\{X_N^{+}\}$ that

$$A_N^{+} = \tau_1 + \ldots + \tau_{N^+-1},$$

where the τ_n are independent and τ_n has an exponential distribution with parameter nb_n, where

$$b_n = (N-n)p_N + \sqrt{A(N-n)p_N(1-p_N)\log N}.$$

Without restriction take N^+ odd and write

$$A_N^+ = R_N^+ + S_N^+,$$

where

$$\begin{aligned} R_N^+ &= \tau_1 + \ldots + \tau_{(N^+-1)/2}, \\ S_N^+ &= \tau_{(N^++1)/2} + \ldots + \tau_{N^+-1}. \end{aligned}$$

From the independence of τ_n we conclude that R_N^+ and S_N^+ are independent. Hence (15) holds if we show that both

$$\lim_{N\to\infty} \Pr\left[Np_N R_N^+ - \log N\right] \le z = \Lambda(z), \tag{17}$$

$$\lim_{N\to\infty} \Pr\left[Np_N S_N^+ - \log N\right] \le z = \Lambda(z), \tag{18}$$

are satisfied. We start with the proof of (17). Denote by $\tau_n^* = \tau_n - E\left[\tau_n\right] = \tau_n - (nb_n)^{-1}$, the centered random variable, and split for $\delta \in (0,1)$,

$$R_N^+ - E\left[R_N^+\right] = V_N^+ + W_N^+,$$

where

$$V_N^+ = \sum_{n\le N^\delta} \tau_n^*, \qquad W_N^+ = \sum_{n=\lfloor N^\delta\rfloor+1}^{(N^+-1)/2} \tau_n^*.$$

By construction both V_N^+ and W_N^+ have mean zero. Moreover for each $A > 0$,

$$\mathrm{var}(Np_N W_N^+) = N^2 p_N^2 \sum_{n=\lfloor N^\delta\rfloor+1}^{(N^+-1)/2} \frac{1}{n^2 b_n^2} \sim N^2 \sum_{n=\lfloor N^\delta\rfloor+1}^{(N^+-1)/2} \frac{1}{n^2(N-n)^2} \to 0,$$

so that by Chebychev's inequality [4, p. 151], $Np_N W_N^+$ converges to 0 in probability. This implies, because $Np_N E\left[R_N^+\right] = \log N + o(1)$, that for (17) it is sufficient to show that

$$\lim_{N\to\infty} \Pr\left[Np_N V_N^+ \le z\right] = \Lambda(z). \tag{19}$$

Put $M = \lfloor N^\delta \rfloor$, $0 < \delta < 1$, and define

$$Z_n = b_n \cdot \tau_n, \quad 1 \le n \le M.$$

Then Z_n has an exponential distribution with parameter n, and so according to the example, $\sum_{n=1}^M Z_n - \log M$ converges in distribution to a random variable Z, with distribution function Λ. Since

$$Np_N\tau_n = \frac{Np_N Z_n}{b_n} = \frac{Z_n}{1 - \frac{n}{N} + \sqrt{A(1-\frac{n}{N})(1-p_N)\frac{\log N}{Np_N}}},$$

we conclude that $\sum_{n=1}^M Np_N\tau_n - \log M \xrightarrow{d} Z$ if $\sum_{n=1}^M (1 - Np_N/b_n) Z_n$ converges to 0 in probability, which in turn follows if both

$$\sum_{n=1}^M (1 - Np_N/b_n) E\left[Z_n\right] \quad \text{and} \quad \sum_{n=1}^M (1 - Np_N/b_n)^2 \mathrm{var}(Z_n),$$

converge to 0. Since $E[Z_n] = \frac{1}{n}$ and $\mathrm{var}(Z_n) = \frac{1}{n^2}$, these two conditions hold for each $A > 0$ if

$$\sqrt{\frac{\log N}{Np_N}} \sum_{n=1}^{M} \frac{1}{n} \to 0,$$

which is immediate from $\sum_{n=1}^{M} \frac{1}{n} \sim \log M = \delta \log N$ and $(\log N)^3/(Np_N) \to 0$. This proves (19) and hence (17). The proof of (18) is similar to that of (17) and is therefore omitted. ¤

Acknowledgements. We are indebted to a referee for the observation that lead us to include Section A.3. With his addition, this paper now contains three different ways to prove the asymptotic law for T_N.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover, 1968.

[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[3] B. Bollobas, *Random Graphs*, Academic Press, 1985.

[4] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, New York, 2nd edition, 1971.

[5] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, *First-passage percolation on the random graph,* Probability Engineering Informational Science (PEIS), vol. 15, pp. 225-237, 2001.

[6] M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes,* Springer, New York, 1983.

[7] E. C. Titchmarsh, *The Theory of Functions,* Oxford University Press, 1964.

[8] P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad, *A Scaling Law for the Hopcount*, report 2000125 (http://wwwtvs.et.tudelft.nl/people/piet/teleconference.html).

[9] P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad, *On the Efficiency of Multicast*, IEEE/ACM Transactions on Networking, vol. 15, dec. 2001.

A Properties of the Laplace transform φ_N and the corresponding d.f.

A.1 Properties of the Laplace transform φ_N

Let φ_N be the Laplace transform (3).

Lemma A.1 *For all $N > 1$ and $Re(x) \geq 0$,*

$$\varphi_{N+2}(x) = \varphi_N(x + N + 1) \left(\frac{N(N+1)}{x + N + 1} \right)^2 . \tag{20}$$

PROOF: From (7) we obtain

$$\begin{aligned}\varphi_{N+2}(x) &= \frac{[(N+1)!]^2}{\prod_{n=1}^{N+1}\left[x+N+1+\frac{N^2}{4}-\left(n-1-\frac{N}{2}\right)^2\right]}\\ &= \frac{[(N-1)!]^2[N(N+1)]^2}{\prod_{n=0}^{N}\left[x+N+1+\frac{N^2}{4}-\left(n-\frac{N}{2}\right)^2\right]}\\ &= \frac{[(N-1)!]^2[N(N+1)]^2}{(x+N+1)^2\prod_{n=1}^{N-1}\left[x+N+1+\frac{N^2}{4}-\left(n-\frac{N}{2}\right)^2\right]},\end{aligned}$$

from which (20) is immediate. ¤

Relation (20) consists of first a division by x^2 in the x-domain corresponding to a double integration in the t-domain, followed by a shift over $N+1$ which corresponds to a multiplication by $e^{-(N+1)t}$. It follows that in the t-domain (20) is equivalent to

$$f_{T_{N+2}}(t) = (N(N+1))^2 e^{-(N+1)t}\int_0^t (t-u) f_{T_N}(u)du. \tag{21}$$

Together with $f_{T_2}(u) = e^{-t}$ and $f_{T_3}(u) = 4te^{-2t}$, $f_{T_N}(u)$ can be determined for all values of N.

An interesting side result of the recursion is the following. In terms of g_N defined in (10), the recursion (21) can be written as

$$\begin{aligned}g_{N+2}(t) &= \frac{1}{N+2} f_{T_{N+2}}\left(\frac{t+2\log(N+2)}{N+2}\right)\\ &= \frac{N^2(N+1)^2}{N+2} e^{-(N+1)\frac{t+2\log(N+2)}{N+2}}\int_0^{\frac{t+2\log(N+2)}{N+2}}\left(\frac{t+2\log(N+2)}{N+2}-u\right) N g_N(Nu-2\log N)\, du.\end{aligned}$$

It was shown in Section 2, that the density g_N converges pointwise to the density g, given by (12).

If it is permitted to interchange the limit and the integral below, we obtain

$$\begin{aligned}g(t) &= \lim_{N\to\infty} g_{N+2}(t) = e^{-t}\lim_{N\to\infty}\frac{N^2}{(N+2)^2}\int_{-2\log(N+2)}^{t}(t-v) g_N\left(\frac{N(v+2\log(N+2))}{N+2}-2\log N\right)dv\\ &= e^{-t}\int_{-\infty}^{t}(t-v)g(v)dv.\end{aligned}$$

The equation $g(t) = e^{-t}\int_{-\infty}^t (t-v)g(v)dv$ is equivalent to the well-known differential equation for the modified Bessel function K_0 [1, 9.6].

The limit and integral above can be interchanged if we have *moment* convergence of the involved sequence of distribution functions $G_N(t) = \int_{-\infty}^t g_N(v)\, dv$ (the minor shifts of the arguments in the density g_N are unimportant, because both G and g are continuous). Moment convergence follows if the sequence G_N or the associated sequence of random variables $NT_N - 2\log N$ is uniformly integrable (see [2, p. 32]). By the inequality of Cauchy-Schwarz, we have for each $a > 0$,

$$\begin{aligned}\int_{|v|\geq a} v g_N(v)\, dv &= E\left[(NT_N - 2\log N)1_{\{|NT_N-2\log N|\geq a\}}\right]\\ &\leq \left\{E\left[(NT_N-2\log N)^2\right]\Pr[|NT_N-2\log N|\geq a)]\right\}^{1/2}.\end{aligned} \tag{22}$$

where 1_x is the indicator function. The right-hand side of (22) is small uniformly in N, because by (5) and (6),

$$E\left[(NT_N - 2\log N)^2\right] = \text{var}\left[NT_N\right] + \left(E\left[NT_N\right] - 2\log N\right)^2 \to \frac{\pi^2}{3} + 4\gamma^2,$$

and because for a sufficiently large $\Pr\left[|NT_N - 2\log N| > a\right] < \varepsilon$, uniformly in N, since $NT_N - 2\log N$ converges in distribution. This proves that $NT_N - 2\log N$ is uniformly integrable.

A.2 Exact expression for $\Pr[T_{2M} \leq t]$

The transform (3) can be inverted to obtain $f_{T_N}(t)$. We prefer the d.f. $\Pr[T_N \leq t]$ over the density function $f_{T_N}(t)$. After substitution of $\int_0^\infty e^{-xt} f_{T_N}(t)\,dt = x\int_0^\infty e^{-xt}\Pr[T_N \leq t]\,dt$ in (3), and with $N = 2M$, we obtain, for $c > 0$,

$$\Pr[T_{2M} \leq t] = \frac{[(2M-1)!]^2}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{tx}dx}{x(x+M^2)\prod_{n=1}^{M-1}(n(2M-n)+x)^2}.$$

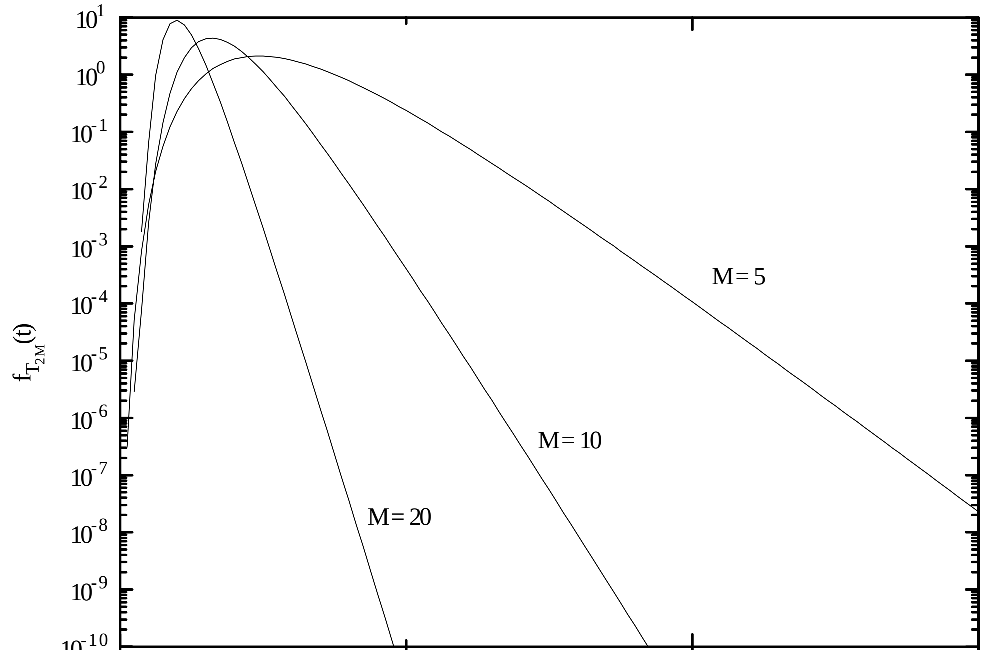

Figure 2: The density $f_{T_{2M}}(t)$ for three sizes of the complete graph. For higher values of M, numerical difficulties start appearing.

We confine further to the even case $N = 2M$. Closing the contour over the negative $\text{Re}(x)$-plane and applying Cauchy's residue theorem yields

$$\Pr[T_{2M} \leq t] = 1 - \frac{e^{-M^2 t}[(2M-1)!]^2}{M^2\prod_{n=1}^{M-1}(M-n)^4} + [(2M-1)!]^2 L,$$

where

$$L = \sum_{n=1}^{M-1} \frac{d}{dx} \left(\frac{e^{tx}(x+n(2M-n))^2}{x(x+M^2)\prod_{j=1}^{M-1}(j(2M-j)+x)^2} \right) \Bigg|_{x=-n(2M-n)}.$$

Let $h_n(x) = [x(x+M^2)\prod_{j=1,j\neq n}^{M-1}(j(2M-j)+x)^2]^{-1}$, then

$$h_n'(x) = -h_n(x) \left(\frac{1}{x} + \frac{1}{x+M^2} + 2\sum_{j=1,j\neq n}^{M-1} \frac{1}{x+j(2M-j)} \right),$$

and

$$\begin{aligned} L &= \sum_{n=1}^{M-1} te^{-n(2M-n)t} h_n(-n(2M-n)) + e^{-n(2M-n)t} h_n'(-n(2M-n)) \\ &= -\sum_{n=1}^{M-1} \frac{e^{-n(2M-n)t}}{n(2M-n)(M-n)^2 \prod_{j=1,j\neq n}^{M-1}(j-n)^2(2M-n-j)^2} \\ &\quad \times \left(t + \frac{1}{n(2M-n)} - \frac{1}{(M-n)^2} - 2\sum_{j=1,j\neq n}^{M-1} \frac{1}{(j-n)(2M-n-j)} \right). \end{aligned}$$

Using

$$\prod_{j=1,j\neq n}^{M-1} (j-n)^2(2M-n-j)^2 = \frac{[(n-1)!(2M-n-1)!]^2}{4(M-n)^4},$$

and

$$\sum_{j=1,j\neq n}^{M-1} \frac{1}{(j-n)(2M-n-j)} = \frac{1}{2(M-n)} \left[\sum_{j=n}^{2M-1-n} \frac{1}{j} - \frac{3}{2(M-n)} \right],$$

we obtain

$$\begin{aligned} L = &-4\sum_{n=1}^{M-1} \frac{(M-n)^2 e^{-n(2M-n)t}}{n(2M-n)[(n-1)!(2M-n-1)!]^2} \\ &\times \left(t + \frac{1}{n(2M-n)} + \frac{1}{2(M-n)^2} - \frac{1}{(M-n)} \sum_{j=n}^{2M-1-n} \frac{1}{j} \right), \end{aligned}$$

and

$$\begin{aligned} \Pr[T_{2M} \le t] &= 1 - \binom{2M-1}{M}^2 e^{-M^2 t} - 4\sum_{n=1}^{M-1} \binom{2M-1}{n}^2 \frac{n(M-n)^2 e^{-n(2M-n)t}}{(2M-n)} \\ &\quad \times \left(t + \frac{1}{n(2M-n)} + \frac{1}{2(M-n)^2} - \frac{1}{(M-n)} \sum_{j=n}^{2M-1-n} \frac{1}{j} \right). \end{aligned} \tag{23}$$

By differentiation we obtain the density,

$$\begin{aligned} f_{T_{2M}}(t) &= M^2 \binom{2M-1}{M}^2 e^{-M^2 t} + 4\sum_{n=1}^{M-1} \binom{2M-1}{n}^2 n^2(M-n)^2 e^{-n(2M-n)t} \\ &\quad \times \left(t + \frac{1}{2(M-n)^2} - \frac{1}{(M-n)} \sum_{j=n}^{2M-1-n} \frac{1}{j} \right). \end{aligned} \tag{24}$$

From (10) and (24), the normalized density follows as,

$$
\begin{aligned}
g_{2M}(t) &= \frac{1}{2M} f_{T_{2M}}\left(\frac{t+2\log(2M)}{2M}\right) \\
&= \frac{M}{2}\frac{\binom{2M-1}{M}^2}{(2M)^M} e^{-Mt/2} + 2\sum_{n=1}^{M-1} \frac{\binom{2M-1}{n}^2}{(2M)^{n(2-\frac{n}{M})}} n^2(1-\frac{n}{M})^2 e^{-nt(1-\frac{n}{2M})} \\
&\quad \times \left(\frac{t}{2} + \log(2M) + \frac{1}{2M(1-n/M)^2} - \frac{1}{(1-n/M)}\sum_{j=n}^{2M-1-n}\frac{1}{j}\right).
\end{aligned} \tag{25}
$$

A.3 The convergence of g_{2M} to the limit (12)

Starting from the explicit form (25), we prove *pointwise* convergence of $g_{2M}(t)$ to its limit $g(t) = 2e^{-t}K_0(2e^{-t/2})$, including an estimate of the rate of convergence. Note that t can be negative, since the support of the random variable $2MT_{2M} - 2\log(2M)$ is equal to $(-2\log(2M), \infty)$, so that for M large enough each negative t is contained in the support.

For large M, using Stirling's asymptotic formula [1, 6.1.38],

$$
M! = \sqrt{2\pi} M^{M+1/2} e^{-M}\left(1 + O\left(\frac{1}{M}\right)\right),
$$

we have

$$
\begin{aligned}
\frac{\binom{2M-1}{M}^2}{(2M)^M} &= \frac{((2M-1)!)^2}{(M!)^2\,((M-1)!)^2\,(2M)^M} \\
&= \frac{2^{3M}}{4\pi M^{M+1}}\left(1 + O\left(\frac{1}{M}\right)\right),
\end{aligned}
$$

so that the first term in (25) tends for $M \to \infty$ to zero, because for fixed t,

$$
\frac{M}{2}\frac{\binom{2M-1}{M}^2}{(2M)^M} e^{-Mt/2} = O\left((Me^{-t/2}/8)^{-M}\right).
$$

Now take an arbitrary δ, with $0 < \delta < \frac{1}{2}$. For $1 \le n \le M^\delta$,

$$
\begin{aligned}
\frac{\binom{2M-1}{n}^2}{(2M)^{n(2-\frac{n}{M})}} e^{-nt(1-\frac{n}{2M})} &= \frac{e^{-nt}\,((2M-1)!)^2\, e^{\frac{n^2}{2M}t}}{(n!)^2\,(2M-1-n)!)^2\,(2M)^{n(2-\frac{n}{M})}} \\
&= \frac{e^{-nt}(2Me^{t/2})^{\frac{n^2}{M}}}{(n!)^2}\prod_{j=1}^{n}\left(1 - \frac{j}{2M}\right)^2 \\
&= \frac{e^{-nt}}{(n!)^2}\exp\left\{\frac{n^2}{M}(\log(2M) + t/2) + \sum_{j=1}^{n} 2\log\left(1 - \frac{j}{2M}\right)\right\} \\
&= \frac{e^{-nt}}{(n!)^2}\left(1 + O\left(\frac{\log M}{M^{1-2\delta}}\right)\right).
\end{aligned}
$$

With the asymptotic expansion of the digamma function $\psi(z)$ [1, 6.3.18], we have for $1 \le n \le M^\delta$,

$$
\log(2M) - \frac{1}{(1-n/M)}\sum_{j=n}^{2M-1-n}\frac{1}{j} = \log(2M) - \frac{1}{(1-n/M)}\sum_{j=1}^{2M-1-n}\frac{1}{j} + \frac{1}{(1-n/M)}\sum_{j=1}^{n-1}\frac{1}{j}
$$

$$
\begin{aligned}
&= \log(2M) - \frac{\psi(2M-n)+\gamma}{(1-n/M)} + \frac{1}{(1-n/M)}\sum_{j=1}^{n-1}\frac{1}{j} \\
&= O\left(\frac{\log M}{M^{1-2\delta}}\right) - \gamma + \sum_{j=1}^{n-1}\frac{1}{j}.
\end{aligned}
$$

Denote by $a_{n,M}(t)$ the summand in (25), i.e.,

$$
g_{2M}(t) = \frac{M}{2}\frac{\binom{2M-1}{M}^2}{(2M)^M}e^{-Mt/2} + 2\sum_{n=1}^{M-1} a_{n,M}(t).
$$

It follows from the above two estimates that for fixed t and $M \to \infty$,

$$
\left| 2\sum_{n=1}^{\lfloor M^\delta \rfloor}\left[a_{n,M}(t) - \frac{e^{-nt}}{((n-1)!)^2}\left(\frac{t}{2} - \gamma + \sum_{j=1}^{n-1}\frac{1}{j}\right)\right]\right| = O\left(\frac{\log M}{M^{1-2\delta}}\right). \tag{26}
$$

On the other hand we have for $M^\delta < n \le M-1$,

$$
|a_{n,M}(t)| \le \frac{\binom{2M-1}{n}^2}{(2M)^{n(2-\frac{n}{M})}} n^2 e^{n|t|}\left(\frac{|t|}{2} + \log(2M) + \frac{1}{2M}\right),
$$

because

$$
\left|(1-n/M)^2\left(\log(2M) + \frac{1}{2M(1-n/M)^2} - \frac{1}{(1-n/M)}\sum_{j=n}^{2M-1-n}\frac{1}{j}\right)\right| \le \log(2M) + \frac{1}{2M}.
$$

Furthermore, we have the identity

$$
\frac{\binom{2M-1}{n}^2}{(2M)^{n(2-n/M)}} = \prod_{k=1}^{n}\left\{\left(\frac{1}{k} - \frac{1}{2M}\right)^2 (2M)^{\frac{2k-1}{M}}\right\}, \tag{27}
$$

which follows easily from $\binom{2M-1}{n} = \binom{2M-1}{n-1}\cdot\frac{2M-n}{n}$. In order to prove that $\sum_{n>M^\delta}|a_{n,M}(t)|$ is small, we proceed with an analysis of the individual term in the product (27).

The function $h : z \mapsto (\frac{1}{z} - \frac{1}{2M})^2(2M)^{\frac{2z}{M}}$, $1 \le z \le M$, is decreasing for $z < z_M$, where

$$
z_M = M - M\sqrt{1 - \frac{2}{\log(2M)}} \sim \frac{M}{\log(2M)},
$$

and increasing for $z > z_M$. Further,

$$
h(M^\delta) = (2M^{1-\delta} - 1)^2(2M)^{-2+2M^{\delta-1}} = O(M^{-2\delta}).
$$

For each M and $2 \le k \le M-1$, each of the factors in the product (27) is smaller than or equal to 1, while for $k = 1$ the factor converges to 1. Hence for $n > M^\delta$,

$$
\prod_{k=1}^{n}\left\{\left(\frac{1}{k} - \frac{1}{2M}\right)^2(2M)^{\frac{2k-1}{M}}\right\} \le \prod_{k=M^{\delta/2}}^{n}\left\{\left(\frac{1}{k} - \frac{1}{2M}\right)^2(2M)^{\frac{2k}{M}}\right\} \le \left(h(M^{\delta/2})\right)^{n-M^{\delta/2}},
$$

so that

$$M\cdot\sum_{n=\lfloor M^\delta\rfloor}^{M-1}|a_{n,M}(t)|\le M\cdot\sum_{n=\lfloor M^\delta\rfloor}^{M-1}\left(h(M^{\delta/2})\right)^{n-M^{\delta/2}}n^2e^{n|t|}\left(\frac{|t|}{2}+\log(2M)+\frac{1}{2M}\right),$$

which is small, for M large enough.

To conclude the proof we use the series expansion [1, 9.6.13],

$$K_0(2e^{-t/2})=\sum_{n=0}^{\infty}\frac{e^{-nt}}{(n!)^2}\left(\frac{t}{2}-\gamma+\sum_{j=1}^{n}\frac{1}{j}\right).$$

Since

$$M\cdot\sum_{n>M^\delta}\frac{e^{-nt}}{(n!)^2}\left(\frac{t}{2}-\gamma+\sum_{j=1}^{n}\frac{1}{j}\right)\to 0,$$

the error bounds above for $M\to\infty$ and each fixed t show that

$$\begin{aligned}
g_{2M}(t) &= 2\sum_{n=1}^{M-1}a_{n,M}(t)+O\left((Me^{-t/2}/8)^{-M}\right)\\
&= 2\sum_{n=1}^{\lfloor M^\delta\rfloor}a_{n,M}(t)+O\left(M^{-1}\right)\\
&= 2\sum_{n=1}^{\lfloor M^\delta\rfloor}\frac{e^{-nt}}{((n-1)!)^2}\left(\frac{t}{2}-\gamma+\sum_{j=1}^{n}\frac{1}{j}\right)+O\left(\frac{\log M}{M^{1-2\delta}}\right)\\
&= 2e^{-t}K_0(2e^{-t/2})+O\left(\frac{\log M}{M^{1-2\delta}}\right).
\end{aligned}$$

The above error estimate is true for each positive δ. Because $(\log M)/M^\delta\to 0$ for each positive δ, we may conclude that the rate of convergence is of order $M^{-1+\varepsilon}$, with ε arbitrary small but positive.