Limit Theorems for the Average Distance and the Degree Distribution of the Threshold Network Model

Akihiro FUJIHARA¹, Yusuke IDE²*, Norio KONNO², Naoki MASUDA³,⁴, Hiroyoshi MIWA¹ and Masato UCHIDA⁵

¹Graduate School of Science and Technology, Kwansei Gakuin University, Hyogo 669-1337, Japan
²Department of Applied Mathematics, Yokohama National University, Yokohama 240-8501, Japan
³Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan
⁴PRESTO, Japan Science and Technology Agency, Saitama 332-0012, Japan
⁵Network Design Research Center, Kyushu Institute of Technology, Tokyo 100-0011, Japan

Received February 27, 2009; final version accepted April 23, 2009

The threshold network model is a type of finite random graph. For this model, we obtain an almost-sure convergence theorem for the average distance and derive a variant of Poisson approximation of the degree distribution by using extreme value theory.

KEYWORDS: complex networks, average distance, Poisson approximation

1. Introduction

Many real-world graphs (networks) are complex and often characterized by small diameters, high clustering, and power-law (scale-free) degree distributions [1, 3, 16]. We analyze the threshold network model defined as follows. Consider \( n \) vertices labeled \( 1, 2, \ldots, n \) and independent and identically distributed (i.i.d.) real-valued random variables \( X_1, \ldots, X_n \) with a common distribution function \( F \). We connect vertices \( i \) and \( j \) with \( i \neq j \) by an edge when \( X_i + X_j > \theta \) for a given threshold \( \theta \in \mathbb{R} \). This model is a subclass of so called hidden variable models [6, 18], and their mean behavior [4, 6, 7, 9, 13, 17, 18] and limit theorems [10–12] for the degree, the clustering coefficients, and the number of subgraphs have been analyzed. See [10–12, 14, 15] for generalizations of the model.

The rest of the paper is organized as follows. In Sect. 2, we state Theorems 1 and 2 which are main results of this paper. We prove Theorem 1 in Sect. 3 and Theorem 2 in Sect. 4.

2. Results

2.1 Average distance

Limit theorems for the degree and the clustering coefficients were obtained in previous literature [10–12]. In this paper, we prove an almost-sure convergence theorem for the average distance defined by

\[
L_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} d(i, j),
\]

where \( d(i, j) \) is the graph distance between vertices \( i \) and \( j \). It is easy to show that \( d(i, j) = 1, 2, \) or \( \infty \) for the threshold network model (see Sect. 3). If \( d(i, j) = \infty \) for some \( i \) and \( j \), then \( L_n = \infty \), irrespective of the number of vertex pairs that realize \( d(i, j) = 1 \) or \( 2 \). Thus we use an indeterminate \( w \) corresponding to a pair of vertices with distance \( \infty \). Let \( D_n \) be the number of edges and \( N_n^{(\infty)} \) be the number of vertex pairs with distance \( \infty \). We obtain

\[
L_n = \frac{1}{\binom{n}{2}} \left[ 1 \times D_n + 2 \times \binom{n}{2} - D_n - N_n^{(\infty)} \right] + w \times N_n^{(\infty)}
\]

\[
= 2 - D_n - (2 - w) \frac{N_n^{(\infty)}}{\binom{n}{2}}.
\] (1)

Using this fact, we obtain
Theorem 1. Let \( x^* = \sup \{ x \in \mathbb{R} : F(x) < 1 \} \). Then we have
\[
\lim_{n \to \infty} L_n = 2 - \Pr(X_1 + X_2 > \theta) - (2 - u)F(\theta - x^*)[2 - F(\theta - x^*)], \quad a.s.
\]

2.2 Degree distribution

An almost-sure convergence theorem for the degree of the model has been proved [10–12]. In this paper, we derive a weak convergence theorem for the degree distribution of a slightly modified model. When \( X_1 = x \), the probability that vertex 1 is connected to another vertex \( j \), i.e.,
\[
\Pr(X_2 > \theta - x) = 1 - F(\theta - x),
\]
does not depend on \( n \). In order to make this probability \( \theta(1/n) \) to avoid the average degree to explode as \( n \to \infty \), we take a sequence of thresholds \( \{ \theta_n \} \) such that
\[
\lim_{n \to \infty} (n - 1)\Pr(X_2 > \theta_n - x) = c(x),
\]
where \( c(x)(\in (0, \infty)) \) depends only on \( x \). This is an analog of the Poisson approximation of the degree distribution of the Erdős-Rényi random graph [5]. However, we can not necessarily take such a sequence of thresholds.

Let \( D_n(1; x) = \sum_{x_i > x} I_{(\theta_n, \infty)}(x + X_i) \) be the degree of vertex 1 with \( X_1 = x \), where \( I_A(\cdot) \) denotes the indicator function of set \( A \subseteq \mathbb{R} \), i.e., \( I_A(y) = 1 \) for \( y \in A \) and \( I_A(y) = 0 \) otherwise, and \( D_n(1) = \sum_{i \geq 1} D_n(1; X_i) \) be the random variable corresponding to the degree of vertex 1. We obtain the following sufficient condition:

Theorem 2. Assume that there exist a sequence \( \{ a_n \} \) of positive numbers and a sequence of thresholds \( \{ \theta_n \} \) such that
\[
\lim_{n \to \infty} F(a_n x + \theta_n)^2 = \exp(-e^{-x}), \quad \text{for any } x \in \mathbb{R}, \text{ and } \lim_{n \to \infty} a_n = 1/A \in (0, \infty).
\]

1. If \( x^* < \infty \) and \( x_* > 0 \), then \( D_n(1) \) does not converge weakly.
2. If \( x^* < \infty \) and \( x_* \leq 0 \), then \( D_n(1) \) converges weakly to \( D(1) \) such that
\[
\Pr(D(1) = 0) = F_-(0) + \int_0^{x^*} \exp(-e^{-x})F(dx),
\]
\[
\Pr(D(1) = k) = \int_0^{x^*} \exp(-e^{-x}) \frac{(e^{-x})^k}{k!} F(dx), \quad (k = 1, 2, \ldots),
\]
where \( F_-(0) = \lim_{x \to 0^+} F(x) \), \( x^* = \sup \{ x \in \mathbb{R} : F(x) < 1 \} \), and \( x_* = \inf \{ x \in \mathbb{R} : F(x) > 0 \} \).
3. If \( x^* = \infty \), then \( D_n(1) \) converges weakly to \( D(1) \equiv \text{Poisson}(e^{x^*}) \), i.e.,
\[
\Pr(D(1) = k) = \int_{x_*}^{\infty} \exp(-e^{-x}) \frac{(e^{-x})^k}{k!} F(dx), \quad (k = 0, 1, \ldots).
\]

Corollary 1. Let \( X_1 \) be absolutely continuous such that it has a probability density function \( f \). Then \( f \) satisfying the assumption of Theorem 2 must have the following form:
\[
\lim_{x \to \infty} \frac{f(x + y)}{f(x)} = e^{-xy}, \quad (2)
\]
for any \( y \in \mathbb{R} \). In this case, \( D_n(1) \) converges weakly to \( D(1) \equiv \text{Poisson}(e^{x^*}) \).

An example of \( f \) that satisfies Eq. (2) is the exponential distribution. In this case, we obtain \( \lim_{k \to \infty} \Pr(D(1) = k)/k^{-2} = 1 \). This result is consistent with a previously derived approximate result [7].

3. Proof of Theorem 1

In this section, we prove Theorem 1. Let \( x_i \) \( (1 \leq i \leq n) \) be a realization of \( X_i \) and \( x_{i(1)} \leq x_{i(2)} \leq \cdots \leq x_{i(k)} (2 \leq k \leq n) \) be a reordered subsequence of the realizations. If \( x_{i(1)} + x_{i(k)} > \theta \), the vertex corresponding to \( x_{i(k)} \) is connected to the other \( k - 1 \) vertices. If \( x_{i(1)} + x_{i(k)} \leq \theta \), then the vertex corresponding to \( x_{i(1)} \) is isolated. Therefore, there exists a vertex connecting all other vertices in the subgraph comprising all nonisolated vertices. Thus \( d(i, j) = 1, 2, \infty \) for any \( 1 \leq i < j \leq n \).

Let \( N_n^{(0)} \) be the number of isolated vertices. Using Eq. (1) and
\[
N_n^{(0)} = \binom{N_n^{(0)}}{2} + N_n^{(0)}(n - N_n^{(0)}) = \frac{N_n^{(0)}(2n - N_n^{(0)} - 1)}{2},
\]
we obtain
Thus the following limit theorem is known:

**Fact 1 (Theorem 1 of [10], Fact 1 of [11]).**

\[
\lim_{n \to \infty} \mathbb{P}\left( X_1 + X_2 > \theta \right) = a.s.
\]

So it is sufficient to know the asymptotic behavior of \( N_n^0(0) \). Let \( N_n^0(i; x) \) be the indicator function for a vertex \( i \in \{1, \ldots, n\} \) with \( X_i = x \), so that \( N_n^0(i; x) = 1 \) if vertex \( i \) is isolated and \( N_n^0(i; x) = 0 \) if vertex \( i \) is not isolated, i.e.,

\[
N_n^0(i; x) = \prod_{1 \leq j \leq n, j \neq i} \{1 - I_{(\theta, \infty)}(x + X_j)\}.
\]

Note that \( N_n^0 = \sum_{i=1}^n N_n^0(i) \), where \( N_n^0(i) \equiv N_n^0(i; X_i) \). Then we have the following lemma:

**Lemma 1.** Let \( x^* = \sup\{x \in \mathbb{R} : F(x) < 1\} \).

1. For any \( x \in \mathbb{R} \),

\[
\lim_{n \to \infty} N_n^0(1; x) = N^0(1; x) = 1 - I_{(-\infty, x^*]}(x), \quad a.s.
\]

2. In particular,

\[
\lim_{n \to \infty} N_n^0(1) = N^0(1) = 1 - I_{(-\infty, x^*]}(X_1), \quad a.s.
\]

**Proof.** Since the sequence of events \( \{N_n^0(1; x) = 1\} \) is nondecreasing in \( n \), we obtain

\[
\mathbb{P}\left( \lim_{n \to \infty} N_n^0(1; x) = 1 \right) = \lim_{n \to \infty} \mathbb{P}(N_n^0(1; x) = 1) = \lim_{n \to \infty} \mathbb{P}\left( \bigcap_{2 \leq j \leq n} \{x + X_j \leq \theta\} \right)
= \lim_{n \to \infty} F(\theta - x)^{n-1}.
\]

Thus we have

\[
\mathbb{P}\left( \lim_{n \to \infty} N_n^0(1; x) = 1 \right) = \begin{cases} 1, & \text{if } F(\theta - x) = 1, \\ 0, & \text{if } F(\theta - x) < 1. \end{cases}
\]

A similar argument gives

\[
\mathbb{P}\left( \lim_{n \to \infty} N_n^0(1; x) = 0 \right) = \begin{cases} 1, & \text{if } F(\theta - x) < 1, \\ 0, & \text{if } F(\theta - x) = 1. \end{cases}
\]

Thus

\[
\lim_{n \to \infty} N_n^0(1; x) = \begin{cases} 1, & \text{if } F(\theta - x) = 1, \\ 0, & \text{if } F(\theta - x) < 1, \end{cases}
\]

with probability 1. Combining Eq. (6) and

\[
[x \in \mathbb{R} : F(\theta - x) = 1] = \{x \in \mathbb{R} : x^* \leq \theta - x\} = (-\infty, \theta - x^*]
\]

leads to Eq. (4). Using Fubini’s theorem and Eq. (4), we have Eq. (5) as follows:

\[
\mathbb{P}\left( \lim_{n \to \infty} N_n^0(1; X_1) = N^0(1; X_1) \right) = \int_{\mathbb{R}} \mathbb{P}\left( \lim_{n \to \infty} N_n^0(1; x) = N^0(1; x) \mid X_1 = x \right) F(dx)
= \int_{\mathbb{R}} 1 \cdot F(dx) = 1.
\]

We obtain the following proposition by combining Lemma 1 and the argument used for proving Theorem 3 of [11]:

\[
L_n = 2 - \frac{D_n}{\langle a \rangle} - (2 - w) \frac{N_n^0(2n - N_n^0) - 1}{n(n - 1)}.
\]
Proposition 1.

\[
\lim_{n \to \infty} \frac{N^{(0)}_n}{n} = \mathbb{E}[N^{(0)}(1; X_1)] = F(\theta - x^*), \quad \text{a.s.}
\]

We obtain Theorem 1 by applying Fact 1 and Proposition 1 to Eq. (3).

4. Proofs of Theorem 2 and Corollary 1

In this section, we prove Theorem 2 and Corollary 1 by using extreme value theory. The theory is concerned with limit behavior of the maximum \(M_n = \max\{X_1, \ldots, X_n\}\) of i.i.d. random variables \(X_i\). First, we review the theory briefly [8]. Suppose there exist a sequence of positive numbers \(\{a_n\}\) and a sequence of real numbers \(\{b_n\}\) such that \((M_n - b_n)/a_n\) has a nondegenerate limit distribution \(G\) called extreme value distribution, i.e., \(\lim_{n \to \infty} F(a_n x + b_n)^n = G(x)\) for every continuity point \(x\) of \(G\).

Fact 2 (Theorem 1.1.3 of [8]). The class of extreme value distributions is \(G\) with \(a > 0\) and \(b \in \mathbb{R}\), where

\[
G(x) = \exp\left(-\left(1 + \gamma x\right)^{-1/\gamma}\right), \quad 1 + \gamma x > 0,
\]

with \(\gamma \in \mathbb{R}\). Here if \(\gamma = 0\) then the right-hand side is interpreted as \(\exp(-e^{-x})\).

These distributions corresponding to \(\gamma = 0\), \(\gamma > 0\), and \(\gamma < 0\) are often called Gumbel, Fréchet, and Weibull distribution, respectively. We can determine \(a_n\) and \(b_n\) based on the following fact:

Fact 3 (Theorem 1.2.1 and Corollary 1.2.4 of [8]). If \(F\) is in the class of domain of attraction of \(G\), i.e., there exist a sequence of positive numbers \(\{a_n\}\) and a sequence of real numbers \(\{b_n\}\) such that

\[
\lim_{n \to \infty} F(a_n x + b_n)^n = G(x),
\]

for every continuity point \(x\) of \(G\), then

1. for \(\gamma > 0\): \(x^a = \infty\), \(a_n = U(n)\) and \(b_n = 0\),
2. for \(\gamma < 0\): \(x^a < \infty\), \(a_n = x^a - U(n)\) and \(b_n = x^a\),
3. for \(\gamma = 0\): \(a_n = g(U(n))\) and \(b_n = U(n)\), where

\[
g(t) = \frac{\int_0^t [1 - F(s)] ds}{1 - F(t)},
\]

where \(U(n) = \inf\{y \in \mathbb{R}: F(y) \geq 1 - 1/n\}\).

To prove Theorem 2, we need the following lemma:

Lemma 2. If \(F\) satisfies Eq. (7) with \(\gamma = 0\) and \(\lim_{n \to \infty} a_n = 1/\lambda \in (0, \infty)\), then

\[
\lim_{n \to \infty} n\mathbb{P}(X_1 > b_n - x) = e^{\lambda x}, \quad \text{for any } x \in \mathbb{R}.
\]

Proof. For \(0 < G_\gamma(x) < 1\), Eq. (7) with \(x\) replaced by \(-x\) yields

\[
\lim_{n \to \infty} n\mathbb{P}(X_1 > -a_n x + b_n) = -\log G_\gamma(-x).
\]

In fact, for every \(n\), the following relation holds:

\[
F(a_n x + b_n)^n = \mathbb{P}(X_1 \leq a_n x + b_n)^n = \left[1 - n\mathbb{P}(X_1 > a_n x + b_n)\right]^n.
\]

If Eq. (7) holds and \(\lim_{n \to \infty} a_n = 1/\lambda \in (0, \infty)\), then

\[
\lim_{n \to \infty} F(x + b_n)^n = \lim_{n \to \infty} \mathbb{P}\left(a_n, \frac{M_n - b_n}{a_n} \leq x\right) = G_\gamma(\lambda x),
\]

by the same argument as the one used for proving Lemma 1 in Section 14 of [2].

Because \(U(n)\) appearing in Fact 3 converges to \(x^a\) as \(n \to \infty\), we obtain \(\lim_{n \to \infty} a_n = \infty\) for \(\gamma > 0\) and \(\lim_{n \to \infty} a_n = 0\) for \(\gamma < 0\). Thus \(\lim_{n \to \infty} a_n = 1/\lambda \in (0, \infty)\) implies \(\gamma = 0\). Therefore, we prove Eq. (8) by Eqs. (9) and (10).

Proof of Theorem 2. In order to obtain the limit distribution of \(D_n(1)\), we calculate the characteristic function \(\mathbb{E}[e^{itD_n(1)}]\) for \(t \in \mathbb{R}\). Since \(D_n(1; x) = \sum_{j=1}^n I_{\theta_j, \infty}(x + X_j)\) and \(\mathbb{P}(I_{\theta_j, \infty}(x + X_j) = 1) = \mathbb{P}(X_2 > \theta_n - x)\) for \(x \in \mathbb{R}\), \(D_n(1; x)\) is binomially distributed. So we have
 Similarly, we obtain for case 2.

\[ E[e^{\mu X_1}] = \sum_{k=0}^{n-1} e^{\mu k} \cdot \left( \frac{n-1}{k} \right) \mathbb{P}(X_2 > \theta_n - x)^{\frac{(n-1)}{k}}. \]

By using Fubini’s theorem, we obtain

\[ E[e^{\mu X_1}] = \int_{x_n}^{\mu} E[e^{\mu X_1}] F(dx) \]

\[ = F(\theta_n - x) \cdot e^{\mu 0} + [1 - F(\theta_n - x)] \cdot e^{\mu(n-1)} \]

\[ + \int_{\theta_n - x}^{\mu} \left[ 1 + \frac{(n-1)\mathbb{P}(X_2 > \theta_n - x)}{n-1} (e^{\mu} - 1) \right]^{n-1} F(dx). \]

(11)

Because \( \lim_{n \to \infty} \theta_n = x^+ \) by the assumption and Fact 3, we have \( \lim_{n \to \infty} F(\theta_n - x^+) = F(-0) \). If \( x_+ > 0 \), then \( \lim_{n \to \infty}(\theta_n - x_+) = x^+ - x_+ < x^+ \) which implies \( \lim_{n \to \infty}[1 - F(\theta_n - x_+)] > 0 \). In this case, the second term of the right-hand side of Eq. (11) does not converge. If \( x_+ \leq 0 \), then \( \lim_{n \to \infty}[1 - F(\theta_n - x_+)] = 0 \) so that this term converges to 0.

So we have, from Lemma 2,

\[ \lim_{n \to \infty} E[e^{\mu X_1}] = \begin{cases} 
\text{does not exist}, & \text{for case 1 of Theorem 2}, \\
F(-0) \cdot e^{\mu 0} + \int_{x_n}^{\mu} \exp(e^{\mu}(e^{\mu} - 1)) F(dx), & \text{for case 2 of Theorem 2}, \\
\int_{x_n}^{\infty} \exp(e^{\mu}(e^{\mu} - 1)) F(dx), & \text{for case 3 of Theorem 2}.
\end{cases} \]

Noting that \( \exp(e^{\mu}(e^{\mu} - 1)) \) is the characteristic function of the Poisson distribution with parameter \( e^{\lambda \mu} \), we obtain, for case 3,

\[ P(D(1) = k) = \int_{x_n}^{\infty} P(\text{Poisson}(e^{\lambda \mu}) = k) F(dx) = \int_{x_n}^{\infty} \exp(-e^{\lambda \mu}) \frac{(e^{\lambda \mu})^k}{k!} F(dx). \]

Similarly, we obtain for case 2.

**Proof of Corollary 1.** Since \( F \) is continuous and nondecreasing, we obtain, by Fact 3, \( \theta_n = F^{-1}(1 - 1/n) \) and

\[ a_n = \frac{\int_{\theta_n}^{\mu} (1 - F(s))ds}{1 - F(\theta_n)} = n \int_{\theta_n}^{\mu} (1 - F(s))ds. \]

If we assume \( x^+ < \infty \), we have

\[ 0 \leq a_n \leq n[1 - F(\theta_n)] \int_{\theta_n}^{\mu} ds = x^+ - \theta_n. \]

(12)

Because \( \lim_{n \to \infty} \theta_n = F^{-1}(1) = x^+ \), Eq. (12) results in \( \lim_{n \to \infty} a_n = 0 \) which is a contradiction. Thus we have \( x^+ = \infty \).

The assumption of Theorem 2 gives

\[ \lim_{n \to \infty} n[1 - F(\theta_n + y)] = e^{-\lambda y}, \]

(13)

for all \( y \in \mathbb{R} \) by Lemma 2. Because \( \theta_n = F^{-1}(1 - 1/n) \), we can rewrite Eq. (13) as

\[ \lim_{n \to \infty} \frac{1 - F(\theta_n + y)}{1 - F(\theta_n)} = e^{-\lambda y}. \]

For each \( x \in \mathbb{R} \), there exists \( n \) with \( \theta_n - 1 \leq x \leq \theta_n \) such that

\[ \frac{1 - F(\theta_n + y)}{1 - F(\theta_n)} \leq \frac{1 - F(x + y)}{1 - F(x)} \leq \frac{1 - F(\theta_{n-1} + y)}{1 - F(\theta_{n-1})}, \]

since \( 1 - F(x) \) is nonincreasing in \( x \) and \( \theta_n \) is nondecreasing in \( n \). Thus we obtain \( \lim_{n \to \infty}[1 - F(x + y)]/[1 - F(x)] = e^{-\lambda y} \). This implies

\[ \lim_{x \to \infty} \frac{f(x + y)}{f(x)} = e^{-\lambda y}, \]

for any \( y \in \mathbb{R} \) by L’Hôpital’s rule.

Conversely, we assume \( \lim_{x \to \infty} f(x + y)/f(x) = e^{-\lambda y} \), for any \( y \in \mathbb{R} \). We can easily prove \( \lim_{n \to \infty} n[1 - F(a_n y + \theta_n)] = e^{-y} \) by choosing \( a_n = 1/\lambda \) and \( \theta_n = F^{-1}(1 - 1/n) \). \( \square \)
Example 1. Consider the exponential distribution with parameter \( \lambda > 0 \) whose distribution function is given by

\[
F(x) = \begin{cases} 
1 - e^{-\lambda x}, & \text{if } x \geq 0, \\
0, & \text{otherwise}.
\end{cases}
\]

In this case, we take \( a_n = 1/\lambda \) and \( \theta_n = (1/\lambda) \log n \) so that

\[
F(a_n x + \theta_n)^n = \left[ 1 - e^{-(x+\log n)} \right]^n = \left( 1 - \frac{e^{-x}}{n} \right)^n \rightarrow \exp(-e^{-x}).
\]

For \( k = 0, 1, \ldots \), we have

\[
P(D(1) = k) = \frac{1}{k!} \int_0^\infty \exp[kx - e^{kx}] e^{-kx} dx = \frac{1}{k!} \int_0^\infty e^{-x} x^{k-2} dx.
\]

Using \( P(D(1) = 2) = 1/(2e) \) and

\[
k(k-1)P(D(1) = k) - (k-1)(k-2)P(D(1) = k - 1) = \frac{1}{e(k-2)!} \quad (k \geq 2),
\]

we obtain

\[
P(D(1) = k) = \frac{1}{k(k-1)} \cdot \frac{1}{e} \sum_{l=2}^{k} \frac{1}{(l-2)!} \quad (k \geq 2).
\]

This gives \( \lim_{k \to \infty} P(D(1) = k)/k^{-2} = 1 \).

Acknowledgements

We thank Takaaki Shimura for valuable comments on Corollary 1. This work was supported in part by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (S) (18100001).

REFERENCES