数理解析研究所講究録
第 1551 巻 2007 年 81-86

Limit theorems for some statistics of a generalized threshold network model

Yusuke Ide, Norio Konno *
Department of Applied Mathematics,
Yokohama National University

Naoki Masuda †
Graduate School of Information Science and Technology,
The University of Tokyo

Abstract

In this report, we state limit theorems for the number of edges, the number of triangles and the clustering coefficients of a generalized threshold network model. We also give examples of these limit theorems.

1 Introduction

The threshold network model is a type of finite random graphs that is generated on n vertices labeled $1, \ldots, n$ with independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$. We connect a pair of vertices i and j with $i \neq j$ by an edge when $X_i + X_j > \theta$ for a given threshold θ. The threshold network model is a subclass of so called hidden variable models, and its mean behavior [1, 2, 5, 7, 8] and limit theorems [4, 6] have been analyzed. Recently, a generalization of the threshold network model was formulated and several limit theorems were studied [3]. Here we review the generalized model. Let $\mathbb{R}^d$ be the d-dimensional Euclidean space. We prepare an i.i.d. sequence of $\mathbb{R}^d$-valued random variables $X_1, \ldots, X_n$ and associate the random variable X_i with vertex i. Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-field of $\mathbb{R}$. Now we introduce Borel measurable functions $f_c^m : (\mathbb{R}^d)^2 \to \mathbb{R}$ with $f_c^m(x, y) = f_c^m(y, x)$ for all $m \in \{1, \ldots, l\}$. For a given finite collection of Borel measurable sets $\mathcal{C} = \{B_1, \ldots, B_l\}$ with $B_m \in \mathcal{B}(\mathbb{R})$ for all $m \in \{1, \ldots, l\}$, we connect vertices i and j $(i \neq j)$ if $f_c^m(X_i, X_j) \in B_m$ for all $m \in \{1, \ldots, l\}$. In other words, we form an edge $\langle i, j \rangle$ if $\prod_{m=1}^{l} I_{B_m}(f_c^m(X_i, X_j)) = 1$ for $i \neq j$, where $I_A(x)$ denotes the indicator function, i.e., $I_A(x) = 1$ for $x \in A$ and $I_A(x) = 0$ otherwise. Thus we obtain a random graph $G_{\mathcal{C}}(X_1, \ldots, X_n)$. References and more details are found in [3].

*Postal address: 79-5, Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan
†Postal address: 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2 Limit Theorems

In this section, we state the limit theorems that are suitable modifications of the theorems proved in [3]. Hereafter, we only consider the one-dimensional ($d = 1$) and $l = 1$ case. Extensions of the following results to general d and l are straightforward. For simplicity, we may write $f_c \equiv f_c^1$ and $B \equiv \mathcal{C} = \{B_1\}$.

2.1 Edges and Triangles

When we choose $h_D(x, y) = I_B(f_c(x, y))$, as the kernel function, we define the following two statistics:

$$D_n = \sum_{1 \leq i < j \leq n} h_D(X_i, X_j), \quad \text{and} \quad D_n(i) = \sum_{\substack{1 \leq j \leq n \\ j \neq i}} h_D(X_i, X_j).$$

Here D_n is the number of edges in the random graph $G_B(X_1, \dots, X_n)$ and $D_n(i)$ is the number of edges connected to vertex i, i.e., the degree of vertex i. Using another kernel function $h_T(x, y, z) = I_B(f_c(x, y)) \cdot I_B(f_c(y, z)) \cdot I_B(f_c(z, x))$, we define the following statistics for the number of triangles:

$$T_n = \sum_{1 \leq i < j < k \leq n} h_T(X_i, X_j, X_k), \quad \text{and} \quad T_n(i) = \sum_{\substack{1 \leq j < k \leq n \\ j,k \neq i}} h_T(X_i, X_j, X_k).$$

Here T_n denotes the number of triangles in the random graph and $T_n(i)$ is the number of triangles including vertex i. Limit theorems for the statistics are the following:

Theorem 1. *As* $n \to \infty$,
(i) for any $x \in \mathbb{R}$, $\dfrac{D_n(1; x)}{n - 1} \to D(1; x) \equiv \mathbb{P}(h_D(x, X_2) = 1)$ *almost surely,*
(ii) $\dfrac{D_n}{\binom{n}{2}} \to D \equiv \mathbb{E}[D(1; X_1)] = \mathbb{P}(h_D(X_1, X_2) = 1)$ *almost surely,*
(iii) for any $x \in \mathbb{R}$, $\dfrac{T_n(1; x)}{\binom{n-1}{2}} \to T(1; x) \equiv \mathbb{P}(h_T(x, X_2, X_3) = 1)$ *almost surely,*
(iv) $\dfrac{T_n}{\binom{n}{3}} \to T \equiv \mathbb{E}[T(1; X_1)] = \mathbb{P}(h_T(X_1, X_2, X_3) = 1)$ *almost surely,*
where

$$D_n(i; x) = \sum_{\substack{1 \leq j \leq n \\ j \neq i}} h_D(x, X_j), \quad \textit{and} \quad T_n(i; x) = \sum_{\substack{1 \leq j < k \leq n \\ j,k \neq i}} h_T(x, X_j, X_k).$$

2.2 Clustering Coefficients

The local clustering coefficient $C_n(i)$ of vertex i is defined by

$$C_n(i) = \frac{T_n(i)}{\binom{D_n(i)}{2}} \cdot I_{\{D_n(i) \geq 2\}} + w \cdot I_{\{D_n(i) = 0,1\}},$$

for an indeterminate w. The global clustering coefficient C_n is defined by

$$C_n = \frac{1}{n}\sum_{i=1}^{n} C_n(i).$$

The following limit theorems for the local and global clustering coefficients were proved:

Theorem 2. *As* $n \to \infty$,
(i) for any $x \in \mathbb{R}$, $C_n(1;x) \to C(1;x) \equiv \frac{T(i;x)}{D(i;x)^2} \cdot I_{\{D(i;x)>0\}} + w \cdot I_{\{D(i;x)=0\}}$ *almost surely,*
(ii) $C_n \to C \equiv \mathbb{E}[C(1;X_1)]$ *almost surely,*

where

$$C_n(i;x) = \frac{T_n(i;x)}{\binom{D_n(i;x)}{2}} \cdot I_{\{D_n(i;x)\geq 2\}} + w \cdot I_{\{D_n(i;x)=0,1\}}.$$

3 Examples

In this section, we give several examples for $D(1;x)$, $T(1;x)$, $C(1;x)$, D, T, C and the distribution of $D(1) \equiv D(1;X_1)$. Let f and $f_{D(1)}$ denote the distributions of X_1 and $D(1)$, respectively. Note that domain of $f_{D(1)}(k)$ is always $0 \leq k \leq 1$.

Case 1 : When we choose $f_c(x,y) = x + y$ and $B = (\theta, \infty)$ for $\theta \in \mathbb{R}$, the random graph becomes the original threshold network model. First, we give a table of examples for the Bernoulli distribution. For simplicity, we omit trivial cases ($\theta < 0, 2 \leq \theta$).

distribution	Bernoulli
$f(x)$	$p \cdot \delta_1(x) + (1-p) \cdot \delta_0(x) \ : \ p \in (0,1)$
$f_{D(1)}(k)$	$\begin{cases} p \cdot \delta_1(k) + (1-p) \cdot \delta_p(k) & \text{if } 0 \leq \theta < 1, \\ p \cdot \delta_p(k) + (1-p) \cdot \delta_0(k) & \text{if } 1 \leq \theta < 2. \end{cases}$
$D(1;x)$	$\begin{cases} D(1;1) = 1,\ D(1;0) = p & \text{if } 0 \leq \theta < 1, \\ D(1;1) = p,\ D(1;0) = 0 & \text{if } 1 \leq \theta < 2. \end{cases}$
$T(1;x)$	$\begin{cases} T(1;1) = p(2-p),\ T(1;0) = p^2 & \text{if } 0 \leq \theta < 1, \\ T(1;1) = p^2,\ T(1;0) = 0 & \text{if } 1 \leq \theta < 2. \end{cases}$
$C(1;x)$	$\begin{cases} C(1;1) = p(2-p),\ C(1;0) = 1 & \text{if } 0 \leq \theta < 1, \\ C(1;1) = 1,\ C(1;0) = w & \text{if } 1 \leq \theta < 2. \end{cases}$
D, T, C	$\begin{cases} D = p(2-p),\ T = p^2(3-2p),\ C = 1 - p(1-p)^2 & \text{if } 0 \leq \theta < 1, \\ D = p^2,\ T = p^3,\ C = p + (1-p) \cdot w & \text{if } 1 \leq \theta < 2. \end{cases}$

Next, we give a table for the exponential distribution [5].

distribution	exponential
$f(x)$	$\lambda e^{-\lambda x},\ x \in (0,\infty)\ :\ \lambda > 0$
$f_{D(1)}(k)$	$\begin{cases} \delta_1(k) & \text{if } \theta \le 0, \\ I_{(e^{-\lambda\theta},1)}(k) \cdot \frac{e^{-\lambda\theta}}{k^2} + e^{-\lambda\theta} \cdot \delta_1(k) & \text{if } \theta > 0. \end{cases}$
$D(1;x)$	$\begin{cases} I_{(0,\infty)}(x) & \text{if } \theta \le 0, \\ I_{(0,\theta)}(x) \cdot e^{-\lambda(\theta-x)} + I_{(\theta,\infty)}(x) & \text{if } \theta > 0. \end{cases}$
$T(1;x)$	$\begin{cases} I_{(0,\infty)}(x) & \text{if } \theta \le 0, \\ I_{(0,\theta/2]}(x) \cdot e^{-2\lambda(\theta-x)} + I_{(\theta/2,\theta]}(x) \cdot [\lambda(2x-\theta)+1]e^{-\lambda\theta} & \\ +I_{(\theta,\infty)}(x) \cdot (\lambda\theta+1)e^{-\lambda\theta} & \text{if } \theta > 0. \end{cases}$
$C(1;x)$	$\begin{cases} I_{(0,\infty)}(x) & \text{if } \theta \le 0, \\ I_{(0,\theta/2]}(x) + I_{(\theta/2,\theta]}(x) \cdot [\lambda(2x-\theta)+1]e^{-\lambda(2x-\theta)} & \\ +I_{(\theta,\infty)}(x) \cdot (\lambda\theta+1)e^{-\lambda\theta} & \text{if } \theta > 0. \end{cases}$
$D,\ T,\ C$	$\begin{cases} D=1,\ T=1,\ C=1 & \text{if } \theta \le 0, \\ D=(\lambda\theta+1)e^{-\lambda\theta},\ T=4e^{-3\lambda\theta/2}-3e^{-2\lambda\theta}, & \\ C=1-\frac{4}{9}e^{-\lambda\theta/2}+\frac{1}{2}(3\lambda\theta+2)e^{-2\lambda\theta} & \text{if } \theta > 0. \end{cases}$

A remarkable feature of $f_{D(1)}$ is existence of the power law k^{-2} which is referred to as the scale-free property. Remark that existence of the delta measure δ_1 is always proved for distributions that are absolutely continuous and have a lower cutoff, i.e., $\operatorname{supp} f = [a,\infty)$, where $a \in \mathbb{R}$ and $\operatorname{supp} f = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$ is the support of f.

Finally, we consider the bilateral exponential distribution. For simplicity, we only show $D(1;x)$ and the distribution of $D(1)$.

distribution	bilateral exponential
$f(x)$	$\frac{1}{2}e^{-\lambda\lvert x\rvert}\ :\ \lambda > 0$
$f_{D(1)}(k)$	$\begin{cases} e^{\lambda\theta} \cdot I_{(0,\frac{1}{2})}(k) + I_{(\frac{1}{2},1-\frac{1}{2}e^{\lambda\theta})}(k) \cdot \frac{e^{-\lambda\theta}}{4(1-k)^2} + e^{-\lambda\theta} \cdot I_{(1-\frac{1}{2}e^{\lambda\theta},1)}(k) & \text{if } \theta < 0, \\ I_{(0,1)}(k) & \text{if } \theta = 0, \\ e^{\lambda\theta} \cdot I_{(0,\frac{1}{2}e^{-\lambda\theta})}(k) + I_{(\frac{1}{2}e^{-\lambda\theta},\frac{1}{2})}(k) \cdot \frac{e^{-\lambda\theta}}{4k^2} + e^{-\lambda\theta} \cdot I_{(\frac{1}{2},1)}(k) & \text{if } \theta > 0. \end{cases}$
$D(1;x)$	$I_{(-\infty,\theta)}(x) \cdot \frac{1}{2}e^{-\lambda(\theta-x)} + I_{(\theta,\infty)}(x) \cdot \left(1-\frac{1}{2}e^{\lambda(\theta-x)}\right)$

In this case, $f_{D(1)}$ is mixture of the uniform distribution and the power law (k^{-2} or $(1-k)^{-2}$). Particularly, when $\theta = 0$, the distribution of $D(1)$ becomes the uniform distribution on $(0,1)$. Note that when $\theta = 0$, the same result also holds for distributions that are absolutely continuous, symmetric, i.e., $f(x) = f(-x)$, and of infinite support, i.e., $\operatorname{supp} f = (-\infty,\infty)$.

Case 2 : Next we consider the case $f_c(x,y) = x + y$, $B = \bigcup_{j=1}^{N}(a_j,b_j]$ for a finite $N \in \{1,2,\dots\}$, where $0 \le a_1 \le b_1 \le \dots \le a_j \le b_j \le a_{j+1} \le \dots \le a_N \le b_N$. We derive the following distribution of $D(1)$ for the exponential distribution with parameter λ:

$$f_{D(1)}(k) = \sum_{j=0}^{N} I_{(e^{\lambda b_j} S_{j+1}, e^{\lambda a_{j+1}} S_{j+1})} \cdot \frac{S_{j+1}}{k^2} + \sum_{j=1}^{N} I_{(e^{\lambda b_j} S_{j+1}, e^{\lambda a_j} S_j)} \cdot \frac{e^{-\lambda b_j} - S_{j+1}}{(1-k)^2},$$

where $b_0 = 0$ and $S_j = \sum_{i=j}^{N} \left(e^{-\lambda a_i} - e^{-\lambda b_i}\right) = \mathbb{P}(X_1 \in \bigcup_{i=j}^{N}(a_i, b_i]) \in [0,1]$ for $j \in \{1, \ldots, N\}$. The original threshold network model is the case $N = 1$ with $a_1 = \theta$ and $b_1 = \infty$, where we set $e^{-\lambda\infty} \equiv 0$. For $B = \bigcup_{j=1}^{\infty}(a_j, b_j]$, we can obtain $f_{D(1)}$ by replacing N with ∞.

Case 3: Let us consider the case in which $f_c(x,y) = x + y$, $B = \bigcup_{j=1}^{N}(a_j, b_j]$ for a finite $N \in \{1, 2, \ldots\}$, where $0 \leq a_1 \leq b_1 \leq \ldots \leq a_j \leq b_j \leq a_{j+1} \leq \ldots \leq a_N \leq b_N \leq 1$, and the distribution of X_1 is the uniform distribution on $(0,1)$. We derive the following distribution of $D(1)$:

$$f_{D(1)}(k) = I_{(0,S_1)}(k) + (1 - b_N) \cdot \delta_0(k) + \sum_{i=1}^{N}(a_i - b_{i-1}) \cdot \delta_{S_i}(k),$$

where $b_0 = 0$ and $S_j = \sum_{i=j}^{N}(b_i - a_i) = \mathbb{P}(X_1 \in \bigcup_{i=j}^{N}(a_i, b_i])$. The uniform distribution $I_{(0,S_1)}(k)$ corresponds to intervals included in B, and the delta measures correspond to gaps, i.e., sets included in $[0,1] \setminus B$. As an example, let us consider the case

$$B = K_n \equiv \bigcup_{\substack{a_m = 0,2 \\ 1 \leq m \leq n}} \left[\sum_{m=1}^{n} \frac{a_m}{3^m}, \sum_{m=1}^{n} \frac{a_m}{3^m} + \frac{1}{3^n}\right].$$

For example, $K_2 = [0, 1/3^2] \cup [2/3^2, 3/3^2] \cup [6/3^2, 7/3^2] \cup [8/3^2, 1]$. We obtain

$$f_{D(1)}(k) = I_{(0,2^n/3^n)}(k) + \frac{1}{3^n} \cdot \sum_{i=1}^{2^n-1} \delta_{\frac{2i-1}{3^n}}(k) + \sum_{i=1}^{2^{n-1}-1} \frac{1}{3 \cdot 3^{|n-1-i|}} \cdot \delta_{\frac{2i}{3^n}}(k).$$

The limit set $K = \bigcap_{n=1}^{\infty} K_n$ is the Cantor set. Because the Lebesgue measure of K equals zero, it is trivial that $f_{D(1)}(k) = \delta_0(k)$ for absolutely continuous distributions.

Case 4 : We study the case $f_c(x,y) = xy$, $B = (\theta, \infty)$, $\theta > 0$ as an example of f_c that is different from addition. The distribution of $D(1)$ for the exponential distribution with parameter λ is the following:

$$f_{D(1)}(k) = I_{(0,1)}(k) \cdot \frac{\lambda^2\theta \cdot e^{\lambda^2\theta/\log k}}{k(\log k)^2}.$$

In this case, the distribution deviates from the power law.

References

[1] BOGUÑÁ, M. AND PASTOR-SATORRAS, R. (2003). Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112.

[2] CALDARELLI, G., CAPOCCI, A., DE LOS RIOS, P. AND MUÑOZ, M. A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702.

[3] IDE, Y., KONNO, N. AND MASUDA, N. (2007). Statistical properties of a generalized threshold network model. *Preprint.*

[4] KONNO, N., MASUDA, N., ROY, R. AND SARKAR, A. (2005). Rigorous results on the threshold network model. *J. Phys. A* **38**, 6277–6291.

[5] MASUDA, N., MIWA, H. AND KONNO, N. (2004). Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights. *Phys. Rev. E* **70**, 036124.

[6] NAJIM, C.A. AND RUSSO, R.P. (2003). On the number of subgraphs of a specified form embedded in a random graph. *Methodol. Comput. Appl. Probab.* **5**, 23–33.

[7] SERVEDIO, V. D. P., CALDARELLI, G. AND BUTTÁ, P. (2004). Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys. Rev. E* **70**, 056126.

[8] SÖDERBERG, B. (2002). General formalism for inhomogeneous random graphs. *Phys. Rev. E* **66**, 066121.