Rigorous results on the threshold network model

2005 J. Phys. A: Math. Gen. 38 6277

(http://iopscience.iop.org/0305-4470/38/28/002)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 6277–6291 doi:10.1088/0305-4470/38/28/002

Rigorous results on the threshold network model

Norio Konno[1], **Naoki Masuda**[2], **Rahul Roy**[3] **and Anish Sarkar**[3]

[1] Faculty of Engineering, Yokohama National University, 79-5, Tokiwadai, Hodogaya, Yokohama, 240-8501, Japan
[2] Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, 2-1, Hirosawa, Wako, Saitama, 351-0198, Japan
[3] Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India

Received 23 March 2005, in final form 25 May 2005
Published 29 June 2005
Online at stacks.iop.org/JPhysA/38/6277

Abstract
We analyse the threshold network model in which a pair of vertices with random weights are connected by an edge when the summation of the weights exceeds a threshold. We prove some convergence theorems and central limit theorems on the vertex degree, degree correlation and the number of prescribed subgraphs. We also generalize some results in the spatially extended cases.

PACS numbers: 89.75.Hc, 89.75.Da, 89.75.Fb

1. Introduction

Recently, real large-scale network data have been analysed in various fields, and the corresponding random graphs have been studied. Many of these graphs exhibit the power-law form of the degree distribution, with the power-law exponent typically between 2 and 3. In view of this, physicists have proposed stochastic algorithms for generating growing networks. The underlying assumption of these growing networks is that a vertex with a certain fixed number of edges is added to the graph one by one at each discrete time step. To obtain a power law, the growth mechanism is usually supplied by the so-called preferential attachment, which stipulates that each newly introduced edge is more likely to be connected to a vertex (which already exists in the graph) that has a larger degree. For a review of the studies on such models see [1, 2].

However, not all real networks are growing; algorithms for generating non-growing realistic graphs could be more appropriate for real situations in which the number of vertices does not change rapidly. In this regard, a type of random graph in which each of n vertices is assigned a random variable (weight) was proposed and its mean behaviour has been analysed [3–6]. Interestingly, the power-law degree distribution can emerge even with a weight distribution that is not equipped with a power law. For example, the mean-field analysis suggests a power-law degree distribution with a scaling exponent -2 when the vertex weights are independent and identically distributed (i.i.d.) random variables obeying the exponential distribution.

Our first study is regarding such graphs. Formally, our model consists of n vertices labelled $1, \ldots, n$ and an i.i.d. sequence of random variables $X_1, \ldots, X_n$ with F denoting their common distribution function. We associate the random variable X_i with the vertex labelled i. Given a fixed threshold value $\theta > 0$, we connect the vertices i and j by an edge $\langle i, j\rangle$ if and only if $X_i + X_j > \theta$ and $i \neq j$. Let G_θ be the random graph thus produced. A simple coupling argument shows that $\mathbb{P}\{\langle i, j\rangle \in G_{\theta'}\} \leqslant \mathbb{P}\{\langle i, j\rangle \in G_\theta\}$ whenever $\theta' \geqslant \theta$.

Let $D_n(i) := \#\{j : \langle i, j\rangle \in G_\theta\}$, i.e., the degree of the vertex i in the graph G_θ. Note that $\{D_n(i) : i \geqslant 1\}$ are identical in distribution and let D_n denote a random variable with this common distribution. The distribution D_n can be obtained as follows: given $(n+1)$ vertices, conditioned on the event $X_1 = x$, vertices $j \in \{2, \ldots, n+1\}$, will connect to the vertex 1 if and only if $X_j > \theta - x$. Therefore, for $0 \leqslant k \leqslant n$, we have

$$\mathbb{P}(D_{n+1} = k) = \int_{-\infty}^{\infty} \binom{n}{k} [1 - F(\theta - x)]^k [F(\theta - x)]^{n-k} F(\mathrm{d}x). \tag{1}$$

Here $F(\mathrm{d}x)$ represents the probability measure on the real line representing the law of X_1. Equation (1) allows us to obtain the following asymptotic result:

Theorem 1. *As $n \to \infty$,*

$$\frac{D_n}{n} \Longrightarrow 1 - F(\theta - X_1). \tag{2}$$

Here $\Longrightarrow$ stands for convergence in distribution.

In a data-analysis context, we are often concerned with the degree correlation between a pair of vertices [1, 2]. Here we discuss the asymptotic properties of the graph. Our first result is

Theorem 2. *$\frac{D_n(1)}{n}$ and $\frac{D_n(2)}{n}$ are asymptotically independent.*

This asymptotic independence breaks under the condition that the vertices 1 and 2 are connected. However, we need to be careful about this statement, in the sense that we will need some conditions on the distribution function F of the weight of a vertex. In particular, we assume that

Assumption 1. *There exist u and v in the support of the distribution function F (i.e., for any $\epsilon > 0$, $F(u+\epsilon) > F(u-\epsilon)$ and $F(v+\epsilon) > F(v-\epsilon)$) such that $u < \theta/2 < v$ and $u + v > \theta$.*

Theorem 3. *Given that the vertices 1 and 2 are connected, under the above assumption $\frac{D_n(1)}{n}$ and $\frac{D_n(2)}{n}$ are not asymptotically independent.*

The importance of the assumption can be understood by looking at the situation when the assumption does not hold. In that case, the vertices can be classified into two groups: those which have weights $\theta/2$ or less and those which have weights more than $\theta/2$. In the random graph constructed, each of the vertices of the former group remains isolated, whereas the vertices of the latter group form a complete subgraph. Also, by the strong law of large numbers, the probability that a vertex belongs to the former group converges to $F(\theta/2)$ and the probability that a vertex belongs to the latter group converges to $1 - F(\theta/2)$. Moreover, given two vertices (1 and 2 (say)) are connected, they must belong to the latter group and then we have

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{D_n(1)}{n} = 1 - F(\theta/2), \lim_{n\to\infty} \frac{D_n(2)}{n} = 1 - F(\theta/2) | \text{vertices 1 and 2 are connected}\right]$$
$$= \mathbb{P}\left[\lim_{n\to\infty} \frac{D_n(1)}{n} = 1 - F(\theta/2) | \text{vertices 1 and 2 are connected}\right] = 1. \tag{3}$$

Thus we obtain conditional asymptotic independence.

A remarkable characteristic of real graphs is the clustering property [1, 2]. The clustering property means an abundance of connected triangles in the random graph. The threshold model exhibits such a clustering property as shown *exactly* in [5, 6].

To formalize this, let $h : \mathbb{R}^3 \to \mathbb{R}$ be given by

$$h(x_1, x_2, x_3) := 1_{\{x_1+x_2>\theta, x_2+x_3>\theta, x_3+x_1>\theta\}}. \tag{4}$$

Also let

$$T_n := \#\{(i, j, k) : 1 \leqslant i < j < k \leqslant n, X_i + X_j > \theta, X_j + X_k > \theta, X_k + X_i > \theta\}, \tag{5}$$

$$\begin{aligned} F_3(\theta) &:= \mathbb{E}\,[h(X_1, X_2, X_3)] \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} F(\mathrm{d}x_1)F(\mathrm{d}x_2)F(\mathrm{d}x_3)h(x_1, x_2, x_3) \\ &= \mathbb{P}(X_1 + X_2 > \theta, X_2 + X_3 > \theta, X_3 + X_1 > \theta), \end{aligned} \tag{6}$$

$$\zeta_1(F) := \mathbb{E}\left[\left(\int_{\mathbb{R}}\int_{\mathbb{R}} F(\mathrm{d}x_2)F(\mathrm{d}x_3)h(X_1, x_2, x_3)\right)^2\right] - (F_3(\theta))^2 > 0. \tag{7}$$

Note that T_n counts all triangles in the graph with n vertices.

The asymptotic results are

Theorem 4. *As* $n \to \infty$,

(a) $\dfrac{T_n}{\binom{n}{3}} \to F_3(\theta)$ *almost surely;*

(b) $\sqrt{n}\left[\dfrac{T_n}{\binom{n}{3}} - F_3(\theta)\right] \Rightarrow \sqrt{3\zeta_1(F)}Z$ *where* Z *is a standard normal random variable.*

The method used to show the above results may be generalized easily to obtain a count of not only triangles but subgraphs in G_θ isomorphic to a fixed graph. In data-analysis contexts, a fixed subgraph is called a motif of the graph. Depending on the types of real networks (e.g., Internet, gene networks, neural networks, social networks), there are some sorts of small motifs that appear in an entire graph significantly more than in the random graphs. These motifs are relevant to functional roles such as signal transduction and information processing apposite to each application [8, 9]. Our results, theorems 8 and 9 in section 4, obtain limit theorems for the motifs of the threshold model.

Besides the extension to general graphs, the above theorem may be extended for local triangle counts, i.e., the number of triangles containing a specified vertex. The clustering coefficient, which is a quantity often used to evaluate the degree of a clustering property, is defined as the normalized number of locally counted triangles averaged over all the vertices. Our local results below show that the vertexwise clustering coefficient satisfies the central limit theorem, which validates the use of expectation in the physics community.

We fix vertex 1 and define $T_n(1)$ as the number of triangles in the set of vertices $\{1, 2, \ldots, n, n+1\}$ of which vertex 1 is a site. In other words,

$$T_n(1) := \#\{2 \leqslant i < j \leqslant n+1 : h(X_1, X_i, X_j) = 1\}. \tag{8}$$

Theorem 5. *As* $n \to \infty$,

$$\frac{T_n(1)}{\binom{n}{2}} \Rightarrow \int_{\mathbb{R}}\int_{\mathbb{R}} F(\mathrm{d}x_1)F(\mathrm{d}x_2)h(X, x_1, x_2) \tag{9}$$

where X *is an independent random variable identical in distribution to* X_1.

Besides the above dimensionless random graph model, a spatial model has been proposed in [7]. Consider a homogeneous Poisson point process of intensity λ on $\mathbb{R}^d$. We insist that the origin, $\mathbf{0}$, is a point of the process. Let $\{\mathbf{0} = \xi_0, \xi_1, \xi_2, \ldots\}$ be an enumeration of the points of the process. Associated with each point ξ_i is a random variable X_i. We assume that $\{X_0, X_1, X_2, \ldots\}$ is an i.i.d. sequence of random variables with common distribution function F. The random graph $G_{\theta,\beta}$ is obtained by connecting ξ_i and ξ_j by an edge if and only if $(X_i + X_j) > \theta|\xi_i - \xi_j|^\beta$ where θ and β, the parameters of the model, are real numbers.

Define the degree of the origin in a sphere of radius r as

$$\Delta_r := \#\{i \geqslant 1 : (X_0 + X_i) > \theta|\xi_i|^\beta \text{and } |\xi_i| \leqslant r\}. \tag{10}$$

Given a fixed $x \in \mathbb{R}$, let

$$f(r; x) = f(|r|; x) := 1 - F(\theta|r|^\beta - x). \tag{11}$$

Define

$$C_r(x) := \int_0^r \tilde{r}^{d-1} f(\tilde{r}; x)\, \mathrm{d}\tilde{r}, \tag{12}$$

and consider the following two cases:

(i) As $r \to \infty$, $C_r(x) \to C(x) := \int_0^\infty \tilde{r}^{d-1} f(\tilde{r}, x)\, \mathrm{d}\tilde{r} < \infty$ for every $x \in \mathbb{R}$.

(ii) There exists a sequence $\{C_r\}$ and a function $g(x)$ such that $C_r \to \infty$ and $\frac{C_r(x) - C_r}{\sqrt{C_r}} \to g(x)$ as $r \to \infty$ for every $x \in \mathbb{R}$.

For the first case, we have

Theorem 6. *If $C_r(x) \to C(x) < \infty$ for every $x \in \mathbb{R}$ as $r \to \infty$, then we have*

$$\Delta_r \Rightarrow \Delta \tag{13}$$

where the characteristic function of the random variable Δ is given by $\phi_\Delta(t) = \int_{\mathbb{R}} F(\mathrm{d}x) \exp(-\lambda c_d C(x)(1 - \exp(\mathrm{i}t)))$ where c_d represents the volume of the $(d-1)$-dimensional unit sphere.

Remark. The random variable Δ represents the degree of the origin in the random graph.

For the second case, we have

Theorem 7. *Suppose that there exists a sequence $\{C_r\}$ such that $C_r \to \infty$ and $\frac{C_r(x) - C_r}{\sqrt{C_r}} \to g(x)$ as $r \to \infty$ for every $x \in \mathbb{R}$. We have*

$$\frac{\Delta_r - \lambda c_d C_r}{\sqrt{\lambda c_d C_r}} \Rightarrow Z + \sqrt{\lambda c_d} g(X_0) \quad as \quad r \to \infty \tag{14}$$

where c_d is as defined in theorem 6.

We end this section with an example of F satisfying the condition in theorem 7. Fix $\beta = 1$ and $d = 2$. Define $F : [0, \infty) \to [0, 1]$ by

$$F(x) := 1 - Cx^{-\alpha} \tag{15}$$

where $0 < \alpha < 2$ and $C > 0$, and

$$C_r := C\theta^{-\alpha} r^{2-\alpha}/(2 - \alpha). \tag{16}$$

Simple computations can be carried out to verify that the conditions of theorem 7 are satisfied in this case with $g(x) = 0$ for all $x \in \mathbb{R}$. In the next few sections we prove our results.

2. Degree D_n of a vertex

To prove theorem 1 observe that, since $(n+1)/n \to 1$, it is enough to show that D_{n+1}/n converges weakly to the required random variable. For $t \in \mathbb{R}$, we have from equation (1)

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{\mathrm{i}tD_{n+1}}{n}\right)\right] &= \sum_{k=0}^{n} \exp\left(\frac{\mathrm{i}tk}{n}\right) \int_{-\infty}^{\infty} \binom{n}{k} [1-F(\theta-x)]^k [F(\theta-x)]^{n-k} F(\mathrm{d}x) \\
&= \int_{-\infty}^{\infty} \sum_{k=0}^{n} \binom{n}{k} \exp\left(\frac{\mathrm{i}tk}{n}\right) [1-F(\theta-x)]^k [F(\theta-x)]^{n-k} F(\mathrm{d}x) \\
&= \int_{-\infty}^{\infty} \left[(1-F(\theta-x))\exp\left(\frac{\mathrm{i}t}{n}\right) + F(\theta-x)\right]^n F(\mathrm{d}x) \\
&= \int_{-\infty}^{\infty} \left[(1-F(\theta-x))\left(1+\frac{\mathrm{i}t}{n}+o\left(\frac{1}{n}\right)\right) + F(\theta-x)\right]^n F(\mathrm{d}x) \\
&= \int_{-\infty}^{\infty} \left[1+\frac{\mathrm{i}t}{n}(1-F(\theta-x))+o\left(\frac{1}{n}\right)\right]^n F(\mathrm{d}x) \\
&\to \int_{-\infty}^{\infty} \exp(\mathrm{i}t(1-F(\theta-x)))F(\mathrm{d}x) \qquad \text{as} \quad n\to\infty, \qquad (17)
\end{aligned}$$

where the limit follows from the dominated convergence theorem because

$$\begin{aligned}
&\left|(1-F(\theta-x))\exp\left(\frac{\mathrm{i}t}{n}\right)+F(\theta-x)\right|^2 \\
&\quad = [(1-F(\theta-x))\cos(t/n)+F(\theta-x)]^2 + (1-F(\theta-x))^2\sin^2(t/n) \\
&\quad = (1-F(\theta-x))^2 + 2F(\theta-x)(1-F(\theta-x))\cos(t/n) + (F(\theta-x))^2 \\
&\quad \leqslant 1. \qquad (18)
\end{aligned}$$

This proves theorem 1.

3. Pair correlation

Now suppose that there are $n+2$ vertices labelled $1, 2, \ldots, n, n+1, n+2$ with the corresponding random variables $X_1, \ldots, X_{n+2}$. Define

$$D_{n,1} := D_{n+2}(n+1) = \#\{j : 1 \leqslant j \leqslant n+2,\ j \neq n+1,\ X_j + X_{n+1} > \theta\} \qquad (19)$$

$$D_{n,2} := D_{n+2}(n+2) = \#\{j : 1 \leqslant j \leqslant n+2,\ j \neq n+2,\ X_j + X_{n+2} > \theta\}. \qquad (20)$$

We have, for $0 \leqslant k, l \leqslant n+1$,

$$\mathbb{P}[D_{n,1}=k, D_{n,2}=l] = \mathbb{P}[D_{n,1}=l, D_{n,2}=k]. \qquad (21)$$

For $k > l$

$$\begin{aligned}
&\mathbb{P}[D_{n,1}=k, D_{n,2}=l] \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)\mathbb{P}[\#\{1 \leqslant j \leqslant n : a+X_j > \theta\} + 1_{\{a+b>\theta\}} = k, \\
&\qquad\qquad \#\{1 \leqslant j \leqslant n : b+X_j > \theta\} + 1_{\{a+b>\theta\}} = l] \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} 1_{\{a+b>\theta\}} F(\mathrm{d}a)F(\mathrm{d}b)\mathbb{P}[\#\{1 \leqslant j \leqslant n : a+X_j > \theta\} = k-1, \\
&\qquad\qquad \#\{1 \leqslant j \leqslant n : b+X_j > \theta\} = l-1]
\end{aligned}$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}}1_{\{a+b\leqslant\theta\}}F(\mathrm{d}a)F(\mathrm{d}b)\mathbb{P}[\#\{1\leqslant j\leqslant n : a+X_j>\theta\}=k,$$
$$\#\{1\leqslant j\leqslant n : b+X_j>\theta\}=l]$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}}1_{\{a+b>\theta\}}g_n(\theta;a,b;k-1,l-1)F(\mathrm{d}a)F(\mathrm{d}b)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}}1_{\{a+b\leqslant\theta\}}g_n(\theta;a,b;k,l)F(\mathrm{d}a)F(\mathrm{d}b), \tag{22}$$

where, for $a\geqslant b$,

$$\begin{aligned}
&g_n(\theta;a,b;k,l)\\
&:=\mathbb{P}[\#\{1\leqslant j\leqslant n : a+X_j>\theta\}=k,\#\{1\leqslant j\leqslant n : b+X_j>\theta\}=l]\\
&=\mathbb{P}[\#\{1\leqslant j\leqslant n : X_j>\theta-a\}=k,\#\{1\leqslant j\leqslant n : X_j>\theta-b\}=l]\\
&=\mathbb{P}[\#\{1\leqslant j\leqslant n : X_j>\theta-b\}=l,\#\{1\leqslant j\leqslant n : \theta-a<X_j\leqslant\theta-b\}=k-l]\\
&=\begin{cases}\dfrac{n!}{l!(k-l)!(n-k)!}[1-F(\theta-b)]^l[F(\theta-b)\\ \quad -F(\theta-a)]^{k-l}[F(\theta-a)]^{n-k} & \text{if } k\geqslant l\geqslant 0\\ 0 & \text{otherwise.}\end{cases}
\end{aligned} \tag{23}$$

For $k=l$ similar calculations show that

$$\mathbb{P}[D_{n,1}=k,D_{n,2}=k] \tag{24}$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}}1_{\{a+b>\theta\}}g_n(\theta;a,b;k-1,k-1)F(\mathrm{d}a)F(\mathrm{d}b)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}}1_{\{a+b\leqslant\theta\}}g_n(\theta;a,b;k,k)F(\mathrm{d}a)F(\mathrm{d}b)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a\leqslant b\}}1_{\{a+b>\theta\}}g_n(\theta;b,a;k-1,k-1)F(\mathrm{d}a)F(\mathrm{d}b)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a\leqslant b\}}1_{\{a+b\leqslant\theta\}}g_n(\theta;b,a;k,k)F(\mathrm{d}a)F(\mathrm{d}b). \tag{24}$$

Now let

$$d_n(1):=\#\{j : 1\leqslant j\leqslant n, X_j+X_{n+1}>\theta\} \tag{25}$$

and

$$d_n(2):=\#\{j : 1\leqslant j\leqslant n, X_j+X_{n+2}>\theta\}; \tag{26}$$

then

$$D_{n,1}\geqslant d_n(1)\geqslant D_{n,1}-1 \tag{27}$$

and

$$D_{n,2}\geqslant d_n(2)\geqslant D_{n,2}-1. \tag{28}$$

Thus the asymptotic distributions of $\left(\frac{D_{n,1}}{n},\frac{D_{n,2}}{n}\right)$ and $\left(\frac{d_n(1)}{n},\frac{d_n(2)}{n}\right)$ are identical, and we work out pair correlation with $d_n(1)$ and $d_n(2)$ instead of $D_{n,1}$ and $D_{n,2}$.

Now, for all $0\leqslant k,l\leqslant n$,

$$\mathbb{P}[d_n(1)=k,d_n(2)=l]=\mathbb{P}[d_n(1)=l,d_n(2)=k], \tag{29}$$

and, for $k > l$,

$$\mathbb{P}[d_n(1) = k, d_n(2) = l] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} g_n(\theta; a, b; k, l) F(\mathrm{d}a) F(\mathrm{d}b), \quad (30)$$

while, for $k = l$,

$$\begin{aligned}
\mathbb{P}[d_n(1) = k, d_n(2) = k] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} g_n(\theta; a, b; k, k) F(\mathrm{d}a) F(\mathrm{d}b) \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a\leqslant b\}} g_n(\theta; b, a; k, k) F(\mathrm{d}a) F(\mathrm{d}b) \\
&= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} g_n(\theta; a, b; k, k) F(\mathrm{d}a) F(\mathrm{d}b) \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} g_n(\theta; a, b; k, k) F(\mathrm{d}a) F(\mathrm{d}b). \quad (31)
\end{aligned}$$

Now let $-\infty < s, t < \infty$. To derive pair independence, let us consider the characteristic function. We have

$$\mathbb{E}[\exp(\mathrm{i}sd_n(1) + \mathrm{i}td_n(2))] = \sum_{k=0}^{n}\sum_{l=0}^{n} \exp(\mathrm{i}sk + \mathrm{i}tl)\mathbb{P}[d_n(1) = k, d_n(2) = l]. \quad (32)$$

We break the above double sum into three parts, when (i) $k > l$, (ii) $k < l$ and (iii) $k = l$.

For (i) we have

$$\begin{aligned}
&\sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk + \mathrm{i}tl)\mathbb{P}[d_n(1) = k, d_n(2) = l] \\
&\quad = \sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk + \mathrm{i}tl) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} g_n(\theta; a, b; k, l) F(\mathrm{d}a) F(\mathrm{d}b) \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a) F(\mathrm{d}b) \sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk + \mathrm{i}tl) g_n(\theta; a, b; k, l). \quad (33)
\end{aligned}$$

The term

$$\begin{aligned}
&\sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk + \mathrm{i}tl) g_n(\theta; a, b; k, l) \\
&\quad = \sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk + \mathrm{i}tl) \frac{n!}{l!(k-l)!(n-k)!} \\
&\qquad \times [1 - F(\theta - b)]^l [F(\theta - b) - F(\theta - a)]^{k-l} [F(\theta - a)]^{n-k} \\
&\quad = \sum_{k,l:0\leqslant l\leqslant k\leqslant n} \exp(\mathrm{i}s(k-l) + \mathrm{i}l(s+t)) \frac{n!}{l!(k-l)!(n-k)!} \\
&\qquad \times [1 - F(\theta - b)]^l [F(\theta - b) - F(\theta - a)]^{k-l} [F(\theta - a)]^{n-k} \\
&\qquad - \sum_{k,l:0\leqslant l=k\leqslant n} \exp(\mathrm{i}s(k-l) + \mathrm{i}l(s+t)) \frac{n!}{l!(k-l)!(n-k)!} \\
&\qquad \times [1 - F(\theta - b)]^l [F(\theta - b) - F(\theta - a)]^{k-l} [F(\theta - a)]^{n-k} \\
&\quad = [\exp(\mathrm{i}(s+t))(1 - F(\theta - b)) + \exp(\mathrm{i}s)(F(\theta - b) - F(\theta - a)) + F(\theta - a)]^n \\
&\qquad - \sum_{k=0}^{n} \exp(\mathrm{i}k(s+t)) \frac{n!}{k!(n-k)!} [1 - F(\theta - b)]^k [F(\theta - a)]^{n-k}
\end{aligned}$$

$$= [\exp(\mathrm{i}(s+t))(1-F(\theta-b)) + \exp(\mathrm{i}s)(F(\theta-b)-F(\theta-a)) + F(\theta-a)]^n$$
$$- [\exp(\mathrm{i}(s+t))(1-F(\theta-b)) + F(\theta-a)]^n. \tag{34}$$

Thus

$$\sum_{k,l:0\leqslant l<k\leqslant n} \exp(\mathrm{i}sk+\mathrm{i}tl)\mathbb{P}[d_n(1)=k, d_n(2)=l]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)([\exp(\mathrm{i}(s+t))(1-F(\theta-b))$$
$$+ \exp(\mathrm{i}s)(F(\theta-b)-F(\theta-a)) + F(\theta-a)]^n$$
$$- [\exp(\mathrm{i}(s+t))(1-F(\theta-b)) + F(\theta-a)]^n). \tag{35}$$

Similarly, for (ii),

$$\sum_{k,l:0\leqslant k<l\leqslant n} \exp(\mathrm{i}sk+\mathrm{i}tl)\mathbb{P}[d_n(1)=k, d_n(2)=l]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)([\exp(\mathrm{i}(s+t))(1-F(\theta-b))$$
$$+ \exp(\mathrm{i}t)(F(\theta-b)-F(\theta-a)) + F(\theta-a)]^n$$
$$- [\exp(\mathrm{i}(s+t))(1-F(\theta-b)) + F(\theta-a)]^n). \tag{36}$$

For (iii),

$$\sum_{k,l:0\leqslant k=l\leqslant n} \exp(\mathrm{i}sk+\mathrm{i}tk)\mathbb{P}[d_n(1)=k, d_n(2)=k]$$
$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)$$
$$\times \left(\sum_{k=0}^{n} \exp(\mathrm{i}sk+\mathrm{i}tk)\frac{n!}{k!(n-k)!}(1-F(\theta-b))^k(F(\theta-a))^{n-k}\right)$$
$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} F(\mathrm{d}a)F(\mathrm{d}b)$$
$$\times \left(\sum_{k=0}^{n} \exp(\mathrm{i}sk+\mathrm{i}tk)\frac{n!}{k!(n-k)!}(1-F(\theta-b))^k(F(\theta-a))^{n-k}\right)$$
$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)[(1-F(\theta-b))\exp(\mathrm{i}(s+t)) + F(\theta-a)]^n$$
$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} F(\mathrm{d}a)F(\mathrm{d}b)[(1-F(\theta-b))\exp(\mathrm{i}(s+t)) + F(\theta-a)]^n. \tag{37}$$

Combining the above and taking $s = u/n$ and $t = v/n$, we have

$$\mathbb{E}[\exp(\mathrm{i}u(d_n(1)/n) + \mathrm{i}v(d_n(2)/n))]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)$$
$$\times \left\{\left[\exp\left(\frac{\mathrm{i}(u+v)}{n}\right)(1-F(\theta-b)) + \exp\left(\frac{\mathrm{i}u}{n}\right)(F(\theta-b)-F(\theta-a)) + F(\theta-a)\right]^n\right.$$
$$\left. + \left[\exp\left(\frac{\mathrm{i}(u+v)}{n}\right)(1-F(\theta-b)) + \exp\left(\frac{\mathrm{i}v}{n}\right)(F(\theta-b)-F(\theta-a)) + F(\theta-a)\right]^n\right\}$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} F(\mathrm{d}a)F(\mathrm{d}b)\left[\exp\left(\frac{\mathrm{i}(u+v)}{n}\right)(1-F(\theta-b))+F(\theta-a)\right]^n$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)\left\{\left[\left(1+\frac{\mathrm{i}(u+v)}{n}+\frac{O(1)}{n^2}\right)(1-F(\theta-b))\right.\right.$$
$$\left.+\left(1+\frac{\mathrm{i}u}{n}+\frac{O(1)}{n^2}\right)(F(\theta-b)-F(\theta-a))+F(\theta-a)\right]^n$$
$$+\left[\left(1+\frac{\mathrm{i}(u+v)}{n}+\frac{O(1)}{n^2}\right)(1-F(\theta-b))\right.$$
$$\left.\left.+\left(1+\frac{\mathrm{i}v}{n}+\frac{O(1)}{n^2}\right)(F(\theta-b)-F(\theta-a))+F(\theta-a)\right]^n\right\}$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} F(\mathrm{d}a)F(\mathrm{d}b)\left[\left(1+\frac{\mathrm{i}(u+v)}{n}+\frac{O(1)}{n^2}\right)(1-F(\theta-b))+F(\theta-a)\right]^n. \tag{38}$$

Letting $n\to\infty$, we see that

$$\lim_{n\to\infty}\mathbb{E}[\exp(\mathrm{i}u(d_n(1)/n)+\mathrm{i}v(d_n(2)/n))]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a>b\}} F(\mathrm{d}a)F(\mathrm{d}b)\{\exp[\mathrm{i}(u+v)(1-F(\theta-b))$$
$$+\mathrm{i}u(F(\theta-b)-F(\theta-a))]+\exp[\mathrm{i}(u+v)(1-F(\theta-b))$$
$$+\mathrm{i}v(F(\theta-b)-F(\theta-a))]\}+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{\{a=b\}} F(\mathrm{d}a)F(\mathrm{d}b)$$
$$\times\exp[\mathrm{i}(u+v)(1-F(\theta-b))]$$
$$= \int_{-\infty}^{\infty}\exp(\mathrm{i}v(1-F(\theta-b)))F(\mathrm{d}b)\int_{-\infty}^{\infty}\exp(\mathrm{i}u(1-F(\theta-a)))F(\mathrm{d}a). \tag{39}$$

Thus combining with the result in section 2 we see that

$$\lim_{n\to\infty}\mathbb{E}[\exp(\mathrm{i}u(d_n(1)/n)+\mathrm{i}v(d_n(2)/n))]$$
$$= \mathbb{E}[\exp(\mathrm{i}u(1-F(\theta-X_1)))]\mathbb{E}[\exp(\mathrm{i}v(1-F(\theta-X_1)))]$$
$$= \lim_{n\to\infty}\mathbb{E}[\exp(\mathrm{i}u(D_n(1)/n))]\lim_{n\to\infty}\mathbb{E}[\exp(\mathrm{i}v(D_n(2)/n))], \tag{40}$$

i.e., we obtain the asymptotic independence as enunciated in theorem 2.

Now we obtain theorem 3. The joint conditional probability distribution of the weights of two fixed vertices, provided that these vertices are connected by an edge, is given by

$$H(\mathrm{d}a,\mathrm{d}b) = \frac{1(a+b>\theta)}{\alpha_F(\theta)}F(\mathrm{d}a)F(\mathrm{d}b) \tag{41}$$

where $\alpha_F(\theta) := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1(\tilde{a}+\tilde{b}>\theta)F(\mathrm{d}\tilde{a})F(\mathrm{d}\tilde{b})$ is the normalizing constant. It is actually the probability that two vertices share an edge. Consequently, the conditional probability distribution of the weight of a vertex provided that it shares an edge with another vertex is given by

$$G(da) = \int_{-\infty}^{\infty} H(\mathrm{d}a,\mathrm{d}b)$$
$$= F(\mathrm{d}a)\frac{\int_{-\infty}^{\infty} 1(a+b>\theta)F(\mathrm{d}b)}{\alpha_F(\theta)}$$
$$= \frac{1-F(\theta-a)}{\alpha_F(\theta)}F(\mathrm{d}a). \tag{42}$$

In order to study the asymptotic dependence of $d_n(1)/n$ and $d_n(2)/n$, we need to consider the difference between the expressions on the left- and right-hand sides of equation (39). Using equations (41) and (42), the asymptotic limit of this difference is

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[\exp(\mathrm{i}u(d_n(1)/n) + \mathrm{i}v(d_n(2)/n))] \\ - \lim_{n\to\infty} \mathbb{E}[\exp(\mathrm{i}u(D_n(1)/n))] \lim_{n\to\infty} \mathbb{E}[\exp(\mathrm{i}v(D_n(2)/n))] \\ = \int_{-\infty}^{\infty} (H(\mathrm{d}a, \mathrm{d}b) - G(\mathrm{d}a)G(\mathrm{d}b))\mathrm{e}^{\mathrm{i}u(1-F(\theta-a))}\mathrm{e}^{\mathrm{i}v(1-F(\theta-b))}. \end{aligned} \tag{43}$$

If the resulting conditional random variables are asymptotically independent, then for all $u, v \in \mathbb{R}$, the right-hand side of equation (43) must vanish. Now we claim that this will imply that the probability measures $H(\mathrm{d}a, \mathrm{d}b)$ and $G(\mathrm{d}a)G(\mathrm{d}b)$ on $\mathbb{R}^2$ are identical. Indeed, let (M_1, M_2) and (R_1, R_2) be two random vectors on $\mathbb{R}^2$ whose distributions are given by the probability measures $H(\mathrm{d}a, \mathrm{d}b)$ and $G(\mathrm{d}a)G(\mathrm{d}b)$, respectively. Consider the map $\tilde{f}$ from $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\tilde{f} : (a, b) \to (1 - F(\theta - a), 1 - F(\theta - b))$. Then, the supposition that equation (43) vanishes implies that the characteristic function of $\tilde{f}(M_1, M_2)$ is same as that of $\tilde{f}(R_1, R_2)$. Hence their distributions are also same. Therefore,

$$\begin{aligned}\mathbb{P}[M_1 \leqslant \beta_1, M_2 \leqslant \beta_2] \\ = \mathbb{P}[1 - F(\theta - M_1) \leqslant 1 - F(\theta - \beta_1), 1 - F(\theta - M_2) \leqslant 1 - F(\theta - \beta_2)] \\ = \mathbb{P}[1 - F(\theta - R_1) \leqslant 1 - F(\theta - \beta_1), 1 - F(\theta - R_2) \leqslant 1 - F(\theta - \beta_2)] \\ = \mathbb{P}[R_1 \leqslant \beta_1, R_2 \leqslant \beta_2], \end{aligned} \tag{44}$$

i.e., $H(\mathrm{d}a, \mathrm{d}b)$ and $G(\mathrm{d}a)G(\mathrm{d}b)$ are identical.

Now, we claim that, because of assumption 1 (of section 1), the probability measures $H(\mathrm{d}a, \mathrm{d}b)$ and $G(\mathrm{d}a)G(\mathrm{d}b)$ cannot be the same. Indeed, if the measures $H(\mathrm{d}a, \mathrm{d}b)$ and $G(\mathrm{d}a)G(\mathrm{d}b)$ are same, then for any subset $A \subseteq \mathbb{R}^2$, $\int_A H(\mathrm{d}a, \mathrm{d}b) = \int_A G(\mathrm{d}a)G(\mathrm{d}b)$. Using equations (41) and (42), we have

$$\int_A \left[\frac{1(a + b > \theta)}{\alpha_F(\theta)} - \frac{1 - F(\theta - a)}{\alpha_F(\theta)}\frac{1 - F(\theta - b)}{\alpha_F(\theta)}\right] F(\mathrm{d}a)F(\mathrm{d}b) = 0. \tag{45}$$

This will imply that

$$1(a + b > \theta)\alpha_F(\theta) - (1 - F(\theta - a))(1 - F(\theta - b)) = 0 \tag{46}$$

$F \times F$ almost surely.

Now fix u and v as in assumption 1 and let ϵ be such that $u + v - 4\epsilon > \theta$ and $u + 4\epsilon < \theta/2 < v - 4\epsilon$. Let $a, b \in (u - \epsilon, u + \epsilon]$. Then

$$1(a + b > \theta)\alpha_F(\theta) = 0. \tag{47}$$

However, $\theta - a < \theta - u + \epsilon < v - 3\epsilon < v - \epsilon$, so $1 - F(\theta - a) \geqslant 1 - F(v - \epsilon) \geqslant F(v + \epsilon) - F(v - \epsilon) > 0$. Similarly, $1 - F(\theta - b) > 0$. Thus

$$(1 - F(\theta - a))(1 - F(\theta - b)) > 0. \tag{48}$$

Hence, on a set of probability at least $(F(v + \epsilon) - F(v - \epsilon))^2 > 0$ we have

$$1(a + b > \theta)\alpha_F(\theta) - (1 - F(\theta - a))(1 - F(\theta - b)) < 0. \tag{49}$$

This contradiction completes the proof.

4. Triangles

Now we study the number of triangles in the graph. The number of triangles T_n can be represented as a U-statistic (see [11]). The kernel function $h : \mathbb{R}^3 \to \mathbb{R}$ defined in equation (4) is clearly a symmetric function of x_1, x_2, x_3. Then, we have

$$T_n = \binom{n}{3} \times \frac{1}{\binom{n}{3}} \sum_{(i,j,k)\in C} h(X_i, X_j, X_k) = \binom{n}{3} U_n \tag{50}$$

where $C := \{(i, j, k) : 1 \leqslant i < j < k \leqslant n\}$ is the collection of all possible triplets and $U_n := \frac{1}{\binom{n}{3}} \sum_{(i,j,k)\in C} h(X_i, X_j, X_k)$ is the U-statistic obtained from the kernel h. Theorem 4(a) can be easily derived from theorem A of [11], p 190.

Next define $h_1 : \mathbb{R} \to \mathbb{R}$ as follows:

$$\begin{aligned} h_1(x) &:= \mathbb{E}[h(x, X_2, X_3)] \\ &= \mathbb{P}[X_2 > \theta - x, X_3 > \theta - x, X_2 + X_3 > \theta] \\ &= \mathbb{P}[\min\{X_2, X_3\} > \theta - x, X_2 + X_3 > \theta]. \end{aligned} \tag{51}$$

Noting that

$$\mathbb{E}[h_1(X_1)] = \mathbb{E}[\mathbb{E}[h(X_1, X_2, X_3)|X_1]] = \mathbb{E}[h(X_1, X_2, X_3)] = F_3(\theta), \tag{52}$$

we have

$$\zeta_1(F) = \mathrm{Var}(h_1(X)) > 0 \tag{53}$$

unless F is degenerate. The asymptotic normality of $T_n/\binom{n}{3}$, i.e., theorem 4(b) follows from theorem A of [11], p 192.

As an aside, we note that the method of the U-statistics employed above is more versatile. It may be applied to configurations involving any subgraph comprising finitely many vertices.

Let $\mathbb{G}$ be a graph on k vertices ($k \leqslant n$) and suppose $\mathbb{V} := \{v_1, \ldots, v_k\}$ and $\mathbb{E}$ be the vertex and edge sets of the graph, respectively. For a permutation $\sigma : \mathbb{V} \to \mathbb{V}$, let us consider the graph $\mathbb{G}_\sigma$ with a vertex set $\mathbb{V}$ and an edge set $\mathbb{E}_\sigma := \{\langle\sigma(v_i), \sigma(v_j)\rangle : \langle v_i, v_j\rangle \in \mathbb{E}\}$. Let $l := \#\{\sigma : \mathbb{E}_\sigma = \mathbb{E}\}$. Thus l counts the number of symmetries of $\mathbb{G}$. To illustrate this, consider the graph with the vertex set as $\mathbb{V} = \{1, 2, 3, 4\}$ and the edge set $\mathbb{E} = \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 3, 4\rangle, \langle 4, 1\rangle\}$. For this graph $l = 8$; indeed any of the eight permutations (1, 2, 3, 4), (2, 3, 4, 1), (3, 4, 1, 2), (4, 1, 2, 3), (1, 4, 3, 2), (4, 3, 2, 1), (3, 2, 1, 4) and (2, 1, 4, 3) results in a graph which is equivalent to the original graph.

Now given a graph G on the vertex set $V := \{1, 2, \ldots, n\}$ with the edge set E consider the subgraph $G(i_1, \ldots, i_k)$ on the vertex set $V(i_1, \ldots, i_k) = \{i_1, \ldots, i_k\}$ and the edge set $E(i_1, \ldots, i_k) = \{\langle i_t, i_s\rangle : \langle i_t, i_s\rangle \in E \text{ and } 1 \leqslant t \neq s \leqslant k\}$. Let $\mathbb{L}_k$ be the set of all *ordered* k-tuples from $\{1, 2, \ldots, n\}$ and $f : \mathbb{L}_k \to \{0, 1\}$ be defined as follows:

$$f(i_1, i_2, \ldots, i_k) = \begin{cases} 1 & \text{if} \quad \langle i_s, i_t\rangle \in E(i_1, i_2, \ldots, i_k) \quad \text{for all} \quad \langle v_s, v_t\rangle \in \mathbb{E} \\ 0 & \text{otherwise.} \end{cases} \tag{54}$$

Thus, the number of subgraphs in G isomorphic to $\mathbb{G}$ is given by $l^{-1}T_n(\mathbb{G})$, where

$$T_n(\mathbb{G}) := \sum_{(i_1,i_2,\ldots,i_k)\in\mathbb{L}_k} f(i_1, i_2, \ldots, i_k). \tag{55}$$

For the graph G_θ, the vertices i_s and i_t are connected by an edge if and only if $X_{i_s} + X_{i_t} > \theta$. Thus the function f in the definition (55) may be replaced by the kernel function $h : \mathbb{R}^k \to \{0, 1\}$ defined by

$$h\big(X_{i_1}, X_{i_2}, \ldots, X_{i_k}\big) = \begin{cases} 1 & \text{if} \quad X_{i_s} + X_{i_t} > \theta \quad \text{for all} \quad \langle v_s, v_t\rangle \in \mathbb{E} \\ 0 & \text{otherwise,} \end{cases} \tag{56}$$

and we have

$$T_n(\mathbb{G}) = \sum_{(i_1,i_2,\dots,i_k)\in\mathbb{L}_k} h\big(X_{i_1}, X_{i_2}, \dots, X_{i_k}\big). \tag{57}$$

This function h need not be symmetric in its argument. Even if it were so, since the sum is over all ordered k-tuples $(i_1, \dots, i_k)$ and not on k-tuples $(i_1, \dots, i_k)$ such that $i_1 < i_2 < \cdots < i_k$, T_n need not be a U-statistic.

To overcome these, we first define a symmetrized version of the kernel h. Let

$$h_{\text{sym}}(x_1, x_2, \dots, x_k) := \frac{1}{k!} \sum_{\tilde\sigma\in P(k)} h(x_{\tilde\sigma(1)}, x_{\tilde\sigma(2)}, \dots, x_{\tilde\sigma(k)}) \tag{58}$$

where $\tilde\sigma : \{x_1, x_2, \dots, x_k\} \to \{x_1, x_2, \dots, x_k\}$ is a permutation, and $P(k)$ is a collection of all such permutations. We note that, even if $x_1 = x_2$, we take $(x_2, x_1, x_3, \dots, x_k)$ to be a permutation of $(x_1, x_2, x_3, \dots, x_k)$. Clearly, h_{sym} is a symmetric function of $x_1, x_2, \dots, x_k$, and we obtain

$$T_n(\mathbb{G}) = \sum_{(i_1,i_2,\dots,i_k)\in\mathbb{L}_k} h(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \sum_{(i_1,i_2,\dots,i_k)\in\mathbb{L}_k} h_{\text{sym}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}). \tag{59}$$

The statistic $T_n(\mathbb{G})$ is *not* a U-statistic. Thus we consider the statistic

$$\begin{aligned} T'_n(\mathbb{G}) &= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} h_{\text{sym}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) \\ &= T_n + \sum_{\substack{1\leqslant i_1,i_2,\dots,i_k\leqslant n \\ \text{not all distinct}}} h_{\text{sym}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) \\ &= T_n(\mathbb{G}) + R_n(\mathbb{G}) \qquad \text{(say)}. \end{aligned} \tag{60}$$

Since $T'_n(\mathbb{G})$ is a von Mises' statistic (see [10], p 39), the asymptotic results about $T'_n(\mathbb{G})$ can be read off from the results about the von Mises' statistics with the kernel function h_{sym}. To relate the statistics $T_n(\mathbb{G})$ and $T'_n(\mathbb{G})$ we observe that the number of terms in the sum defining R_n is of the order of n^{k-1}; thus noting that $h_{\text{sym}} \leqslant 1$, we have $R_n = O(n^{k-1})$ as $n \to \infty$. Let

$$F(\mathbb{G}) := \mathbb{E}[h(X_1, X_2, \dots, X_k)]. \tag{61}$$

Then, by the i.i.d. nature of $\{X_i : i \geqslant 1\}$, we have $\mathbb{E}[h_{\text{sym}}(X_1, X_2, \dots, X_k)] = F(\mathbb{G})$.

Theorem 8. *As $n \to \infty$,*

$$\frac{T_n(\mathbb{G})}{n^k} \to F(\mathbb{G}) \textit{ almost surely.} \tag{62}$$

Proof. From theorem 3.3.1 of [10], p 102, we have $\frac{T'_n(\mathbb{G})}{n^k} \to F(\mathbb{G})$ almost surely. Our observation that $R_n = O(n^{k-1})$ as $n \to \infty$ completes the proof of the theorem. □

To obtain the central limit theorem, as in (51) let

$$h_1(x) := \mathbb{E}[h_{\text{sym}}(x, X_2, X_3, \dots, X_k)] \tag{63}$$

and $\zeta_1(\mathbb{G}) = \text{Var}(h_1(X))$ where X is an independent random variable identical in distribution to X_1. Then, from theorem 4.2.5 ([10], p 134) we have the central limit theorem for $T'_n(\mathbb{G})$. Now $R_n(\mathbb{G})/n^k \to 0$ in probability as $n \to \infty$.

Thus we obtain

Theorem 9. *As $n \to \infty$,*

$$\sqrt{n}\left[\frac{T_n(\mathbb{G})}{n^k} - F(\mathbb{G})\right] \Rightarrow \sqrt{k\zeta_1(\mathbb{G})}Z. \tag{64}$$

5. Local properties

In this section, we study $T_n(1)$ as defined in equation (8).

For fixed $x \in \mathbb{R}$, the kernel $h(x, x_1, x_2)$ as defined in equation (4) is a symmetric function of x_1 and x_2. Define a U-statistic based on the kernel $h(x, \cdot, \cdot)$ by

$$T_n(1; x) := \sum_{2 \leqslant i \neq j \leqslant n+1} h(x, X_i, X_j). \tag{65}$$

We have by the strong law for U-statistics (theorem A, [11], p 190)

$$\frac{T_n(1; x)}{\binom{n}{2}} \to \mathbb{E}[h(x, X_i, X_j)] \qquad \text{almost surely, as} \quad n \to \infty. \tag{66}$$

The random variable $T_n(1; x)$ may be easily identified as the number of triangles of G_θ with a fixed vertex 1 and $X_1 = x$. Formally, we may write, for any $t \in \mathbb{R}$

$$\begin{aligned} \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{T_n(1)}{\binom{n}{2}}\right)\right] &= \int_{\mathbb{R}} F(\mathrm{d}x)\mathbb{E}\left[\exp\left(\mathrm{i}t\frac{T_n(1)}{\binom{n}{2}}\right)\Big| X_1 = x\right] \\ &\to \int_{\mathbb{R}} F(\mathrm{d}x)\exp(\mathrm{i}t\mathbb{E}[h(x, X_2, X_3)]) \end{aligned} \tag{67}$$

as $n \to \infty$, because $T_n(1; x)/\binom{n}{2} \to \mathbb{E}[h(x, X_2, X_3)]$ almost surely and hence also in distribution. The limit is justified by the usual application of the dominated convergence theorem since the integrand, being a characteristic function, is bounded by 1. The right-hand side of (67) is the characteristic function of the random variable $\int_{\mathbb{R}}\int_{\mathbb{R}} F(\mathrm{d}x_1)F(\mathrm{d}x_2)h(X, x_1, x_2)$. This proves theorem 5.

As in the previous section, we may generalize theorem 5 for the subgraph $\mathbb{G}$ defined earlier.

6. The spatial model

Consider the Poisson spatial model as elaborated in section 1. We first thin the underlying Poisson process. Fix $x \in \mathbb{R}$. For $i \geqslant 1$, each point ξ_i of the original process is included in the thinned process with a probability $f(|\xi_i|; x)$ independently of the other points, where $f(r; x)$ is as defined in equation (11). The thinned process is an inhomogeneous Poisson process with the intensity function $g(y)$ for $y \in \mathbb{R}$ given by $g(y) = \lambda f(|y|; x)$. For Δ_r as defined in equation (10), we have

Proposition 1. *The conditional distribution of Δ_r given that $X_0 = x$, is Poisson with the parameter $\lambda_r(x)$ where*

$$\lambda_r(x) := \lambda \int_{|\tilde{r}| \leqslant r} f(\tilde{r}; x)\,\mathrm{d}\tilde{r}. \tag{68}$$

We first prove theorem 6 where $C_r(x) \to C(x) < \infty$ (see section 1 for the relevant definitions). Because $t \to \int_{\mathbb{R}} F(\mathrm{d}x)\exp\big(-\lambda c_d C(x)(1 - \exp(\mathrm{i}t))\big)$ is indeed a characteristic function, it is enough to prove that the characteristic function of Δ_r converges to the above quantity. By proposition 1, the conditional distribution of Δ_r given $X_0 = x$ is Poisson with the parameter $\int_{|\tilde{r}| \leqslant r} \lambda f(\tilde{r}, x)\,\mathrm{d}\tilde{r} = \lambda c_d \int_0^r \tilde{r}^{d-1}[1 - F(\theta\tilde{r}^\beta - x)]\,\mathrm{d}\tilde{r} = \lambda c_d C_r(x)$ where c_d is the volume of the $(d-1)$-dimensional unit sphere. Therefore, we have

$$\begin{aligned}\phi_{\Delta_r}(t) &:= \mathbb{E}[\exp(\mathrm{i}t\Delta_r)] \\ &= \int_{\mathbb{R}} F(\mathrm{d}x)\mathbb{E}[\exp(\mathrm{i}t\Delta_r)|X_0 = x] \\ &= \int_{\mathbb{R}} F(\mathrm{d}x)\exp(-\lambda c_d C_r(x)(1-\exp(\mathrm{i}t))). \end{aligned} \tag{69}$$

Now, since $C_r(x) \to C(x)$, the usual dominated convergence theorem assures that

$$\phi_{\Delta_r}(t) \to \phi_{\Delta}(t) \qquad \text{as} \quad r \to \infty. \tag{70}$$

This completes the proof of theorem 6.

Remark. Assuming $\theta > 0$, we may rewrite $C(x)$ in the following way:

$$\begin{aligned}&\int_0^\infty r^{d-1}[1 - F(\theta r^\beta - x)]\,\mathrm{d}r \\ &\quad= \int_0^\infty r^{d-1}\int_{-\infty}^\infty 1(\tilde{x} > \theta r^\beta - x)F(\mathrm{d}\tilde{x})\,\mathrm{d}r \\ &\quad= \int_{-\infty}^\infty \int_0^\infty r^{d-1}1(r < \max\{0, [(\tilde{x}+x)/\theta]^{1/\beta}\})\,\mathrm{d}r\,F(\mathrm{d}\tilde{x}) \\ &\quad= \frac{1}{d}\int_{-\infty}^\infty [\max\{0, [(\tilde{x}+x)/\theta]^{d/\beta}\}]F(\mathrm{d}\tilde{x}) \\ &\quad= \frac{1}{\mathrm{d}\theta^{d/\beta}}\int_{-\infty}^\infty [\max\{0, (\tilde{x}+x)^{d/\beta}\}]F(\mathrm{d}\tilde{x}) \\ &\quad= \frac{1}{\mathrm{d}\theta^{d/\beta}}\mathbb{E}[\max\{0, (X_0+x)^{d/\beta}\}]. \end{aligned} \tag{71}$$

Thus $C(x) < \infty$ if $\mathbb{E}[|X_0|^{d/\beta}] < \infty$.

To show theorem 7, it is enough to prove that the characteristic function of $(\Delta_r - \lambda c_d C_r)/\sqrt{\lambda c_d C_r}$ converges to the product of the characteristic functions of a standard normal random variable and $\sqrt{\lambda c_d} g(X_0)$,

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\mathrm{i}t\frac{\Delta_r - \lambda c_d C_r}{\sqrt{\lambda c_d C_r}}\right)\right] \\ &\quad= \int_{\mathbb{R}} F(\mathrm{d}x)\mathbb{E}\left[\exp\left(\mathrm{i}t\frac{\Delta_r - \lambda c_d C_r}{\sqrt{\lambda c_d C_r}}\right)\middle| X_0 = x\right] \\ &\quad= \int_{\mathbb{R}} F(\mathrm{d}x)\exp(-\mathrm{i}t\sqrt{\lambda c_d C_r})\mathbb{E}\left[\exp\left(\mathrm{i}\frac{t}{\sqrt{\lambda c_d C_r}}\Delta_r\right)\middle| X_0 = x\right] \\ &\quad= \int_{\mathbb{R}} F(\mathrm{d}x)\exp(-\mathrm{i}t\sqrt{\lambda c_d C_r} - \lambda c_d C_r(x)(1-\exp(\mathrm{i}t/\sqrt{\lambda c_d C_r}))) \end{aligned} \tag{72}$$

using the fact that the conditional distribution of Δ_r given $X_0 = x$ follows a Poisson distribution with the parameter $\lambda c_d C_r(x)$. Consider the logarithm of the integrand:

$$\begin{aligned}&-\mathrm{i}t\sqrt{\lambda c_d C_r} - \lambda c_d C_r(x)\left[1 - \exp\left(\frac{\mathrm{i}t}{\sqrt{\lambda c_d C_r}}\right)\right] = -\mathrm{i}t\sqrt{\lambda c_d C_r} - \lambda c_d C_r(x) \\ &\qquad \times\left[-\frac{\mathrm{i}t}{\sqrt{\lambda c_d C_r}} - \frac{1}{2}\left(\frac{\mathrm{i}t}{\sqrt{\lambda c_d C_r}}\right)^2 + o\left(\left(\frac{\mathrm{i}t}{\sqrt{\lambda c_d C_r}}\right)^2\right)\right] \\ &\quad= \mathrm{i}t\sqrt{\lambda c_d}\left[\frac{C_r(x) - C_r}{\sqrt{C_r}}\right] - \frac{t^2}{2}\frac{C_r(x)}{C_r} - \lambda c_d C_r(x) o\left(\frac{1}{C_r}\right). \end{aligned} \tag{73}$$

Under our condition, the first term converges to $\mathrm{i}t\sqrt{\lambda c_d}g(x)$. The condition also implies that $C_r(x)/C_r \to 1$ as $r \to \infty$, and thus the second term converges to $-t^2/2$. The third term can be written as $[C_r(x)/C_r] \times [C_r o(1/C_r)] \to 0$ as $r \to \infty$. Applying the dominated convergence theorem, we now obtain that

$$\mathbb{E}\left[\exp\left(\mathrm{i}t\frac{\Delta_r - \lambda c_d C_r}{\sqrt{\lambda c_d C_r}}\right)\right] \to \exp\left(-\frac{t^2}{2}\right)\int_{\mathbb{R}} F(\mathrm{d}x)\exp(\mathrm{i}t\sqrt{\lambda c_d}g(x)). \quad (74)$$

Now note that $\exp(-t^2/2)$ is the characteristic function of a standard normal random variable and $\int_{\mathbb{R}} F(\mathrm{d}x)\exp(\mathrm{i}t\sqrt{\lambda c_d}g(x))$ is the characteristic function of $\sqrt{\lambda c_d}g(X_0)$. This completes the proof of theorem 7.

Acknowledgments

We thank Hiroyoshi Miwa for helpful discussions. NM is supported by Special Postdoctoral Researchers Program of RIKEN and RR is supported by DST grant no MS/152/01.

References

[1] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[2] Newman M E J 2003 *SIAM Rev.* **45** 167
[3] Caldarelli G, Capocci A, De Los Rios P and Muñoz M A 2002 *Phys. Rev. Lett.* **89** 258702
[4] Söderberg B 2002 *Phys. Rev.* E **66** 066121
[5] Boguña M and Pastor-Satorras R 2003 *Phys. Rev.* E **68** 036112
[6] Masuda N, Miwa H and Konno N 2004 *Phys. Rev.* E **70** 036124
[7] Masuda N, Miwa H and Konno N 2005 *Phys. Rev.* E **71** 036108
[8] Itzkovitz S, Milo R, Kashtan N, Ziv G and Alon U 2003 *Phys. Rev.* E **68** 026127
[9] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 *Science* **298** 824
[10] Koroljuk V S and Borovskich Yu V 1994 *Theory of U-Statistics* (Dordrecht: Kluwer)
[11] Serfling R J 1980 *Approximation Theorems of Mathematical Statistics* (New York: Wiley)