MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **93** (2025) 767–789
ISSN: 0340–6253
doi: 10.46793/match.93-3.767Y

Limiting Distribution for the Randić Index of a Random Geometric Graph

Mingao Yuan

[a] *Department of Statistics, North Dakota State University, Fargo, North Dakota, USA*

mingao.yuan@ndsu.edu

(Received February 22, 2024)

Abstract

The Randić index is a popular topological graph index that measures the extent of branching of a graph. It has many applications in chemistry and network data analysis. In this paper, we study the limiting distribution of the Randić index in a random geometric graph. We prove that the centered and scaled Randić index converges in law to an infinite sum of functions of independent chi-square random variables. It is interesting that the limiting distribution is not the standard normal distribution as in the Erdös-Rényi random graph case. However, the Randić index of the random geometric graph is asymptotically the same as the Erdös-Rényi random graph.

1 Introduction

A network or graph consists of a set of nodes or vertices and a set of edges. Edges in a graph represent interactions between nodes. Networks are widely used to understand many real-world problems [7, 21]. Networks can be employed to investigate the relationship between papers, authors, and scientific work [23]. In biology, network is used to detect gene-gene interactions [9]. In sociology, networks are used to model relationships among social actors and study dependence structures among social units [22].

The Randić index is a summary statistic that measures the extent of branching of a network [4, 24, 25]. It has been used to understand quantitative structure-property and structure-activity relations in chemistry and pharmocology [25, 26]. The Randić index also finds many applications in network data analysis. For example, it is used to measure how heterogeneous the degrees of the nodes are [14, 15]. It is also used to measure robustness of networks [11, 12] and quantify similarity of networks [15, 16]. Moreover, the Randić index has many interesting mathematical properties [5, 6, 8, 10, 18].

One of the important research topics is to study the Randić index in random graphs. Recently, [1, 19, 20] perform simulation studies of the Randić index in the Erdős-Rényi random graph and a random geometric graph. It is observed that the Randić index is approximately equal to one half of the number of nodes in the graph [1, 19, 20]. [2] conduct a simulation study of the relationship between the Randić index and the Shannon entropy in a random geometric graph. [28, 29] derives the limit and asymptotic distribution of the Randić index in a heterogeneous Erdős-Rényi random graph.

In this paper, we are interested in limiting distribution of the Randić index of a random geometric graph. In this random geometric graph, each node is independently assigned a position in the unit sphere and an edge exists between two nodes if and only if their distance is less than some pre-specified constant. The random geometric graph can capture the dependence structure and inherent geometric features of many real networks [13, 17]. Due to dependence of edges, it is more challenging to theoretically analyze the Randić index of the random geometric graph. We prove that the centered and scaled Randić index converges in distribution to an infinite sum of functions of independent chi-square random variables. This result is different from that in the Erdős-Rényi random graph, where the limiting distribution is the standard normal distribution [29]. Moreover, we show that the Randić index is asymptotically equal to one half of the number of nodes in the random geometric graph, which is the same as in the Erdős-Rényi random graph [28, 29]. In this sense, the Randić index itself cannot detect geometry in a network, but its limiting distribution can be used to detect geometry. This highlights the necessity to study asymptotic properties of the Randić index in random graphs.

The structure of the article is as follows. In Section 2 we present the main results. In Section 3, we present the proof.

Notation: We adopt the Bachmann–Landau notation throughout this paper. Let a_n and b_n be two positive sequences. Denote $a_n = \Theta(b_n)$ if $c_1 b_n \leq a_n \leq c_2 b_n$ for some positive constants c_1, c_2. Denote $a_n = \omega(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$. Denote $a_n = O(b_n)$ if $a_n \leq cb_n$ for some positive constants c. Denote $a_n = o(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$. Let X_n, X be random variables. Then $X_n \Rightarrow X$ means X_n converges in distribution to X as n goes to infinity. Denote $X_n = O_P(a_n)$ if $\frac{X_n}{a_n}$ is bounded in probability. Denote $X_n = o_P(a_n)$ if $\frac{X_n}{a_n}$ converges to zero in probability as n goes to infinity. Let $\mathbb{E}[X]$ and $Var(X)$ denote the expectation and variance of a random variable X respectively. $\mathbb{P}[E]$ denote the probability of an event E. $\exp(x)$ denote the exponential function e^x. For positive integer n, denote $[n] = \{1, 2, \ldots, n\}$. Given a finite set E, $|E|$ represents the number of elements in E, E^C represents the complement of the set E. For a positive integers i, j, k, $i \neq j \neq k$ means $i \neq j, j \neq k, k \neq i$. Given positive integer t, $\sum_{i_1 \neq i_2 \neq \ldots \neq i_t}$ means summation over all integers $i_1, i_2, \ldots, i_t$ in $[n]$ such that $|\{i_1, i_2, \ldots, i_t\}| = t$. $\sum_{i_1 < i_2 < \cdots < i_t}$ means summation over all integers $i_1, i_2, \ldots, i_t$ in $[n]$ such that $i_1 < i_2 < \cdots < i_t$. $I[E]$ is the indicator function of an event E, that is, $I[E] = 1$ if E occurs, $I[E] = 0$ otherwise.

2 Main result

Given a positive integer n, an *undirected* graph on $\mathcal{V} = [n]$ is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ is a set of edges. An edge $e \in \mathcal{E}$ is a subset of $\mathcal{V}$ such that $|e| = 2$. The elements in $\mathcal{V}$ are called nodes or vertices of the graph. A graph can be conveniently represented as an adjacency matrix A, where $A_{ij} = 1$ if $\{i, j\}$ is an edge, $A_{ij} = 0$ otherwise and $A_{ii} = 0$. Since graph $\mathcal{G}$ is undirected, the adjacency matrix A is symmetric. The degree d_i of vertex i is the number of edges that connect it, that is, $d_i = \sum_j A_{ij}$.

If $A_{ij}(1 \leq i < j \leq n)$ are random variables, the graph is said to be random. The most popular random graph is the well-known Erdős-Rényi random graph, where $A_{ij}(1 \leq i < j \leq n)$ are independent Bernoulli random variables with success probability p. Next we introduce a variant of the Erdős-Rényi random graph–the random geometric graph, which is relevant to the modelling of real networks with geometry and dependence structures [13, 17].

Definition 1. Let n, m be positive integers and r be a positive real number. Let $X = (X_1, \ldots, X_n)$ be a vector of independent and uniformly distributed random variables on the unit sphere $\mathbb{S}^m$. The random geometric graph $\mathcal{G}(n, m, r)$ is defined as follows:

$$A_{ij} = I[||X_i - X_j||_2 \leq r], \quad i < j$$

where $A_{ji} = A_{ij}$ for $1 \leq i < j \leq n$ and $A_{ii} = 0$ for $i = 1, 2, \ldots, n$.

In $\mathcal{G}(n, m, r)$, the vector X models the latent position of each node in the unit sphere. The presence of an edge between two nodes depends on their distance. If the distance between two nodes is less than r, then an edge exists between them. The parameter r models the sparsity of the random graph. Larger r produces a graph with more edges and smaller r leads to a graph with less edges. If $r \geq 2$, then the graph is the complete graph, that is, there is an edge between every pair of nodes. If $r = 0$, then there is no edge in the graph. Due to the random latent position vector X, $A_{ij}(1 \leq i < j \leq n)$ are not independent. It is therefore more difficult to study properties of $\mathcal{G}(n, m, r)$ than the Erdős-Rényi random graph. More general random geometric graphs can be found in [13, 17].

Definition 2. The Randić index of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as [24]

$$\mathcal{R}_n = \sum_{\{i,j\} \in \mathcal{E}} \frac{1}{\sqrt{d_i d_j}}, \tag{1}$$

The Randić index in random graphs has been widely studied [1, 3, 19, 20, 28, 29]. Especially, [28, 29] derives the limit and asymptotic distribution of the Randić index in a heterogeneous Erdős-Rényi random graph. It is shown that the scaled and centered Randić index converges in distribution to the standard normal distribution [28, 29].

In this paper, we study limiting distribution of the Randić index in the random geometric graph $\mathcal{G}(n, m, r)$.

Theorem 1. *For the random geometric graph $\mathcal{G}(n, m, r)$ with fixed m and constant $r \in (0, 1)$, we have*

$$16\mu^3 \left(\mathcal{R}_n - \mathbb{E}[\mathcal{R}_n]\right) \Rightarrow \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1),$$

as n goes to infinity, where

$$\mathbb{E}[\mathcal{R}_n] = \frac{n}{2}\left(1 + O\left(\frac{1}{n}\right)\right),$$

$\mu = \mathbb{P}(||X_1 - X_2||_2 \leq r)$, Z_j are independent standard normal random variables and λ_j are the eigenvalues of the kernel function $g(x_1, x_2)$ defined as

$$\begin{aligned} g(x_1, x_2) \quad = \quad & 2\mathbb{E}\Big[I[d(x_1, X_3) \leq r] I[d(X_3, X_4) \leq r] I[d(X_4, x_2) \leq r]\Big] \\ & -2\mu\mathbb{E}\Big[I[d(x_1, X_3) \leq r] I[d(X_3, x_2) \leq r]\Big]. \end{aligned}$$

Moreover,

$$\mathcal{R}_n = \frac{n}{2}\left(1 + o_P(1)\right).$$

Remark. Let F be the uniform distribution on $\mathbb{S}^m$ and $L_2(\mathbb{S}^m, F)$ be the space of square-integrable functions. For a symmetric function $f(x_1, x_2) \in L_2(\mathbb{S}^m, F)$, define an operator T as

$$T(g)(x) = \int_{\mathbb{S}^m} f(x, y) g(y) dF(y), \quad g \in L_2(\mathbb{S}^m, F).$$

The eigen-vectors g_i and eigen-values λ_i of T is defined as

$$T(g_i) = \lambda_i g_i.$$

According to Theorem 1, the centered and scaled Randić index of the random geometric graph $\mathcal{G}(n, m, r)$ converges in distribution to an infinite sum of functions of independent chi-square random variables. However, in the Erdős-Rényi random graph, the centered and scaled Randić index converges in distribution to the standard normal distribution [29]. This result signifies the difference between the random geometric graph $\mathcal{G}(n, m, r)$ and the Erdős-Rényi random graph.

Moreover, the Randić index of the random geometric graph $\mathcal{G}(n, m, r)$ is asymptotically equal to $\frac{n}{2}$, that is,

$$\mathcal{R}_n = \frac{n}{2}\left(1 + o_P\left(1\right)\right).$$

This result theoretically confirms the empirical observation obtained in [1]. Note that the Randić index in the Erdős-Rényi random graph is also asymptotically equal to $\frac{n}{2}$ [28, 29]. In this sense, the Randić index itself cannot distinguish the random geometric graph $\mathcal{G}(n, m, r)$ from the Erdős-Rényi random graph. However, its limiting distribution can detect the geometry in a network. This highlights the necessity to study limiting distribution of the Randić index in random graphs.

Due to the dependence of edges in the random geometric graph $\mathcal{G}(n, m, r)$, the proof of Theorem 1 is more challenging than in the Erdős-Rényi random graph case. Our proof strategy is to express $\mathcal{R}_n$ as a sum of leading term and reminder term, followed by showing the leading term is a degenerate U-statistic and the reminder term is negligible. Since the limiting distribution of degenerate U-statistic is known [27], then the proof is complete.

3 Proof

For given indices $i < j$, denote $d_{i(j)} = 1 + \sum_{l \notin \{i,j\}} A_{il}$ and $d_{j(i)} = 1 + \sum_{l \notin \{i,j\}} A_{jl}$. For convenience, we still write $d_{i(j)}$ as d_i. Then the Randić index of a graph $\mathcal{G}$ can be written as

$$\mathcal{R}_n = \sum_{i<j} \frac{A_{ij}}{\sqrt{d_i d_j}}. \tag{2}$$

Before prove Theorem 1, we provide a lemma first.

Lemma 1. *Let* $a_n = \log n$ *and* $\epsilon_n = (\log n)^{-3}$*. For the random geometric graph* $\mathcal{G}(n, m, r)$*, we have*

$$\mathbb{P}(d_1 d_2 < \epsilon_n n^2 \mu^2) = e^{-\frac{n\mu}{a_n}(1+o(1))}.$$

Proof of Lemma 1. Recall that we denote $d_{i(j)} = 1 + \sum_{l \notin \{i,j\}} A_{il}$ and $d_{j(i)} = 1 + \sum_{l \notin \{i,j\}} A_{jl}$, and we still write $d_{i(j)}$ as d_i for convenience. Note that

$$\left(\sum_{l \notin \{i,j\}} A_{il}\right)\left(\sum_{l \notin \{i,j\}} A_{jl}\right) = \sum_{k,l \notin \{i,j\}, k \neq l} A_{ik} A_{jl} + \sum_{k \notin \{i,j\}} A_{ik} A_{jk}.$$

Simple algebra yields

$$\begin{aligned} d_i d_j &= (1 + \sum_{k \notin \{i,j\}} A_{ik})(1 + \sum_{k \notin \{i,j\}} A_{jk}) \\ &= 1 + \sum_{k \notin \{i,j\}} A_{ik} + \sum_{k \notin \{i,j\}} A_{jk} \\ &\quad + \sum_{k,l \notin \{i,j\}, k \neq l} A_{ik} A_{jl} + \sum_{k \notin \{i,j\}} A_{ik} A_{jk}. \end{aligned} \tag{3}$$

Then

$$\begin{aligned} \mathbb{P}(d_1 d_2 < \epsilon_n n^2 \mu^2) &\leq \mathbb{P}\left(\sum_{k,l \notin \{1,2\}, k \neq l} A_{1k} A_{2l} < \epsilon_n n^2 \mu^2\right) \\ &\leq \mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}} \sum_{l=\frac{n}{a_n}+1}^{n} A_{1k} A_{2l} < \epsilon_n n^2 \mu^2\right). \end{aligned} \tag{4}$$

We claim that A_{1k} $(3 \leq k \leq n)$ are independent. To see this, we use the moment generating function to prove that A_{13}, A_{14} are independent. Since $X_1, X_2, \ldots, X_n$ are independent and uniformly distributed on sphere $\mathbb{S}^m$, then

$$\mathbb{P}\left(||X_1 - X_i||_2 \leq r \middle| X_1\right) = \mu, \quad i \in \{2, 3, 4, \ldots, n\}.$$

The joint moment generating function of A_{13}, A_{14} is equal to

$$\begin{aligned}
& \mathbb{E}\left[e^{t_1 A_{13}+t_2 A_{14}}\right] \\
= & \mathbb{E}\Big[\mathbb{E}\left[e^{t_1 I[||X_1-X_3||_2\le r]}\big|X_1\right]\mathbb{E}\left[e^{t_2 I[||X_1-X_4||_2\le r]}\big|X_1\right]\Big] \\
= & \mathbb{E}\Big[(e^{t_1}\mu+1-\mu)[(e^{t_2}\mu+1-\mu)\Big] \\
= & \mathbb{E}\left[e^{t_1 A_{13}}\right]\mathbb{E}\left[e^{t_2 A_{14}}\right], \qquad t_1, t_2 \in \mathbb{R}. \qquad (5)
\end{aligned}$$

Hence A_{13}, A_{14} are independent. Similarly, we can prove A_{1k} $(3 \le k \le n)$ are independent.

Moreover, A_{1k} $(3 \le k \le n/a_n)$ are independent of A_{2l} $(n/a_n + 1 \le l \le n)$. Let S be a subset of $\{3, 4, \dots, \frac{n}{a_n}\}$ and $\delta_n = (\log n)^{-1}$. Denote $A_{1S} = \{A_{1k}|k \in S\}$. Then

$$\begin{aligned}
& \mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}}\sum_{l=\frac{n}{a_n}+1}^{n} A_{1k}A_{2l} < \epsilon_n n^2\mu^2\right) \\
= & \sum_{t=1}^{\frac{n\delta_n}{a_n}}\sum_{|S|=t}\mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}}\sum_{l=\frac{n}{a_n}+1}^{n} A_{1k}A_{2l} < \epsilon_n n^2\mu^2\Big|A_{1S}=1, A_{1S^C}=0\right) \\
& \times\mathbb{P}(A_{1S}=1, A_{1S^C}=0) \\
& + \sum_{t=\frac{n\delta_n}{a_n}+1}^{\frac{n}{a_n}}\sum_{|S|=t}\mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}}\sum_{l=\frac{n}{a_n}+1}^{n} A_{1k}A_{2l} < \epsilon_n n^2\mu^2\Big|A_{1S}=1, A_{1S^C}=0\right) \\
& \times\mathbb{P}(A_{1S}=1, A_{1S^C}=0) \\
\le & \sum_{t=1}^{\frac{n\delta_n}{a_n}}\sum_{|S|=t}\mathbb{P}(A_{1S}=1, A_{1S^C}=0) \\
& + \sum_{t=\frac{n\delta_n}{a_n}+1}^{\frac{n}{a_n}}\sum_{|S|=t}\mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}}\sum_{l=\frac{n}{a_n}+1}^{n} A_{1k}A_{2l} < \epsilon_n n^2\mu^2\Big|A_{1S}=1, A_{1S^C}=0\right). \qquad (6)
\end{aligned}$$

Note that for $t \le \frac{n}{a_n}$, we have

$$\sum_{|S|=t}\mathbb{P}(A_{1S}=1, A_{1S^C}=0) = \binom{\frac{n}{a_n}}{t}\mu^t(1-\mu)^{\frac{n}{a_n}-t} \le e^{f(t)},$$

where

$$f(t) = t\log\frac{n}{a_n} - t\log t + t + t\log\mu + \left(\frac{n}{a_n} - t\right)\log(1-\mu).$$

The derivative of $f(t)$ is equal to

$$f'(t) = \log\frac{n\mu}{a_n(1-\mu)} - \log t.$$

Recall that $f(t)$ is increasing if $f'(t) \geq 0$, and $f(t)$ is decreasing if $f'(t) \leq 0$. When $t \leq \frac{n\mu}{a_n(1-\mu)}$, it is easy to verify that $f'(t) \geq 0$. Hence $f(t)$ is increasing if $t \leq \frac{n\mu}{a_n(1-\mu)}$. Then for $t \leq \frac{n\delta_n}{a_n} < \frac{n\mu}{a_n(1-\mu)}$, we have

$$f(t) \leq f\left(\frac{n\delta_n}{a_n}\right) \leq \frac{n}{a_n}\delta_n \log\frac{\mu}{\delta_n(1-\mu)} + \frac{n\delta_n}{a_n} - \frac{n}{a_n}\mu = -\frac{n}{a_n}\mu(1+o(1)).$$

Hence

$$\sum_{t=1}^{\frac{n\delta_n}{a_n}} \sum_{|S|=t} \mathbb{P}(A_{1S} = 1, A_{1S^C} = 0) \leq e^{-\frac{n\mu}{a_n}(1+o(1))}. \tag{7}$$

On the other hand, given $\frac{n\delta_n}{a_n} \leq t \leq \frac{n}{a_n}$ and $|S| = t$, we have

$$\begin{aligned} & \mathbb{P}\left(\sum_{k=3}^{\frac{n}{a_n}} \sum_{l=\frac{n}{a_n}+1}^{n} A_{1k}A_{2l} < \epsilon_n n^2\mu^2 \middle| A_{1S} = 1, A_{1S^C} = 0\right) \\ = \ & \mathbb{P}\left(\sum_{l=\frac{n}{a_n}+1}^{n} A_{2l} < \frac{\epsilon_n n^2\mu^2}{t}\right) \\ \leq \ & \mathbb{P}\left(\sum_{l=\frac{n}{a_n}+1}^{n} A_{2l} < \frac{\epsilon_n n a_n \mu^2}{\delta_n}\right). \end{aligned} \tag{8}$$

Let k be a non-negative integer less than $\frac{\epsilon_n n a_n \mu^2}{\delta_n}$. Then

$$\mathbb{P}\left(\sum_{l=\frac{n}{a_n}+1}^{n} A_{2l} = k\right) = \binom{n(1-\frac{1}{a_n})}{k}\mu^k(1-\mu)^{n(1-\frac{1}{a_n})-k} \leq e^{g(k)},$$

where

$$
\begin{aligned}
g(k) &= k\log\left(n\left(1-\frac{1}{a_n}\right)\right) - k\log k + k + k\log\mu \\
&\quad + \left(n\left(1-\frac{1}{a_n}\right) - k\right)\log(1-\mu).
\end{aligned}
$$

The derivative of $g(k)$ is equal to

$$
g'(k) = \log\frac{n\left(1-\frac{1}{a_n}\right)\mu}{1-\mu} - \log k.
$$

Then $g(k)$ is increasing if $k \leq \frac{n\left(1-\frac{1}{a_n}\right)\mu}{1-\mu}$. For $k \leq \frac{\epsilon_n n a_n \mu^2}{\delta_n} < \frac{n\left(1-\frac{1}{a_n}\right)\mu}{1-\mu}$, we have

$$
\begin{aligned}
g(k) &\leq g\left(\frac{\epsilon_n n a_n \mu^2}{\delta_n}\right) \\
&\leq \frac{\epsilon_n n a_n \mu^2}{\delta_n}\log\frac{1}{\mu(1-\mu)^{\frac{\epsilon_n a_n}{\delta_n}}} + \frac{\epsilon_n n a_n \mu^2}{\delta_n} - n\mu\left(1-\frac{1}{a_n}\right) \\
&= -n\mu(1+o(1)).
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbb{P}\left(\sum_{l=\frac{n}{a_n}+1}^{n} A_{2l} = k\right) &= \binom{n(1-\frac{1}{a_n})}{k}\mu^k(1-\mu)^{n(1-\frac{1}{a_n})-k} \\
&\leq e^{-n\mu(1+o(1))}.
\end{aligned}
$$

Hence,

$$
\mathbb{P}\left(\sum_{l=\frac{n}{a_n}+1}^{n} A_{2l} < \frac{\epsilon_n n a_n \mu^2}{\delta_n}\right) \leq \frac{\epsilon_n n a_n \mu^2}{\delta_n} e^{-n\mu(1+o(1))} = e^{-n\mu(1+o(1))}. \quad (9)
$$

By (4), (6)-(9), the proof is complete.

■

Proof of Theorem 1: For convenience, we denote $\bar{A}_{ij} = A_{ij} - \mathbb{E}[A_{ij}]$, $\mu = \mathbb{E}[A_{ij}]$ and let $b_n = \mathbb{E}[d_i] = 1 + (n-2)\mu$. Given $i < j$, denote

$\bar{d}_i = \sum_{k\notin\{i,j\}} \bar{A}_{ij}$. By (3), we have

$$\begin{aligned}\mathbb{E}[d_i d_j] &= 1 + 2(n-2)\mu + (n-2)(n-3)\mu^2 + (n-2)\mu^2 \\ &= 1 + 2(n-2)\mu + (n-2)^2\mu^2 \\ &= \mathbb{E}[d_i]\mathbb{E}[d_j]. \end{aligned} \tag{10}$$

By (10), straightforward calculation yields

$$\begin{aligned} d_i d_j - \mathbb{E}[d_i d_j] &= d_i d_j - \mathbb{E}[d_i]\mathbb{E}[d_j] \\ &= \bar{d}_i \bar{d}_j + \bar{d}_i \mathbb{E}[d_j] + \mathbb{E}[d_i]\bar{d}_j. \end{aligned} \tag{11}$$

By Taylor expansion, we have

$$\begin{aligned} \mathcal{R}_n &= \sum_{i<j} \frac{A_{ij}}{\sqrt{\mathbb{E}[d_i d_j]}} - \frac{1}{2}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])}{\big(\mathbb{E}[d_i d_j]\big)^{\frac{3}{2}}} \\ &\quad + \frac{3}{8}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^2}{\big(\mathbb{E}[d_i d_j]\big)^{\frac{5}{2}}} \\ &\quad - \frac{5}{16}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^3}{Z_{ij}^{\frac{7}{2}}}, \end{aligned} \tag{12}$$

where Z_{ij} is between $d_i d_j$ and $\mathbb{E}[d_i d_j]$. Then

$$\begin{aligned} &\mathcal{R}_n - \mathbb{E}[\mathcal{R}_n] \\ = &\sum_{i<j} \frac{\bar{A}_{ij}}{\sqrt{\mathbb{E}[d_i d_j]}} - \frac{1}{2}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j]) - \mathbb{E}[A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])]}{\big(\mathbb{E}[d_i d_j]\big)^{\frac{3}{2}}} \\ &+ \frac{3}{8}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^2 - \mathbb{E}[A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^2]}{\big(\mathbb{E}[d_i d_j]\big)^{\frac{5}{2}}} \\ &- \frac{5}{16}\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^3}{Z_{ij}^{\frac{7}{2}}} \\ &+ \frac{5}{16}\mathbb{E}\Big[\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^3}{Z_{ij}^{\frac{7}{2}}}\Big], \end{aligned} \tag{13}$$

Next we isolate the leading terms in (13).

Consider the last two terms of (13) first. Let $\epsilon_n = (\log n)^{-3}$ as in Lemma 1. It is easy to get

$$
\begin{aligned}
& \mathbb{E}\left[\left|\sum_{i<j}\frac{A_{ij}(d_id_j-\mathbb{E}[d_id_j])^3}{Z_{ij}^{\frac{7}{2}}}\right|\right] \\
\le\ & \mathbb{E}\left[\sum_{i<j}\frac{A_{ij}|d_id_j-\mathbb{E}[d_id_j]|^3}{Z_{ij}^{\frac{7}{2}}}I[Z_{ij}\ge \epsilon_n n^2]\right] \\
& +\mathbb{E}\left[\sum_{i<j}\frac{A_{ij}|d_id_j-\mathbb{E}[d_id_j]|^3}{Z_{ij}^{\frac{7}{2}}}I[Z_{ij}< \epsilon_n n^2]\right]. \qquad (14)
\end{aligned}
$$

By a similar argument as in Lemma 3.2 of [30], we have $\mathbb{E}[\bar{d}_i^{2s}] = O(n^s)$ for positive integer s. By (11) and the Cauchy-Schwarz inequality, one has

$$
\begin{aligned}
\mathbb{E}\left[\left|d_id_j-\mathbb{E}[d_id_j]\right|^3\right] &\le 3^3\left(\mathbb{E}[|\bar{d}_i^3\bar{d}_j^3|]+\mathbb{E}[|\bar{d}_i^3|]b_n^3+\mathbb{E}[|\bar{d}_j^3|]b_n^3\right) \\
&\le 3^3\left(\sqrt{\mathbb{E}[\bar{d}_i^6]\mathbb{E}[\bar{d}_j^6]}+\sqrt{\mathbb{E}[\bar{d}_i^6]}b_n^3+\sqrt{\mathbb{E}[\bar{d}_j^6]}b_n^3\right) \\
&= O\left(n^3\sqrt{n^3}\right).
\end{aligned}
$$

Hence

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i<j}\frac{A_{ij}|d_id_j-\mathbb{E}[d_id_j]|^3}{Z_{ij}^{\frac{7}{2}}}I[Z_{ij}\ge \epsilon_n n^2]\right] &\le \sum_{i<j}\frac{\mathbb{E}\left[|d_id_j-\mathbb{E}[d_id_j]|^3\right]}{\epsilon_n^{\frac{7}{2}}n^7} \\
&= O\left(\frac{1}{\sqrt{n}\epsilon_n^{\frac{7}{2}}}\right). \qquad (15)
\end{aligned}
$$

Suppose $Z_{ij} < \epsilon_n n^2$. Note that $\epsilon_n n^2 = o(n^2)$ and $\mathbb{E}[d_id_j] = \Theta(n^2)$. If $Z_{ij} < d_id_j$, then Z_{ij} cannot be between d_id_j and $\mathbb{E}[d_id_j]$. Hence $Z_{ij} \ge d_id_j$. Moreover, $d_id_j \ge 1$. Then by Lemma 1, we get

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{i<j}\frac{A_{ij}|d_id_j-\mathbb{E}[d_id_j]|^3}{Z_{ij}^{\frac{7}{2}}}I[Z_{ij}< \epsilon_n n^2]\right] \\
\le\ & \mathbb{E}\left[\sum_{i<j}A_{ij}|d_id_j-\mathbb{E}[d_id_j]|^3 I[d_id_j< \epsilon_n n^2]\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad n^5 \max_{i<j} \mathbb{P}(d_i d_j < \epsilon_n n^2) \\
&= \quad e^{-\frac{n\mu}{\log n}(1+o(1))}. \qquad (16)
\end{aligned}
$$

Combining (14), (15) and (16) yields

$$
-\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^3}{Z_{ij}^{\frac{7}{2}}} + \mathbb{E}\Big[\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])^3}{Z_{ij}^{\frac{7}{2}}}\Big] = o_P(1). \quad (17)
$$

Now consider the first two terms of (13). By (10) and (11), we have

$$
\sum_{i<j} \frac{A_{ij}(d_i d_j - \mathbb{E}[d_i d_j])}{\big(\mathbb{E}[d_i d_j]\big)^{\frac{3}{2}}} = \sum_{i<j} \frac{A_{ij}\bar{d}_i \bar{d}_j}{b_n^3} + \sum_{i<j} \frac{A_{ij}(\bar{d}_i + \bar{d}_j)}{b_n^2}. \qquad (18)
$$

Since A_{ij} is independent of $\bar{d}_i$ given X_i, then

$$
\mathbb{E}[A_{ij}\bar{d}_i] = \mathbb{E}\big[\mathbb{E}[A_{ij}|X_i]\mathbb{E}[\bar{d}_i|X_i]\big] = 0
$$

and

$$
\mathbb{E}\left[\sum_{i<j} \frac{A_{ij}(\bar{d}_i + \bar{d}_j)}{b_n^2}\right] = 0.
$$

Moreover, simple algebra yields

$$
\sum_{i<j} \frac{A_{ij}(\bar{d}_i + \bar{d}_j)}{b_n^2} = \sum_{i<j} \frac{\bar{A}_{ij}(\bar{d}_i + \bar{d}_j)}{b_n^2} + \sum_{i<j} \frac{\mu(\bar{d}_i + \bar{d}_j)}{b_n^2}. \qquad (19)
$$

It is easy to verify that

$$
\sum_{i<j} \frac{\mu(\bar{d}_i + \bar{d}_j)}{b_n^2} = \frac{2(n-2)\mu}{b_n} \sum_{i<j} \frac{\bar{A}_{ij}}{b_n}. \qquad (20)
$$

Note that

$$
\mathbb{E}\left[\left(\sum_{i<j} \frac{\bar{A}_{ij}}{b_n}\right)^2\right] = O(1).
$$

Then

$$\sum_{i<j}\frac{\bar{A}_{ij}}{b_n}-\frac{(n-2)\mu}{b_n}\sum_{i<j}\frac{\bar{A}_{ij}}{b_n}=\frac{1}{b_n}\sum_{i<j}\frac{\bar{A}_{ij}}{b_n}=O_P\left(\frac{1}{n}\right). \tag{21}$$

The first term of (19) is equal to

$$\sum_{i<j}\frac{\bar{A}_{ij}(\bar{d}_i+\bar{d}_j)}{b_n^2}=\frac{2}{b_n^2}\sum_{i<j<k}(\bar{A}_{ij}\bar{A}_{ik}+\bar{A}_{ji}\bar{A}_{jk}+\bar{A}_{ki}\bar{A}_{kj}). \tag{22}$$

The first term of (18) can be written as

$$\begin{aligned}\sum_{i<j}\frac{A_{ij}\bar{d}_i\bar{d}_j}{b_n^3} &= \frac{1}{2b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}+\frac{1}{2b_n^3}\sum_{i\neq j\neq k}\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}\\ &\quad+\frac{\mu}{2b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ik}\bar{A}_{jl}+\frac{\mu}{2b_n^3}\sum_{i\neq j\neq k}\bar{A}_{ik}\bar{A}_{jk}.\end{aligned}$$

Then

$$\begin{aligned}&\sum_{i<j}\frac{A_{ij}\bar{d}_i\bar{d}_j}{b_n^3}-\mathbb{E}\left[\sum_{i<j}\frac{A_{ij}\bar{d}_i\bar{d}_j}{b_n^3}\right]\\ =\ &\frac{1}{2b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}+\frac{1}{2b_n^3}\sum_{i\neq j\neq k}\left(\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}-\mathbb{E}[\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}]\right)\\ &+\frac{\mu}{2b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ik}\bar{A}_{jl}+\frac{\mu}{2b_n^3}\sum_{i\neq j\neq k}\bar{A}_{ik}\bar{A}_{jk}.\end{aligned} \tag{23}$$

We show the last three terms of (23) are $o_P(1)$. Note that

$$\mathbb{E}\left[\left(\frac{1}{b_n^3}\sum_{i\neq j\neq k}\bar{A}_{ik}\bar{A}_{jk}\right)^2\right]=\frac{1}{b_n^6}\sum_{\substack{i\neq j\neq k\\ i_1\neq j_1\neq k_1}}\mathbb{E}\left[\bar{A}_{ik}\bar{A}_{jk}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}\right].$$

If $i\notin\{i_1,j_1,k_1\}$, then X_i is independent of $X_k,X_j,X_{i_1},X_{j_1},X_{k_1}$. Recall that $\mathbb{E}\left[\bar{A}_{ik}|X_k\right]=0$. Then

$$\mathbb{E}\left[\bar{A}_{ik}\bar{A}_{jk}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}\right]=\mathbb{E}\Big[\mathbb{E}\left[\bar{A}_{ik}|X_k\right]\bar{A}_{jk}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}\Big]=0.$$

Hence $i \in \{i_1, j_1, k_1\}$. Similarly, $j \in \{i_1, j_1, k_1\}$. Since $|\bar{A}_{ik}\bar{A}_{jk}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}| \leq 1$, then

$$\mathbb{E}\left[\left(\frac{1}{b_n^3}\sum_{i\neq j\neq k}\bar{A}_{ik}\bar{A}_{jk}\right)^2\right] = O\left(\frac{n^4}{b_n^6}\right) = O\left(\frac{1}{n^2}\right). \tag{24}$$

Similarly, it is easy to get

$$\mathbb{E}\left[\left(\frac{1}{b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ik}\bar{A}_{jl}\right)^2\right] = O\left(\frac{1}{n^2}\right). \tag{25}$$

The second moment of the second term of (23) is equal to

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{b_n^3}\sum_{i\neq j\neq k}\left(\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk} - \mathbb{E}[\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}]\right)\right)^2\right] \\
= \;& \frac{1}{b_n^6}\sum_{\substack{i\neq j\neq k\\ i_1\neq j_1\neq k_1}}\mathbb{E}\Big[\left(\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk} - \mathbb{E}[\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}]\right) \\
&\times\left(\bar{A}_{i_1j_1}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1} - \mathbb{E}[\bar{A}_{i_1j_1}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}]\right)\Big].
\end{aligned}$$

If $\{i, j, k\} \cap \{i_1, j_1, k_1\} = \emptyset$, then $\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}$ and $\bar{A}_{i_1j_1}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}$ are independent. Hence

$$\begin{aligned}
&\mathbb{E}\Big[\left(\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk} - \mathbb{E}[\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}]\right) \\
&\times\left(\bar{A}_{i_1j_1}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1} - \mathbb{E}[\bar{A}_{i_1j_1}\bar{A}_{i_1k_1}\bar{A}_{j_1k_1}]\right)\Big] = 0.
\end{aligned}$$

Then $|\{i, j, k, i_1, j_1, k_1\}| \leq 5$ and

$$\mathbb{E}\left[\left(\frac{1}{b_n^3}\sum_{i\neq j\neq k}\left(\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk} - \mathbb{E}[\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}]\right)\right)^2\right] = O\left(\frac{1}{n}\right). \tag{26}$$

Based on (18)-(26), we get

$$\begin{aligned}
&\sum_{i<j}\frac{A_{ij}}{\sqrt{\mathbb{E}[d_id_j]}}-\frac{1}{2}\sum_{i<j}\frac{A_{ij}(d_id_j-\mathbb{E}[d_id_j])}{\left(\mathbb{E}[d_id_j]\right)^{\frac{3}{2}}}\\
= & -\frac{1}{b_n^2}\sum_{i<j<k}(\bar{A}_{ij}\bar{A}_{ik}+\bar{A}_{ji}\bar{A}_{jk}+\bar{A}_{ki}\bar{A}_{kj})-\frac{1}{4b_n^3}\sum_{i\neq j\neq k\neq l}\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}\\
&+O_P\left(\frac{1}{\sqrt{n}}\right). \qquad (27)
\end{aligned}$$

Now consider the third term of (13). By (10) and (11), we have

$$\begin{aligned}
\sum_{i<j}\frac{A_{ij}(d_id_j-\mathbb{E}[d_id_j])^2}{\left(\mathbb{E}[d_id_j]\right)^{\frac{5}{2}}} = & \sum_{i<j}\frac{A_{ij}\bar{d}_i^2\bar{d}_j^2}{b_n^5}+\sum_{i<j}\frac{2A_{ij}\bar{d}_i\bar{d}_j(\bar{d}_i+\bar{d}_j)b_n}{b_n^5}\\
&+\sum_{i<j}\frac{A_{ij}(\bar{d}_i+\bar{d}_j)^2b_n^2}{b_n^5}. \qquad (28)
\end{aligned}$$

Note that $0\leq A_{ij}\leq 1$. By the Cauchy-Schwarz inequality, one has

$$\mathbb{E}\left[\sum_{i<j}\frac{A_{ij}\bar{d}_i^2\bar{d}_j^2}{b_n^5}\right]\leq\sum_{i<j}\frac{\sqrt{\mathbb{E}[\bar{d}_i^4]\mathbb{E}[\bar{d}_j^4]}}{b_n^5}=O\left(\frac{1}{n}\right), \qquad (29)$$

$$\mathbb{E}\left[\sum_{i<j}\left|\frac{A_{ij}\bar{d}_i^2\bar{d}_jb_n}{b_n^5}\right|\right]\leq\sum_{i<j}\frac{\sqrt{\mathbb{E}[\bar{d}_i^4]\mathbb{E}[\bar{d}_j^2]}}{b_n^4}=O\left(\frac{1}{\sqrt{n}}\right). \qquad (30)$$

Then the first two terms of (28) are $o_P(1)$.

Now we study the third term of (28). Straightforward calculation yields

$$\begin{aligned}
&\sum_{i<j}\frac{A_{ij}(\bar{d}_i+\bar{d}_j)^2b_n^2}{b_n^5}\\
= & \sum_{i\neq j}\frac{A_{ij}\bar{d}_i^2}{b_n^3}+\sum_{i\neq j}\frac{A_{ij}\bar{d}_i\bar{d}_j}{b_n^3}\\
= & \sum_{i\neq j}\frac{\bar{A}_{ij}\bar{d}_i^2}{b_n^3}+\sum_{i\neq j}\frac{\bar{A}_{ij}\bar{d}_i\bar{d}_j}{b_n^3}+\sum_{i\neq j}\frac{\mu\bar{d}_i^2}{b_n^3}+\sum_{i\neq j}\frac{\mu\bar{d}_i\bar{d}_j}{b_n^3}. \qquad (31)
\end{aligned}$$

Next we get the leading terms of (31). Note that

$$\sum_{i\neq j}\frac{\mu\bar{d}_i^2-\mathbb{E}[\mu\bar{d}_i^2]}{b_n^3}=\sum_{i\neq j\neq k\neq l}\frac{\mu\bar{A}_{ik}\bar{A}_{il}}{b_n^3}+\sum_{i\neq j\neq k}\frac{\mu\bar{A}_{ik}^2-\mathbb{E}[\mu\bar{A}_{ik}^2]}{b_n^3}. \tag{32}$$

It is easy to get

$$\sum_{i\neq j\neq k\neq l}\frac{\mu\bar{A}_{ik}\bar{A}_{il}}{b_n^3}=\frac{(n-3)\mu}{b_n}\sum_{i\neq k\neq l}\frac{\bar{A}_{ik}\bar{A}_{il}}{b_n^2}, \tag{33}$$

$$\mathbb{E}\left[\left(\sum_{i\neq j\neq k}\frac{\mu(\bar{A}_{ik}^2-\mathbb{E}[\bar{A}_{ik}^2])}{b_n^3}\right)^2\right]=O\left(\frac{1}{n^2}\right), \tag{34}$$

$$\sum_{i\neq j}\frac{\mu\bar{d}_i\bar{d}_j}{b_n^3}=\sum_{i\neq j\neq k\neq l}\frac{\mu\bar{A}_{ik}\bar{A}_{jl}}{b_n^3}+\sum_{i\neq j\neq k}\frac{\mu\bar{A}_{ik}\bar{A}_{jk}}{b_n^3}=O_P\left(\frac{1}{n}\right), \tag{35}$$

$$\sum_{i\neq j}\frac{\bar{A}_{ij}\bar{d}_i^2}{b_n^3}=\sum_{i\neq j\neq k\neq l}\frac{\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{il}}{b_n^3}+\sum_{i\neq j\neq k}\frac{\bar{A}_{ij}\bar{A}_{ik}^2}{b_n^3}=O_P\left(\frac{1}{\sqrt{n}}\right), \tag{36}$$

$$\sum_{i\neq j}\frac{\bar{A}_{ij}\bar{d}_i\bar{d}_j}{b_n^3}=\sum_{i\neq j\neq k\neq l}\frac{\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}}{b_n^3}+\sum_{i\neq j\neq k}\frac{\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jk}}{b_n^3}. \tag{37}$$

By (13), (17), (27), (28)-(37), we get

$$\mathcal{R}_n-\mathbb{E}[\mathcal{R}_n] \quad = \quad \frac{1}{8}\sum_{i\neq j\neq k\neq l}\frac{\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}}{b_n^3}-\frac{1}{8}\sum_{i\neq j\neq k}\frac{\bar{A}_{ij}\bar{A}_{ik}}{b_n^2}+o_P(1). \tag{38}$$

Next we derive the asymptotic distribution of the first two terms of (38) by showing the first two terms of (38) is a degenerate U-statistic. Let $d(X_1,X_2)=||X_1-X_2||_2$,

$$h_1(X_1,X_2,X_3,X_4) \quad = \quad \big(I[d(X_1,X_3)\leq r]-\mu\big)\big(I[d(X_3,X_4)\leq r]-\mu\big)$$

$$
\begin{aligned}
&\times\big(I[d(X_4,X_2)\le r]-\mu\big)\\
&+\big(I[d(X_1,X_4)\le r]-\mu\big)\big(I[d(X_4,X_3)\le r]-\mu\big)\\
&\times\big(I[d(X_3,X_2)\le r]-\mu\big)\\
&+\big(I[d(X_1,X_2)\le r]-\mu\big)\big(I[d(X_2,X_4)\le r]-\mu\big)\\
&\times\big(I[d(X_4,X_3)\le r]-\mu\big)\\
&+\big(I[d(X_1,X_4)\le r]-\mu\big)\big(I[d(X_4,X_2)\le r]-\mu\big)\\
&\times\big(I[d(X_2,X_3)\le r]-\mu\big)\\
&+\big(I[d(X_1,X_2)\le r]-\mu\big)\big(I[d(X_2,X_3)\le r]-\mu\big)\\
&\times\big(I[d(X_3,X_4)\le r]-\mu\big)\\
&+\big(I[d(X_1,X_3)\le r]-\mu\big)\big(I[d(X_3,X_2)\le r]-\mu\big)\\
&\times\big(I[d(X_2,X_4)\le r]-\mu\big)\\
&+\big(I[d(X_2,X_1)\le r]-\mu\big)\big(I[d(X_1,X_4)\le r]-\mu\big)\\
&\times\big(I[d(X_4,X_3)\le r]-\mu\big)\\
&+\big(I[d(X_2,X_4)\le r]-\mu\big)\big(I[d(X_4,X_1)\le r]-\mu\big)\\
&\times\big(I[d(X_1,X_3)\le r]-\mu\big)\\
&+\big(I[d(X_2,X_1)\le r]-\mu\big)\big(I[d(X_1,X_3)\le r]-\mu\big)\\
&\times\big(I[d(X_3,X_4)\le r]-\mu\big)\\
&+\big(I[d(X_2,X_3)\le r]-\mu\big)\big(I[d(X_3,X_1)\le r]-\mu\big)\\
&\times\big(I[d(X_1,X_4)\le r]-\mu\big)\\
&+\big(I[d(X_3,X_1)\le r]-\mu\big)\big(I[d(X_1,X_2)\le r]-\mu\big)\\
&\times\big(I[d(X_2,X_4)\le r]-\mu\big)\\
&+\big(I[d(X_3,X_2)\le r]-\mu\big)\big(I[d(X_2,X_1)\le r]-\mu\big)\\
&\times\big(I[d(X_1,X_4)\le r]-\mu\big),
\end{aligned}
\tag{39}
$$

and

$$
\begin{aligned}
&h_2(X_1,X_2,X_3,X_4)\\
=\;&\big(I[d(X_1,X_2)\le r]-\mu\big)\big(I[d(X_2,X_3)\le r]-\mu\big)
\end{aligned}
$$

$$+\big(I[d(X_1,X_3)\le r]-\mu\big)\big(I[d(X_3,X_2)\le r]-\mu\big)$$
$$+\big(I[d(X_3,X_1)\le r]-\mu\big)\big(I[d(X_1,X_2)\le r]-\mu\big)$$
$$+\big(I[d(X_1,X_2)\le r]-\mu\big)\big(I[d(X_2,X_4)\le r]-\mu\big)$$
$$+\big(I[d(X_1,X_4)\le r]-\mu\big)\big(I[d(X_4,X_2)\le r]-\mu\big)$$
$$+\big(I[d(X_2,X_1)\le r]-\mu\big)\big(I[d(X_1,X_4)\le r]-\mu\big)$$
$$+\big(I[d(X_1,X_3)\le r]-\mu\big)\big(I[d(X_3,X_4)\le r]-\mu\big)$$
$$+\big(I[d(X_1,X_4)\le r]-\mu\big)\big(I[d(X_4,X_3)\le r]-\mu\big)$$
$$+\big(I[d(X_3,X_1)\le r]-\mu\big)\big(I[d(X_1,X_4)\le r]-\mu\big)$$
$$+\big(I[d(X_2,X_3)\le r]-\mu\big)\big(I[d(X_3,X_4)\le r]-\mu\big)$$
$$+\big(I[d(X_2,X_4)\le r]-\mu\big)\big(I[d(X_4,X_3)\le r]-\mu\big)$$
$$+\big(I[d(X_3,X_2)\le r]-\mu\big)\big(I[d(X_2,X_4)\le r]-\mu\big). \tag{40}$$

Then

$$\frac{1}{8}\sum_{i\neq j\neq k\neq l}\frac{\bar{A}_{ij}\bar{A}_{ik}\bar{A}_{jl}}{b_n^3}=\frac{\binom{n}{4}}{4b_n^3}\frac{1}{\binom{n}{4}}\sum_{i<j<k<l}h_1(X_i,X_j,X_k,X_l), \tag{41}$$

and

$$-\frac{1}{8}\sum_{i\neq j\neq k}\frac{\bar{A}_{ij}\bar{A}_{ik}}{b_n^2}=-\frac{\binom{n}{4}}{4b_n^2(n-3)}\frac{1}{\binom{n}{4}}\sum_{i<j<k<l}h_2(X_i,X_j,X_k,X_l). \tag{42}$$

Let $h(X_i,X_j,X_k,X_l)=h_1(X_i,X_j,X_k,X_l)-\mu h_2(X_i,X_j,X_k,X_l)$. Combining (38), (39), (40), (41) and (42) yields

$$\mathcal{R}_n-\mathbb{E}[\mathcal{R}_n] \;=\; \frac{\binom{n}{4}}{4nb_n^3}(nU_n)+o_P(1), \tag{43}$$

where

$$U_n=\frac{1}{\binom{n}{4}}\sum_{i<j<k<l}h(X_i,X_j,X_k,X_l).$$

Next we show U_n is a degenerate U-statistic. It is easy verify that

$$\mathbb{E}[h(X_1,X_2,X_3,X_4)]=0,$$

$$\mathbb{E}[h(X_1, X_2, X_3, X_4)|X_1] = 0.$$

Let

$$g(x_1, x_2) = \mathbb{E}[h(X_1, X_2, X_3, X_4)|X_1 = x_1, X_2 = x_2],$$

$$g_1(x_1, x_2) = \mathbb{E}[h_1(X_1, X_2, X_3, X_4)|X_1 = x_1, X_2 = x_2],$$

$$g_2(x_1, x_2) = \mathbb{E}[h_2(X_1, X_2, X_3, X_4)|X_1 = x_1, X_2 = x_2].$$

Straightforward calculation yields

$$\begin{aligned}
& g(x_1, x_2) \\
= \; & g_1(x_1, x_2) - g_2(x_1, x_2) \\
= \; & 2\mathbb{E}\Big[\big(I[d(x_1, X_3) \le r] - \mu\big)\big(I[d(X_3, X_4) \le r] - \mu\big) \\
& \times\big(I[d(X_4, x_2) \le r] - \mu\big)\Big] \\
& -2\mu\mathbb{E}\Big[\big(I[d(x_1, X_3) \le r] - \mu\big)\big(I[d(X_3, x_2) \le r] - \mu\big)\Big] \\
= \; & 2\mathbb{E}\Big[I[d(x_1, X_3) \le r]I[d(X_3, X_4) \le r]I[d(X_4, x_2) \le r]\Big] \\
& -2\mu\mathbb{E}\Big[I[d(x_1, X_3) \le r]I[d(X_3, x_2) \le r]\Big].
\end{aligned}$$

Since $r \in (0, 1)$, there exists positive constant ϵ such that $(2+\epsilon)r < 2$. For x_1, x_2 satisfying $2r < d(x_1, x_2) < (2 + \epsilon)r$, we have

$$\mathbb{E}\Big[I[d(x_1, X_3) \le r]I[d(X_3, x_2) \le r] = 0,$$

but

$$\mathbb{E}\Big[I[d(x_1, X_3) \le r]I[d(X_3, X_4) \le r]I[d(X_4, x_2) \le r]\Big] > 0.$$

Let $E = \{2r < d(X_1, X_2) < (2 + \epsilon)r\}$. Since

$$\mathbb{E}\Big[g(X_1, X_2)I[E] + g(X_1, X_2)I[E^c]\Big] = \mathbb{E}[g(X_1, X_2)] = 0,$$

then

$$\mathbb{E}\Big[g(X_1, X_2)I[E]\Big] = -\mathbb{E}\Big[g(X_1, X_2)I[E^c]\Big].$$

Note that $\mathbb{P}(E) > 0$ (Lemma 36 in [17]). Then the variance of $g(X_1, X_2)$

can be bounded as follows

$$\begin{aligned} Var[g(X_1,X_2)] &= Var\Big[g(X_1,X_2)I[E]+g(X_1,X_2)I[E^c]\Big] \\ &\geq 2Cov\Big(g(X_1,X_2)I[E],g(X_1,X_2)I[E^c]\Big) \\ &= -2\mathbb{E}\Big[g(X_1,X_2)I[E]\Big]\mathbb{E}\Big[g(X_1,X_2)I[E^c]\Big] \\ &= 2\Big(\mathbb{E}\Big[g(X_1,X_2)I[E]\Big]\Big)^2 > 0. \end{aligned}$$

Hence, U_n is a degenerate U-statistic. By the result in Section 5.5.2 of [27],

$$nU_n \Rightarrow 6\sum_{j=1}^{\infty}\lambda_j(Z_j^2-1),$$

where Z_j are independent standard normal random variables and λ_j are the eigenvalues of the kernel function $g(x_1,x_2)$. By (43) and the fact $\frac{\binom{n}{4}}{4nb_n^3} = \frac{1}{96\mu^3}(1+o(1))$, the proof is complete.

■

References

[1] R. Aguilar-Sánchez, I. F. Herrera-González, J. A. Méndez-Bermúdez, J. M. Sigarreta, Computational properties of general indices on random networks, *Symmetry* **12** (2020) #1341.

[2] R. Aguilar-Sánchez, J. A. Méndez-Bermúdez, F. A. Rodrigues, J. M. Sigarreta, Topological versus spectral properties of random geometric graphs, *Phys. Rev. E.* **102** (2020) #042306.

[3] R. Aguilar-Sánchez, J. A. Méndez-Bermúdez, J. M. Rodriguez, J. M. Sigarreta, Multiplicative topological indices: Analytical properties and application to random networks, *AIMS Math.* **9** (2024) 3646–3670.

[4] D. Bonchev, N. Trinajstić, On topological characterization of molecular branching, *Int. J. Quantum Chem.* **12** (1978) 293–303.

[5] B. Bollobás, P. Erdős, Graphs of extremal weights, *Ars Comb.* **50** (1998) 225–233.

[6] B. Bollobás, P. Erdős, A. Sarkar, Extremal graphs for weights, *Discr. Math.* **200** (1999) 5–19.

[7] C. F. Chiasserini, M. Garetto, E. Leonardi, Social network de-anonymization under scale-free user relations, *IEEE/ACM Trans. Netw.* **24** (2016) 3756–3769.

[8] M. Cavers, S. Kirkland, On the normalized Laplacian energy and general Randić index r_1 of graphs, *Lin. Algebra Appl.* **433** (2010) 172–190.

[9] M. Costanzo, B. Vandersluis, E. N. Koch, A. Baryshnikova, C. Pons, G. Tan, C. Boone, A global genetic interaction network maps a wiring diagram of cellular function, *Science* **353** (2016) #aaf1420.

[10] K. C. Das, S. Sun, I. Gutman, Normalized Laplacian eigenvalues and Randić energy of graphs, *MATCH Commun. Math. Comput. Chem.* **77** (2017) 45–59.

[11] P. De Meo, F. Messina, D. Rosaci, G. M. L. Sarne, A. V. Vasilakos, Estimating graph robustness through the Randić index, *IEEE Trans Cybern.* **48** (2018) 3232–3242.

[12] S. Dattola, N. Mammone, F. C. Morabito, D. Rosaci, G.M.L. Sarne, F. L. Foresta, Testing graph robustness indexes for EEG analysis in Alzheimer's disease diagnosis, *Electronics* **10** (2021) #1440.

[13] Q. Duchemin, Y. De Castro, Random geometric graph: some recent developments and perspectives, in: R. Adamczak, N. Gozlan, K. Lounici, M. Madiman (Eds.), *High Dimensional Probability IX*, Springer, Cham, 2023, pp. 347–392.

[14] E. Estrada, Quantifying network heterogeneity, *Phys. Rev. E* **82** (2010) #066102.

[15] E. Estrada, Topological analysis of SARS CoV-2 main protease, *Chaos* **30** (2020) #061102.

[16] D. Fourches, A. Tropsha, Using graph indices for the analysis and comparison of chemical datasets, *Mol. Inf.* **32** (2013) 827–842.

[17] S. Galhotra, A. Mazumdar, S. Pal, B. Saha, Community recovery in the geometric block model, *J. Machine Learn. Res.* **24** (2023) 1–53.

[18] X. Li, Y. Shi, A survey on the Randić index, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 127–156.

[19] C. T. Martínez-Martínez, J. A. Méndez-Bermúdez, J. M. Rodriguez, J. Sigarreta, Computational and analytical studies of the Randić index in Erdő-Rényi models, *Appl. Math. Comput.* **377** (2020) #125137.

[20] C. T. Martínez-Martínez, J. A. Méndez-Bermúdez, J. M. Rodriguez, J. Sigarreta, Computational and analytical studies of the harmonic index on Erdős-Rényi models, *MATCH Commun. Math. Comput. Chem.* **85** (2021) 395–426.

[21] M. Newman, The structure and function of complex networks, *SIAM Rev.* **45** (2003) 167–256.

[22] A. J. O'Malley, P. V. Marsden, The analysis of social networks, *Health Serv. Outcomes Res. Method.* **8** (2008) 222–269.

[23] F. Radicchi, S. Fortunato, A. Vespignani, *Models of Science Dynamics: Encounters between Complexity Theory and Information Sciences*, Springer, Berlin, 2012.

[24] M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.

[25] M. Randić, On history of the Randić index and emerging hostility toward chemical graph theory, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 5–124.

[26] M. Randić, M. Nović, D. Plavšić, *Solved and Unsolved Problems in Structural Chemistry*, CRC Press, Boca Raton, 2016.

[27] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[28] M. Yuan, On the Randić index and its variants of network data, *TEST* **33** (2023) 155–179.

[29] M. Yuan, Asymptotic distribution of degree-based topological indices, *MATCH Commun. Math. Comput. Chem.* **91** (2024) 135–196.

[30] M. Yuan, X. Zhao, Asymptotic distributions of the average clustering coefficient and its variant, 2023, `https://arxiv.org/pdf/2311.10979.pdf`