Check for updates

J. Appl. Probab. **63,** 919–949 (2026)
*doi:*10.1017/jpr.2025.10063

ANALYSIS OF CLUSTERING AND DEGREE INDEX IN RANDOM GRAPHS AND COMPLEX NETWORKS

ÜMİT IŞLAK,* *** *Boğaziçi University and Middle East Technical University*
BARIŞ YEŞİLOĞLU,* ** *Boğaziçi University*

Abstract

The purpose of this paper is to analyze the degree index and the clustering index in dense random graphs. The degree index in our setup is a certain measure of degree irregularity whose basic properties are well studied in the literature, and the corresponding theoretical analysis in a random graph setup turns out to be tractable. On the other hand, the clustering index, based on a similar reasoning, is first introduced in this paper. Computing exact expressions for the expected clustering index turns out to be more challenging even in the case of Erdős–Rényi graphs, and our results are on obtaining relevant upper bounds. These are also complemented with observations based on Monte Carlo simulations. In addition to the Erdős–Rényi case, we also present a simulation-based analysis for random regular graphs, the Barabási–Albert model, and the Watts–Strogatz model.

Keywords: Erdős–Rényi graphs; clustering coefficient; degree irregularity

2020 Mathematics Subject Classification: Primary 60C05; 05C80
Secondary 05C82; 05C07

1. Introduction

Since the introduction of Erdős–Rényi graphs around 1960 [10, 11], random graphs have gained substantial importance in both applied and theoretical fields. Meanwhile, various statistics of interest for the underlying networks (or graphs) have been studied, including, among many others, the average degree, clustering coefficient, efficiency, and modularity. See, for example, [8, 13, 15, 22] for relevant definitions and certain applications. Due to the development of graph-based machine-learning algorithms in recent years, these statistics have become even more popular, especially in classification problems [23]. The purpose of this paper is to study two other types of graph characteristics in random graphs.

In order to describe the two characteristics that we study, let $\mathcal{G} = (V, E)$ be a graph where $V = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. Let $N(k)$ be the set of neighbors of $k \in V$ in $\mathcal{G}$. We denote

Received 16 August 2024; accepted 30 October 2025.
* Postal address: Boğaziçi University, Mathematics Department, Istanbul, Turkey.
** Postal address: Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey. Email: umit.islak1@bogazici.edu.tr and uislak@metu.edu.tr
*** Postal address: Boğaziçi University, Mathematics Department, Istanbul, Turkey.

© The Author(s), 2026. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

the degree of k by d_k. With this notation, we define the *α-level degree index* of $\mathcal{G}$ to be

$$\mathrm{DI}_\alpha(\mathcal{G}) = \sum_{1\le i<j\le n} |d_i - d_j|^\alpha.$$

Here, $\alpha > 0$ is a real number, which is chosen to be in $\{1, 2\}$ in this paper. In addition, let us note that $\sum_{i<j}$ will mean $\sum_{1\le i<j\le n}$ in the following. Noting that $\mathrm{DI}_\alpha(\mathcal{G}) = 0$ means that the graph under study is regular, this quantity is considered as a measure of irregularity of the graph [24]. On the other hand, in networks with high disparity in node connectivity, the index $\mathrm{DI}_\alpha(\mathcal{G})$ would yield large values. The index overall provides a global measure of degree variability in the network and may, for example, help differentiate between an Erdős–Rényi graph and a Barabási–Albert network even if they have the same number of nodes and edges.

A definition of degree irregularity that has found significant use in the literature is the *Albertson index* introduced in [5], in which the corresponding definition is similar to ours, but the sum under consideration is over the edge set instead of the vertices. Although the irregularity used in the present paper is quite natural to consider, the special case $\mathrm{DI}_1(\mathcal{G})$ was relatively recently introduced in [1], where the authors call it the *total irregularity* of the graph. In addition, a form of the special case of $\mathrm{DI}_2(\mathcal{G})$, in which the corresponding sum is over the edge set instead of the vertex set, was proposed in [2]. Aside from these, there are other notions of graph degree irregularity; see [3] for a partial list and [6] for a broad look at the topic.

The second graph characteristic of interest for us will be the clustering index, which, as far as we know, has not previously been studied in the literature. Before introducing this index, recall that for given $i \in V$, when $d_i \ge 2$, the *local clustering coefficient* of i is defined to be

$$C(i) = \frac{1}{d_i(d_i-1)} \sum_{j,k\in N(i);j\ne k} \mathbf{1}(j \leftrightarrow k).$$

In the case where $d_i \in \{0, 1\}$, we set $C(i) = 0$. In general, the clustering coefficient of a node $i \in V(\mathcal{G})$ is a measure of the likelihood of the neighbors of i to be neighbors among themselves [25]. The clustering coefficient is studied extensively in the literature, and several variations and generalizations are proposed; see [16, 18, 26] for some exemplary work. The clustering coefficient is also studied in a random graph setup. For instance, [17] analyzed the clustering coefficient in Erdős–Rényi graphs and certain random regular graphs. An investigation was performed in [12] in the case of a generalized small world model. On the other hand, a natural generalization of the standard clustering coefficient was discussed in [26] and both Erdős–Rényi and small world settings were analyzed.

Now, similar to the degree index, for $\alpha > 0$, we define the *α-level clustering index* of $\mathcal{G}$ to be

$$\mathrm{CI}_\alpha(\mathcal{G}) = \sum_{i<j} |C(i) - C(j)|^\alpha.$$

Here, again, α in our case is either 1 or 2. As noted previously, although such an index seems natural to define, it is first introduced here to the best of the authors' knowledge. Before continuing with our contributions, let us now include some discussion of this newly introduced index.

The clustering index $\mathrm{CI}_\alpha(\mathcal{G})$ captures the heterogeneity in local clustering across the graph, offering insights that traditional global or average measures may miss. For instance, it can reveal whether certain nodes are embedded in tightly knit communities where others are more

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

isolated. This distinction is crucial in real-world networks, where some nodes may form dense subgraphs, while others may lie on the periphery with sparse neighborhoods [7]. In many systems, the core nodes tend to have high clustering, whereas the peripheral nodes exhibit low clustering. A high clustering index reflects this inequality, which could remain hidden when only considering global or average clustering coefficients.

Moreover, in networks with community structure [20], nodes within the same community often display similar clustering levels. Hence, a low clustering index may indicate structural homogeneity, whereas a high value can signal the presence of multiple structurally distinct communities. As a last note, when analyzing temporal networks, the global clustering might remain stable over time, even as the roles of nodes shift. In contrast, the clustering index is sensitive to such changes in clustering variability, making it a possible tool for detecting structural transitions or evolving roles in dynamic graphs.

Moving to the contributions of this article, our primary goal is to analyze $\mathrm{DI}_\alpha(G)$ and $\mathrm{CI}_\alpha(G)$ when G is an Erdős–Rényi graph. For such random graphs, the degree index case turns out to be relatively straightforward, and precise formulas for the expected degree index are obtainable. In particular, the results in Section 4 show that when G is an Erdős–Rényi graph with n nodes and a fixed attachment probability p, we have

$$\mathrm{E}[\mathrm{DI}_1(\mathcal{G})] \sim \frac{2}{\sqrt{\pi}}\binom{n}{2}\sqrt{(n-2)p(1-p)} \quad \text{and} \quad \mathrm{E}[\mathrm{DI}_2(\mathcal{G})] = 6\binom{n}{3}p(1-p).$$

On the other hand, the computations for the case of the expected clustering index are more involved, and we are not able to obtain exact formulas. However, for the same random graph family, we show that the bounds

$$\mathrm{E}[\mathrm{CI}_1(\mathcal{G})] \leq K_1 n \quad \text{and} \quad \mathrm{E}[\mathrm{CI}_2(\mathcal{G})] \leq K_2$$

hold for every n, for some constants K_1, K_2 depending on p, but not on n. Our heuristic arguments yield that matching lower bounds are true for both cases, but we have not yet verified these rigorously. Moreover, we complement our theoretical results by making observations on the two indices of interest for other random graph models via Monte Carlo simulations. The models we study include random regular graphs, the Barabási–Albert model, and the Watts–Strogatz model, and relevant simulations and discussions are given in Section 5.1.

The remainder of the paper is organized as follows. The next section contains some basic observations on the clustering index, which can be easily adapted to the degree index. This section also includes a brief discussion on the comparison of the extremal cases of the degree index and the clustering index. In Section 3 we analyze the clustering index in Erdős–Rényi graphs, and provide upper bounds for the expected clustering index for both $\alpha = 1$ and 2. In Section 4, a similar study is done for the degree index. Section 5 is then devoted to a simulation-based study for the indices under study for random regular graphs, the Barabási–Albert model, and the Watts–Strogatz model. Lastly, we conclude the paper in Section 6 with some discussions on possible future work.

2. Some Basic Observations on the Clustering Index

In this section, we make some elementary observations on the clustering index, beginning with the analysis of the extremal values. In the following, whenever it is clear from the context,

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press



Complete Graph K_4 Null Graph N_4

FIGURE 1. A graph with eight vertices where $\mathrm{CI}_\alpha = 8^2/4 = 16$.

we assume that the node set of the underlying graph is $V = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. The following elementary lemma is used to determine the extremal values of $\mathrm{CI}_\alpha(\mathcal{G})$.

Lemma 2.1. *Let $a_1, a_2, \ldots, a_n$ be real numbers on $[0, 1]$. Then for any $\alpha \geq 1$, we have*

$$\sum_{1 \leq i < j \leq n} |a_i - a_j|^\alpha \leq \frac{n^2}{4}.$$

Proof. Clearly, it suffices to show that $\sum_{1 \leq i < j \leq n} |a_i - a_j| \leq n^2/4$. Now, without loss of generality, we assume $0 \leq a_1 \leq \cdots \leq a_n \leq 1$, and define $b_i := a_i - a_{i-1}$ for every $i \in \{2, \ldots, n\}$. Then, for every $j > i$, we have $|a_i - a_j| = a_j - a_i = \sum_{k=i+1}^{j} b_k$. Therefore,

$$S := \sum_{1 \leq i < j \leq n} |a_i - a_j| = \sum_{1 \leq i < j \leq n} \sum_{k=i+1}^{j} b_k = \sum_{k=2}^{n} (k-1)(n+1-k) b_k.$$

The reason for the last equality is that for the terms containing b_k, we must have $i \leq k-1$ and $j \geq k$. The number of such pairs is $(k-1)(n+1-k)$. The maximum value $(k-1)(n+1-k)$ can take is $n^2/4$. Therefore, $S \leq (n^2/4) \sum_{k=2}^{n} b_k = (n^2/4)\,(a_n - a_1) \leq n^2/4$. □

Now, the next corollary provides the extremal values of $\mathrm{CI}_\alpha(\mathcal{G})$. The lower bound in the corollary is clear, and the upper bound follows from Lemma 2.1.

Corollary 2.1. *For any graph $\mathcal{G} = (V, E)$ and for any $\alpha \geq 1$, we have*

$$0 \leq \mathrm{CI}_\alpha(\mathcal{G}) \leq \frac{n^2}{4}.$$

Example 2.1. In this example, we discuss some elementary graphs for which the bounds stated in Corollary 2.1 are attained. Let $\alpha \geq 1$ be arbitrary. The lower bound in Corollary 2.1 is attained, for example, for the following graphs:

- the null graph and the complete graph;
- any forest and, in particular, any tree.

As a simple example where the clustering index is maximized, let $m \geq 2$, $n = 2m$, and consider the graph that is the graph union of an m-null set, and K_m (complete graph with m vertices). Figure 1 exemplifies such a construction when $n = 8$.

Here, the only contribution to the sum defining CI_α is from the cases where a node from K_m and another node from the m-null set are considered. Noting that the clustering coefficient

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

of the node from the m-null set is 0 by definition, the observation that CI_α takes the value $n^2/4$ follows immediately.

Another related example is given in Example 2.3 where extremal situations in clustering and degree indices are compared and discussed.

Noting that the observations in Corollary 2.1 can be easily adapted to the degree index, we continue with the following result which provides a way to express the clustering index in terms of the order statistics of the corresponding local clustering coefficients.

Lemma 2.2. (Expression in terms of order statistics.) *Let $\mathcal{G} = (V, E)$ be some graph. Order the local clustering coefficients as $C_{(1)} \le C_{(2)} \le \cdots \le C_{(n)}$. Then,*

$$\sum_{i<j} |C(i) - C(j)| = \sum_{i=2}^{n-1} (2i - 1 - n)C_{(i)} + (n-1)(C_{(n)} - C_{(n-1)}).$$

Proof. Following the proof of Lemma 2.1 with $a_i = C_{(i)}$ gives

$$\sum_{i<j} |C(i) - C(j)| = \sum_{k=2}^{n} (k-1)(n+1-k)(C_{(k)} - C_{(k-1)}).$$

Expanding the right-hand side as $\sum_{k=2}^{n} (k-1)(n+1-k)C_{(k)} - \sum_{k=2}^{n} (k-1)(n+1-k)C_{(k-1)}$, changing the index in the second sum properly, and performing some elementary manipulations yield the result. □

Let us also examine extremal cases in random graphs. Clearly, any random tree model provides a random graph model for which the lower bound in Corollary 2.1 is achieved. The following example, which is a randomized version of the construction given in Example 2.1, provides a simple random graph model where the upper bound in Corollary 2.1 is achieved asymptotically.

Example 2.2. Begin with two empty sets of nodes V_1 and V_2 at time 0. At each time $t \in \mathbb{N}$, a newly generated node v_t is inserted in V_1 or V_2 with probabilities p and $1-p$, respectively. If v_t is inserted in V_1, we do not attach it to any present nodes. If it is inserted in V_2, then we attach it to all vertices present in V_2. Then, at time $n \in \mathbb{N}$, the vertices in V_1 form a null graph, and the vertices in V_2 form a complete graph.

Letting N be the number of vertices in V_1 at time n, N is binomially distributed with parameters n and p, and we have

$$\begin{aligned}\mathrm{E}\left[\sum_{i<j} |C(i) - C(j)|\right] &= \mathrm{E}\left[\mathrm{E}\left[\sum_{i<j} |C(i) - C(j)| \,\middle|\, N\right]\right] = \mathrm{E}[N(n-N)] = n\mathrm{E}[N] - \mathrm{E}[N^2] \\ &= n^2 p - np(1-p) - n^2p^2 = (n^2 - n)p(1-p).\end{aligned}$$

When $p = 1/2$, we have $\mathrm{E}[\mathrm{CI}_1(\mathcal{G})] = \frac{1}{4}(n^2 - n)$. In particular, for large n, $\mathrm{E}[\mathrm{CI}_1(\mathcal{G})] \approx n^2/4$.

In the following last example, we provide a comparison of the clustering index and the degree index. In particular, we are interested in cases where the degree index is high and the clustering coefficient is low, and vice versa.

Example 2.3. Consider a collection of disjoint polygons, i.e. a graph where each node is contained in a unique polygon. We have $C(i) = 1$ if i is contained in a triangle and $C(i) = 0$ if

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

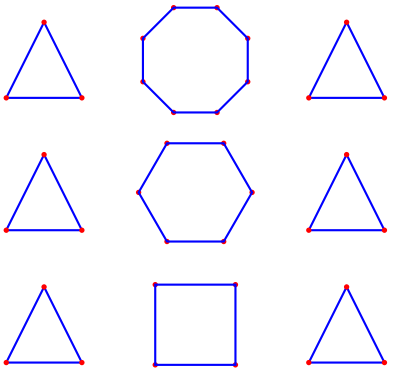


FIGURE 2. A graph with $d_i = 2$ for any i (hence, $\mathrm{DI}_\alpha = 0$) but $\mathrm{CI}_\alpha(\mathcal{G}) = n^2/4$.

i is contained in a polygon other than a triangle. The degree of any node in this graph is 2. Assuming we have k vertices contained in some triangle and $n-k$ vertices contained in a polygon which is not a triangle, we have $\mathrm{CI}_\alpha = \sum_{1\le i<j\le n} |C(i) - C(j)|^\alpha = nk - k^2$. If n is even, then the maximum is attained when $k = n/2$, which yields $n^2/4$. Thus we have found a graph for which the degree index is zero, but the clustering index is $n^2/4$, the maximal possible value.

Next, we also discuss an example where the clustering index is zero while the degree index $\sum_{1\le i<j\le n} |d_i - d_j|$ is $n^2(n/2-3)/4$. (Lemma 2.1 can be used to show that the degree index can be at most $n^3/4$. Thus, our example here is off by a factor of $1/2$, but still demonstrates that the degree index can be very large while the clustering index is zero.)

Let n be an integer divisible by 6. Consider a complete graph with $n/2$ vertices, and add $n/6$ disjoint triangles to the graph. Then $C(i) = 1$ for any i, and so the clustering index is zero. On the other hand, $|d_i - d_j| = n/2 - 3$ if one of the i,j is in our original complete graph while the other belongs to a triangle. Since there are $n^2/4$ of these non-zero terms, we conclude $\sum_{1\le i\le j\le n} |d_i - d_j| = \frac{n^2(n/2-3)}{4}$.

3. Clustering Index of Erdős–Rényi Graphs

3.1. Sublinearity of $\mathrm{E}[\mathrm{CI}_1(\mathcal{G})]$ and $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$

In this subsection we focus on the case where $\mathcal{G}$ is an Erdős–Rényi graph, and show that $\mathrm{E}[\mathrm{CI}_1(\mathcal{G})]$ and $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$ are both sublinear. The result for the latter is improved later and, in particular, $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$ will be shown to be bounded by a constant independent of n. We begin with the following proposition concerning the first two moments of the local clustering coefficient.

Proposition 3.1. *Let $\mathcal{G} = (V, E)$ be an Erdős–Rényi graph with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$. Let $i \in V$ be any node.*

(i) *We have*

$$\mathrm{E}[C(i)] = p(1 - (1-p)^{n-1} - (n-1)p(1-p)^{n-2}).$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

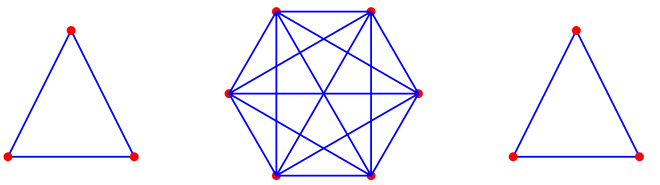


FIGURE 3. A graph with $C(i) = 1$ for any i (hence, $\mathrm{CI}_\alpha(\mathcal{G}) = 0$) but $\mathrm{DI}_1(\mathcal{G}) = n^2(n/2 - 3)/4$.

In particular, $\mathrm{E}[C(i)] \geq p - D/n^2$ *for each n, for some constant D depending on p, but not on n.*

(ii) *We have*

$$|\mathrm{E}[C(i)^2] - p^2| \leq \frac{D_1}{n^2},$$

where D_1 *is a positive constant depending on p, but not on n.*

(iii) *We have* $\mathrm{var}(C(i)) = \mathcal{O}(1/n^2)$.

Proof.

(i) First, letting $\phi = \mathrm{P}(d_i \leq 1) = \mathrm{P}(d_i = 0) + \mathrm{P}(d_i = 1) = (1-p)^{n-1} + (n-1)p(1-p)^{n-2}$, we have

$$\mathrm{E}[C(i)] = \phi \mathrm{E}[C(i) \mid d_i \leq 1] + (1-\phi)\mathrm{E}[C(i) \mid d_i \geq 2]. \tag{3.1}$$

Now, by the definition of $C(i)$, $\mathrm{E}[C(i) \mid d_i \leq 1] = 0$. We claim $\mathrm{E}[C(i) \mid d_i \geq 2] = p$. We have

$$\begin{aligned}\mathrm{E}[C(i) \mid d_i \geq 2] &= \mathrm{E}[\mathrm{E}[C(i) \mid d_i, d_i \geq 2] \mid d_i \geq 2] \\ &= \mathrm{E}\left[\frac{1}{\binom{d_i}{2}} \mathrm{E}\left[\sum_{k,\ell \in N(i)} \mathbf{1}(k \leftrightarrow \ell) \,\middle|\, d_i, d_i \geq 2\right] \,\middle|\, d_i \geq 2\right],\end{aligned}$$

where we recall that $N(i)$ is the set of neighbors of node i. Now given d_i with $d_i \geq 2$, $\sum_{k,\ell \in N(i)} \mathbf{1}(k \leftrightarrow \ell)$ is binomially distributed with parameters $\binom{d_i}{2}$ and p, whose expectation is $\binom{d_i}{2}p$. Therefore, continuing the last observations, we conclude

$$\mathrm{E}[C(i) \mid d_i \geq 2] = \mathrm{E}\left[\frac{1}{\binom{d_i}{2}}\binom{d_i}{2}p\right] = p.$$

Substituting this into (3.1), we obtain

$$\begin{aligned}\mathrm{E}[C(i)] &= p(1 - (1-p)^{n-1} - (n-1)p(1-p)^{n-2}) \\ &= p - p(1-p)^{n-1} - (n-1)p^2(1-p)^{n-2}.\end{aligned}$$

Now, for $p \in (0, 1)$, the sequence $p(1-p)^{n-1} + (n-1)p^2(1-p)^{n-2}$ decays exponentially fast as $n \to \infty$ and, thus, it is upper bounded by D/n^2 for some $D > 0$. The assertion that $\mathrm{E}[C(i)] \geq p - D/n^2$ is now clear.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

(ii) Note that, in the following, D denotes a constant independent of n that does not necessarily have the same value in its two appearances.

Let A_i be the event that $|d_i - (n-1)p| < n^{2/3}$. Then by McDiarmid (or Azuma–Hoeffding) inequality [19], $\mathrm{P}(\overline{A_i}) \le M_1 e^{-M_2 n^{1/3}}$ for some positive constants M_1, M_2 independent of n. Using A_i, we now write,

$$\mathrm{E}[C(i)^2] = \mathrm{E}[C(i)^2 \mathbf{1}(A_i)] + \mathrm{E}[C(i)^2 \mathbf{1}(\overline{A_i})].$$

Let us analyze the two terms on the right-hand side separately. First, focusing on the second term, noting the trivial bound $|C(i)| \le 1$, observe that

$$\mathrm{E}[C(i)^2 \mathbf{1}(\overline{A_i})] \le \mathrm{P}(\overline{A_i}) \le M_1 e^{-M_2 n^{1/3}} \le \frac{D}{n^2}.$$

Hence, for the second term, we have

$$0 \le \mathrm{E}[C(i)^2 \mathbf{1}(\overline{A_i})] \le \frac{D}{n^2}. \tag{3.2}$$

Moving on to the first term, we have

$$\begin{aligned}
\mathrm{E}[C(i)^2 \mathbf{1}(A_i)] &= \mathrm{E}[\mathrm{E}[C(i)^2 \mathbf{1}(A_i) \mid d_i]] \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{1}{\binom{d_i}{2}^2} \left(\sum_{\{j,k\} \subset N(i)} \mathbf{1}(j \leftrightarrow k) \right)^2 \mathbf{1}(A_i) \,\middle|\, d_i \right] \right] \\
&= \mathrm{E}\left[\frac{\mathbf{1}(A_i)}{\binom{d_i}{2}^2} \mathrm{E}\left[\sum_{\{j,k\} \subset N(i)} \mathbf{1}(j \leftrightarrow k) + \sum_{\{j,k\},\{j',k'\} \subset N(i), \{j,k\} \neq \{j',k'\}} \mathbf{1}(j \leftrightarrow k, j' \leftrightarrow k') \,\middle|\, d_i \right] \right] \\
&= \mathrm{E}\left[\frac{\mathbf{1}(A_i)}{\binom{d_i}{2}^2} \left(\binom{d_i}{2} p + \left(\binom{d_i}{2}^2 - \binom{d_i}{2} \right) p^2 \right) \right] \\
&= p^2 \mathrm{E}[\mathbf{1}(A_i)] + (p - p^2) \mathrm{E}\left[\frac{\mathbf{1}(A_i)}{\binom{d_i}{2}} \right] = p^2 \mathrm{P}(A_i) + (p - p^2) \mathrm{E}\left[\frac{\mathbf{1}(A_i)}{\binom{d_i}{2}} \right].
\end{aligned} \tag{3.3}$$

Now, when A_i is true, $(n-1)p - n^{2/3} \le d_i \le (n-1)p + n^{2/3}$. Keeping this in mind, and using the last relation we obtained, we get

$$\mathrm{E}[C(i)^2 \mathbf{1}(A_i)] \le p^2 \cdot 1 + \frac{p - p^2}{\binom{(n-1)p - n^{2/3}}{2}} \le p^2 + \frac{D}{n^2}.$$

In addition, again by (3.3), we have

$$\mathrm{E}[C(i)^2 \mathbf{1}(A_i)] \ge p^2 \mathrm{P}(A_i) \ge p^2 \left(1 - M_1 e^{-M_2 n^{1/3}}\right) \ge p^2 - \frac{D}{n^2}.$$

Hence, we conclude that

$$|\mathrm{E}[C(i)^2 \mathbf{1}(A_i)] - p^2| \le \frac{D}{n^2}. \tag{3.4}$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

Combining (3.2) and (3.4), we arrive at

$$|\mathrm{E}[C(i)^2 - p^2]| \leq \frac{D_1}{n^2},$$

which holds for every n, where D_1 is depending on p, but not on n, as asserted.

(iii) Follows immediately from (i) and (ii). □

Proposition 3.1 can now be used to show that $\mathrm{E}[\mathrm{CI}_1(\mathcal{G})]$ is sublinear.

Theorem 3.1. *Let $\mathcal{G}$ be an Erdős–Rényi graph with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$. Then we have*

$$\mathrm{E}[\mathrm{CI}_1(\mathcal{G})] \leq Kn, \quad n \geq 1,$$

where K is a constant depending on p, but not on n.

Proof. Let $i, j \in \mathbb{N}$. Note that $C(i)$ and $C(j)$ have the same distribution in Erdős–Rényi graphs due to the underlying symmetry. Using this along with the Cauchy–Schwarz inequality, and the previous proposition, we observe

$$\begin{aligned}\mathrm{E}[|C(i) - C(j)|] &= \mathrm{E}[|C(i) - \mathrm{E}[C(i)] + \mathrm{E}[C(j)] - C(j)|] \\ &\leq 2\mathrm{E}|C(i) - \mathrm{E}|C(i)|| \leq 2(\mathrm{E}|C(i) - \mathrm{E}|C(i)||^2)^{1/2} = 2\sqrt{\mathrm{var}(C(i))} \leq \frac{D}{n},\end{aligned}$$

for some constant D. Thus,

$$\mathrm{E}[\mathrm{CI}_1(\mathcal{G})] \leq \binom{n}{2}\frac{D}{n} \leq Kn.$$

□

Noting that the trivial inequality $|C(i) - C(j)|^2 \leq |C(i) - C(j)|$ always holds, one obtains the following corollary.

Corollary 3.1. *In the setting of the previous theorem,*

$$\mathrm{E}[\mathrm{CI}_2(\mathcal{G})] \leq Kn, \quad n \geq 1,$$

where K is a constant depending on p, but independent of n.

In the next subsection, we show that in the case of Erdős–Rényi graphs for fixed p, $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$ is indeed bounded by a constant that does not depend on n. Note that our simulation results further suggest that for Erdős–Rényi graphs, the growth of $\mathrm{CI}_1(\mathcal{G})$ behaves linearly in the number of nodes, and $\mathrm{CI}_2(\mathcal{G})$ converges to a positive constant as the number of nodes increases.

3.2. Further analysis of $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$

Now that we know $\mathrm{E}[\mathrm{CI}_2(\mathcal{G})]$ is sublinear via Corollary 3.1 when $\mathcal{G}$ is an Erdős–Rényi graph, we focus on this expectation in more detail and provide a more precise analysis for this case. This will require understanding of expectations of the form $\mathrm{E}[C(i)C(j)]$ for nodes $i \neq j$.

Proposition 3.2. *Let $\mathcal{G}$ be an Erdős–Rényi graph with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, and i, j be two distinct nodes. Then,*

$$\mathrm{E}[C(i)C(j)] \geq p^2 - \frac{D_2}{n^2},$$

for each n, where D_2 is a positive constant depending on p, but not on n.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

We defer the proof of Proposition 3.2 to the end of this subsection and first discuss the main result.

Theorem 3.2. *Let $\mathcal{G}$ be an Erdős–Rényi graph with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$. Then we have*

$$\mathrm{E}[\mathrm{CI}_2(\mathcal{G})] \leq K$$

for some constant K independent of n.

Proof. By Propositions 3.1 and 3.2, we have

$$\begin{aligned}\mathrm{E}[\mathrm{CI}_2(\mathcal{G})] &= \binom{n}{2}\mathrm{E}[(C(i) - C(j))^2] = \binom{n}{2}\left(\mathrm{E}[C(i)^2] + \mathrm{E}[C(j)^2] - 2\mathrm{E}[C(i)C(j)]\right) \\ &= n(n-1)\big(\mathrm{E}[C(i)^2] - \mathrm{E}[C(i)C(j)]\big) \leq n^2\left(p^2 + \frac{D_1}{n^2} - p^2 + \frac{D_2}{n^2}\right) = D_1 + D_2 =: K.\end{aligned}$$

□

Let us now prove Proposition 3.2.

Proof of Proposition 3.2. Recall from the proof of Proposition 3.1 that A_i is the event that $|d_i - (n-1)p| < n^{2/3}$. Define A_j similarly. We have

$$\mathrm{E}[C(i)C(j)] = \mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] + S,$$

where

$$S = \mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(\overline{A_j})] + \mathrm{E}[C(i)C(j)\mathbf{1}(\overline{A_i})\mathbf{1}(A_j)] + \mathrm{E}[C(i)C(j)\mathbf{1}(\overline{A_i})\mathbf{1}(\overline{A_j})].$$

Noting that A_i and A_j are equally likely, and using trivial upper bound for the clustering coefficients, we see that

$$S \leq 3\mathrm{P}(\overline{A_i}).$$

Recalling now $\mathrm{P}(\overline{A_i}) \leq M_1 e^{-M_2 n^{1/3}}$ for some constants $M_1, M_2 > 0$ from the Proof of Proposition 3.1, we conclude that

$$S \leq D/n^2, \quad n \geq 1,$$

where D is a positive constant independent of n. (Again, D in distinct appearances may denote different constants.) Next, we focus on the estimation of $\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)]$.

Let M be the event that the edge between i, j exists. In addition, let T_i and T_j denote the number of triangles that contain i and j as a node, respectively. Then

$$\begin{aligned}\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] &= p\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j) \mid M] + (1-p)\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j) \mid \overline{M}] \\ &= p\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}\mathrm{E}[T_iT_j \,\big|\, d_i, d_j, M] \,\middle|\, M\right] \\ &\quad + (1-p)\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}\mathrm{E}[T_iT_j \,\big|\, d_i, d_j, \overline{M}] \,\middle|\, \overline{M}\right]. \qquad (3.5)\end{aligned}$$

Now we examine the conditional expectation of T_iT_j given d_i, d_j, M and again of T_iT_j given d_i, d_j, $\overline{M}$ in more detail. For this purpose, we decompose T_i and T_j as

$$T_i = T_{i1} + T_{i2} + T_{i3} \quad \text{and} \quad T_j = T_{j1} + T_{j2} + T_{j3},$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

where:

- $T_{i1} = T_{j1}$ denotes the number of triangles that have both i and j as vertices;
- $T_{i2} = T_{j2}$ denotes the number of edges $e_{k\ell}$, with k and ℓ distinct from i and j, and the triangles $\{i, k, \ell\}$ and $\{j, k, \ell\}$ both appear in the graph;
- T_{i3} and T_{j3} denote the remaining triangles containing i and j as a node, respectively.

A few observations about this decomposition are in order. First note that conditional on $\overline{M}$, both T_{i1} and $T_{j1} = 0$. Second, apart from the products $T_{i1}T_{j1} = T_{i1}^2$ and $T_{i2}T_{j2} = T_{i2}^2$, the products appearing in T_iT_j are products of independent random variables given M, d_i, d_j or $\overline{M}, d_i, d_j$. Hence, to estimate these terms, we can investigate the remainder we obtain from $\mathrm{E}[T_i]\mathrm{E}[T_j]$ in the following way. We have

$$\begin{aligned}
\mathrm{E}[T_iT_j \mid M, d_i, d_j] &= \sum_{k=1}^{3}\sum_{\ell=1}^{3} \mathrm{E}[T_{ik}T_{j\ell} \mid M, d_i, d_j] \\
&= \left(\sum_{(k,\ell)\notin\{(1,1),(2,2)\}} \mathrm{E}[T_{ik} \mid M, d_i, d_j]\mathrm{E}[T_{j\ell} \mid M, d_i, d_j] \right) \\
&\quad + \mathrm{E}[T_{i1}^2 \mid M, d_i, d_j] + \mathrm{E}[T_{i2}^2 \mid M, d_i, d_j] \\
&= \sum_{k=1}^{3}\sum_{\ell=1}^{3} \mathrm{E}[T_{ik} \mid M, d_i, d_j]\mathrm{E}[T_{j\ell} \mid M, d_i, d_j] \\
&\qquad - \mathrm{E}[T_{i1} \mid M, d_i, d_j]\mathrm{E}[T_{j1} \mid M, d_i, d_j] - \mathrm{E}[T_{i2} \mid M, d_i, d_j]\mathrm{E}[T_{j2} \mid M, d_i, d_j] \\
&\qquad + \mathrm{E}[T_{i1}^2 \mid M, d_i, d_j] + \mathrm{E}[T_{i2}^2 \mid M, d_i, d_j] \\
&= \mathrm{E}[T_i \mid M, d_i, d_j]\mathrm{E}[T_j \mid M, d_i, d_j] \\
&\qquad + (\mathrm{E}[T_{i1}^2 \mid M, d_i, d_j] - (\mathrm{E}[T_{i1} \mid M, d_i, d_j])^2) \\
&\qquad + (\mathrm{E}[T_{i2}^2 \mid M, d_i, d_j] - (\mathrm{E}[T_{i2} \mid M, d_i, d_j])^2) \\
&= \mathrm{E}[T_i \mid M, d_i, d_j]\mathrm{E}[T_j \mid M, d_i, d_j] \\
&\qquad + \mathrm{E}[T_{i1}^2 - (\mathrm{E}[T_{i1}|M, d_i, d_j])^2|M, d_i, d_j] \\
&\qquad + \mathrm{E}[T_{i2}^2 - (\mathrm{E}[T_{i2}|M, d_i, d_j])^2|M, d_i, d_j]. \qquad (3.6)
\end{aligned}$$

In addition, recalling that conditional on $\overline{M}$, both T_{i1} and $T_{j1} = 0$, and following similar steps we obtain

$$\mathrm{E}[T_iT_j \mid \overline{M}, d_i, d_j] = \mathrm{E}[T_i \mid \overline{M}, d_i, d_j]\mathrm{E}[T_j \mid \overline{M}, d_i, d_j] + \mathrm{E}[T_{i2}^2 - (\mathrm{E}[T_{i2}|\overline{M}, d_i, d_j])^2|\overline{M}, d_i, d_j]. \qquad (3.7)$$

Now, note that

$$\mathrm{E}[T_i \mid M, d_i, d_j] = \mathrm{E}[T_i \mid \overline{M}, d_i, d_j] = p\binom{d_i}{2}. \qquad (3.8)$$

To see this, given the degree d_i of ith node along with d_j and M, observe that there are $\binom{d_i}{2}$ possible triangles containing node i, and one just needs to count the number of edges between

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

the pairs of neighbors of i to find T_i. Hence, there are $\binom{d_i}{2}$ candidates for triangles containing node i, each of which is present with probability p, independent of others. Thus, one has

$$\mathrm{E}[T_i \mid M, d_i, d_j] = p\binom{d_i}{2}.$$

Noting that given the degree of d_i, the presence or absence of the edge between node i and j does not change the distribution of T_i, we similarly have

$$\mathrm{E}[T_i \mid \overline{M}, d_i, d_j] = p\binom{d_i}{2}.$$

Combining our observations in (3.5), (3.6), (3.7) and (3.8), we obtain,

$$\begin{aligned}
\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] = {} & p\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}p^2\binom{d_i}{2}\binom{d_j}{2} \mid M\right] \\
& + (1-p)\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}p^2\binom{d_i}{2}\binom{d_j}{2} \mid \overline{M}\right] \\
& + p\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}\mathrm{E}[T_{i1}^2 - (\mathrm{E}[T_{i1}|M, d_i, d_j])^2|M, d_i, d_j] \mid M\right] \\
& + p\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}\mathrm{E}[T_{i2}^2 - (\mathrm{E}[T_{i2}|M, d_i, d_j])^2|M, d_i, d_j] \mid M\right] \\
& + (1-p)\mathrm{E}\left[\frac{\mathbf{1}(A_i)\mathbf{1}(A_j)}{\binom{d_i}{2}\binom{d_j}{2}}\mathrm{E}[T_{i2}^2 - (\mathrm{E}[T_{i2}|\overline{M}, d_i, d_j])^2|\overline{M}, d_i, d_j] \mid \overline{M}\right].
\end{aligned}$$

The first two terms on the right-hand side simplify to

$$p^2\left(p\mathrm{E}[\mathbf{1}(A_i)\mathbf{1}(A_j) \mid M] + (1-p)\mathrm{E}\left[\mathbf{1}(A_i)\mathbf{1}(A_j) \mid \overline{M}\right]\right),$$

which is equal to $p^2\mathrm{E}\left[\mathbf{1}(A_i)\mathbf{1}(A_j)\right]$. In addition, all the other terms contain conditional variances of T_{i1} and T_{i2}, which are bound to be non-negative random variables and, hence, these terms are non-negative. Thus, we conclude that

$$\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] \geq p^2\mathrm{E}\left[\mathbf{1}(A_i)\mathbf{1}(A_j)\right].$$

In addition, again recalling from the proof of Proposition 3.1 that $\mathrm{P}(\overline{A_i}) \leq M_1 e^{-M_2 n^{1/3}}$ for some constants M_1, M_2 independent of n, and noting a similar bound holds for A_j, we have $\mathrm{P}(A_i \cap A_j) \geq 1 - D/n^2$ for some constant D independent of n. Using this, we reach at

$$\mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] \geq p^2 - \frac{D}{n^2},$$

where D is again independent of n.

Lastly, recalling

$$\mathrm{E}[C(i)C(j)] = \mathrm{E}[C(i)C(j)\mathbf{1}(A_i)\mathbf{1}(A_j)] + S,$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

where $0 \le S \le D/n^2$, we get

$$\mathrm{E}[C(i)C(j)] \ge p^2 - \frac{D_2}{n^2},$$

as asserted. □

4. Degree Index in Erdős–Rényi Graphs

As noted earlier, the degree index (or total degree irregularity) has been introduced and studied previously, but we are not aware of a relevant analysis for random graphs. This section is devoted to such an analysis, and this turns out to be more tractable compared with the clustering index case due to the lack of $\binom{d_i}{2}$ terms in the denominator. We begin the discussion by studying $\mathrm{E}[\mathrm{DI}_2(\mathcal{G})]$.

Theorem 4.1. *For an Erdős–Rényi graph $\mathcal{G}$ with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$,*

$$\mathrm{E}[\mathrm{DI}_2(\mathcal{G})] = 6\binom{n}{3}p(1-p).$$

In particular,

$$\frac{\mathrm{E}[\mathrm{DI}_2(\mathcal{G})]}{6\binom{n}{3}} = p(1-p),$$

which is independent of n.

Proof. Let $\mathcal{G}$ be an Erdős–Rényi graph with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$. We are interested in calculating

$$\mathrm{E}\big[(d_i - d_j)^2\big] = 2\mathrm{E}\big[d_i^2\big] - 2\mathrm{E}[d_i d_j].$$

Let $k = n - 2$ for convenience. Denoting the binomial distribution by Bin, since $d_i \sim \mathrm{Bin}(n-1, p) = \mathrm{Bin}(k+1, p)$, we have

$$\mathrm{E}[d_i^2] = (\mathrm{E}[d_i])^2 + \mathrm{var}(d_i) = (n-1)^2p^2 + (n-1)p(1-p) = (k+1)^2p^2 + (k+1)p(1-p).$$

Next, let us examine $\mathrm{E}[d_i d_j]$. Let M denote the event that e_{ij} appears in the graph. Conditioning on M and $\overline{M}$ yields

$$\mathrm{E}[d_i d_j] = p\mathrm{E}[d_i d_j \mid M] + (1-p)\mathrm{E}[d_i d_j \mid \overline{M}].$$

If we assume that e_{ij} does not appear in the graph, then the number of possible neighbors of i or j is now $n-2$ instead of $n-1$. In addition, since we are in the Erdős–Rényi model, the presence of any other edge is independent of the presence of e_{ij}. Moreover, apart from whether e_{ij} appears or not, d_i and d_j are independent. By the above observations, $d_i \mid M$ and $d_j \mid M$ are independent and identically distributed (i.i.d.) with the common distribution $1 + \mathrm{Bin}(n-2, p) = 1 + \mathrm{Bin}(k, p)$. Similarly, $d_i \mid \overline{M}$ and $d_j \mid \overline{M}$ are i.i.d. this time with the common distribution $\mathrm{Bin}(n-2, p) = \mathrm{Bin}(k, p)$. Therefore, we have

$$\begin{aligned}\mathrm{E}[d_i d_j] &= p\mathrm{E}[d_i d_j \mid M] + (1-p)\mathrm{E}[d_i d_j \mid \overline{M}] = p(kp+1)^2 + (1-p)(kp)^2 \\ &= k^2p^3 + 2kp^2 + p + k^2p^2 - k^2p^3 = k^2p^2 + 2kp^2 + p,\end{aligned}$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

which implies

$$\begin{aligned}\mathrm{E}[d_i^2] - \mathrm{E}[d_i d_j] &= (k+1)^2p^2 + (k+1)p(1-p) - k^2p^2 - 2kp^2 - p \\ &= k^2p^2 + 2kp^2 + p^2 + kp - kp^2 + p - p^2 - k^2p^2 - 2kp^2 - p \\ &= kp - kp^2 = kp(1-p) = (n-2)p(1-p).\end{aligned}$$

Thus,

$$\mathrm{E}[(d_i - d_j)^2] = 2(n-2)p(1-p).$$

Recalling that $\mathrm{DI}_2(\mathcal{G}) = \sum_{1\le i<j\le n} |d_i - d_j|^2$, result follows. □

Next, we focus on $\mathrm{DI}_1(\mathcal{G})$. In this case a direct computation readily gives

$$\begin{aligned}\mathrm{E}[\mathrm{DI}_1(\mathcal{G})] &= \mathrm{E}\left[\sum_{1\le i<j\le n} |d_i - d_j|\right] \\ &= \sum_{\ell=0}^{n-2}\sum_{k=0}^{n-2} |k-\ell| \binom{n-1}{k}\binom{n-1}{\ell} p^k(1-p)^{n-1-k}p^\ell(1-p)^{n-1-\ell}.\end{aligned}$$

For clearer expressions, we analyze the symmetric $p = 1/2$ case and the general p case separately.

Keeping the notation as above, let us now observe that

$$\mathrm{E}[|d_i - d_j|] = p\mathrm{E}[|d_i - d_j| \mid M] + (1-p)\mathrm{E}[|d_i - d_j| \mid \overline{M}].$$

Note that the distribution of $d_i - d_j$ is the same regardless of whether the edge between i and j, e_{ij}, is present in the graph or not. And since the only dependence between d_i and d_j was a result of e_{ij}, we have

$$\mathrm{E}[|d_i - d_j|] = \mathrm{E}[|B_1 - B_2|],$$

where B_1, B_2 are independent binomial random variables with parameters $n-2$ and p. For general p case, an easy upper bound on $\mathrm{E}[|d_i - d_j|]$ can be obtained as follows:

$$\begin{aligned}\mathrm{E}[|d_i - d_j|] = \mathrm{E}|B_1 - B_2| &\le \mathrm{E}|B_1 - (n-2)p| + \mathrm{E}|B_2 - (n-2)p| = 2\mathrm{E}|B_1 - (n-2)p| \\ &\le 2(\mathrm{E}|B_1 - (n-2)p|^2)^{1/2} = 2\sqrt{\mathrm{var}(B_1)} = 2\sqrt{(n-2)p(1-p)}.\end{aligned}$$

Improving this, one may obtain an exact asymptotic result for $\mathrm{E}[\mathrm{DI}_1(\mathcal{G})]$. This will follow from the following elementary result on the mean absolute difference of binomials.

Lemma 4.1. *Let X_n, Y_n be independent binomial random variables with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$.*

(i) *If $p = 1/2$, then*

$$\mathrm{E}|X_n - Y_n| = \frac{n\binom{2n}{n}}{2^{2n}}.$$

Moreover, $\mathrm{E}|X_n - Y_n| \sim \sqrt{n/\pi}$, as $n \to \infty$.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

(ii) *For general* $p \in (0, 1)$*, we have*

$$\left| \mathrm{E}|X_n - Y_n| - \frac{2}{\sqrt{\pi}}\sqrt{np(1-p)} \right| \leq D$$

for every $n \geq 1$*, where* D *is a constant depending on* p*, but not on* n*. In particular,* $\mathrm{E}|X_n - Y_n| \sim (2/\sqrt{\pi})\sqrt{np(1-p)}$*, as* $n \to \infty$*.*

Although the lemma is elementary and there are similar results in the literature (see, for example, [9]), we were unable to find the exact form used in our case. Therefore, we include the proof of the lemma in the appendix.

Now, given Lemma 4.1, the following result on DI_1 follows immediately.

Theorem 4.2. *(i) If* $\mathcal{G}$ *is an Erdős–Rényi graph with parameters* $n \in \mathbb{N}$ *and* $p = 1/2$*, then*

$$\mathrm{E}[\mathrm{DI}_1(\mathcal{G})] = \frac{(n-2)\binom{n}{2}\binom{2n-4}{n-2}}{2^{2n-4}}.$$

In particular, $\mathrm{E}[\mathrm{DI}_1(\mathcal{G})] \sim \sqrt{(n-2)/\pi}\binom{n}{2}$*, as* $n \to \infty$*.*

(ii) If $\mathcal{G}$ *is an Erdős–Rényi graph with parameters* $n \in \mathbb{N}$ *and* $p \in (0, 1)$*, then*

$$\mathrm{E}[\mathrm{DI}_1(\mathcal{G})] \sim \frac{2}{\sqrt{\pi}}\binom{n}{2}\sqrt{(n-2)p(1-p)}$$

as $n \to \infty$*.*

Remark 4.1. One may consider the normalized versions of the degree indices we have studied here in order to be able to do comparisons among distinct random graphs models in a clear way and to make meaningful inferences about the graph properties. Motivated by Theorems 4.1 and 4.2, a natural candidate for the scaled version of the statistic we study is

$$\mathrm{DI}^*_\alpha = \frac{\mathrm{DI}_\alpha(\mathcal{G})}{n^{2+\alpha/2}}.$$

Note that in the case of Erdős–Rényi graphs, as $n \to \infty$, $\mathrm{E}[\mathrm{DI}^*_\alpha]$ converges to $\sqrt{p(1-p)/\pi}$ and $p(1-p)$ for α equals 1 and 2, respectively.

In addition, a second natural possible normalization for $\mathrm{DI}_\alpha(\mathcal{G})$ could be a scaling by $\binom{n}{2}$, where n is the number of nodes. However, such a choice does not take α into account, and the scaled limits do not tend to constant values.

5. Monte Carlo Simulations for Other Random Graph Models

The calculations for the degree and clustering indices become more involved when one leaves the framework of Erdős–Rényi random graphs and deals with other models. In this section, we focus on three other random graph models and do a Monte Carlo simulation study for these. The models we are interested in are the Watts–Strogatz model, Barabási–Albert model, and, lastly, random regular graphs. The degree index in random regular graphs is already zero, and one is only interested in the clustering index for this special case.

5.1. Random graph models and implementation

Here, instead of providing a detailed technical background on the random graph models of interest, we briefly discuss the practical aspects of the implementation of Monte Carlo simulations. In particular, we focus on the selection of parameters so that the experiments are

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

performed with matching edge densities for distinct random graph models. Before moving further, let us note that we used the NetworkX module in Python for all the simulations performed in the following. Now, let us review the models under study, keeping in mind that we are willing to have a given edge density $p^* \in (0, 1)$ in our simulations.

1. *Erdős–Rényi model*. The function erdos_renyi_graph which has parameters n and p as discussed in previous sections produces an Erdős–Rényi graph in NetworkX. One just chooses $p = p^*$ in order to satisfy the edge density criterion.

2. *Watts–Strogatz model*. This model was introduced by Watts and Strogatz in 1998, and it catches certain common properties of real-life complex networks by having the small world property, high clustering, and short average path lengths [25]. The description of watts_strogatz_graph function in NetworkX reference page is as follows:

   > *First create a ring over n nodes. Then each node in the ring is joined to its k nearest neighbors (or $k - 1$ neighbors if is odd). Then shortcuts are created by replacing some edges as follows: for each edge (u, v) in the underlying 'n-ring with k nearest neighbors' with probability p (rewiring probability) replace it with a new edge (u,w) with uniformly random choice of existing node w.*

The number of edges in the Watts–Strogatz model is constant and is given by $nk/2$. Thus, in order to have the edge density p^*, one needs to have $p^* = (nk/2)/\binom{n}{2}$. Simplifying, we obtain $p^* = k/(n-1)$. Thus, given p^* and n, one needs to choose k properly so that this criterion is satisfied.

3. *Barabási–Albert model*. This model, which was introduced by Albert and Barabási in 2002, is a random graph model [4] that uses preferential attachment (i.e. higher-degree vertices tend to have even higher degrees) in order to generate scale-free networks. In NetworkX, samples from this model are generated via the barabasi_albert_graph(n, m, seed=None, initial_graph=None) function. The descriptions of the inputs in NetworkX references page are as follows:

   - n int Number of nodes;
   - m int Number of edges to attach from a new node to existing nodes;
   - Initial_graph Graph or None (default) Initial network for Barabási–Albert algorithm; it should be a connected graph for most use cases; a copy of initial_graph is used; if None, starts from a star graph on $(m + 1)$ nodes.

In our case, we use the default as the initial graph and so begin with a star on $m + 1$ nodes. Now, given the edge density p^* and the number of nodes n, our goal is to select m so that the resulting Barabási–Albert graph will have approximate edge density p^*.

Now, since we begin with $m + 1$ nodes, the steps required to reach n nodes is $N = n - (m + 1)$. Then the number of edges inserted throughout these steps is $Nm = (n - (m + 1))m$. Thus, also considering the initial nodes present, at the end of the evolution of the graph, we have

$$(n - (m + 1))m + m$$

many edges, and so the edge density is

$$\frac{(n - (m + 1))m + m}{\binom{n}{2}} = \frac{nm - m^2}{\frac{n(n-1)}{2}}.$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

Hence, we would like to have

$$\frac{nm - m^2}{\frac{n(n-1)}{2}} \approx p^*,$$

or, rearranging,

$$2m^2 - 2nm + n(n-1)p^* \approx 0.$$

Now the corresponding Δ is $4n^2 - (4)(2)n(n-1)p^* = 4n^2 - 8n(n-1)p^*$. Since $\Delta < 0$ when p^* is large for the number of nodes we intend to simulate, we will only consider the cases $p^* = 0.1$ and $p^* = 0.5$.

When $\Delta > 0$, the root we use is given by

$$m^* = \frac{2n - \sqrt{4n^2 - 8n(n-1)p^*}}{4}. \tag{5.1}$$

For example, when $n = 200$ and $p^* = 0.1$, placing these in the given formula, we find $m^* = 10.501$, which is rounded to 11. Note that for fixed p^*, m^* depends linearly on n. In particular, when $p^* = 0.5$, $m^* \approx n/2$ and when $p^* = 0.1$, $m^* \approx n/20$.

4. *Random regular graphs*. Random regular graphs are generated using random_regular_graph(d, n) function in NetworkX. Here, n is the number of nodes and d is the common degree. In this case, to achieve an edge density p^*, one needs to have

$$p^* = \frac{\frac{nd}{2}}{\binom{n}{2}},$$

which can be rewritten as $p^* = d/(n-1)$. Thus, given n and p^*, we choose d appropriately to achieve the given edge density.

5.2. Results for the clustering index

Now we are ready to present the simulation results. In each fixed selection of parameters, 120–600 sample graphs of each model were generated for varying node counts, and these were averaged to approximate the value of interest. For each plot, the number of nodes was chosen from the set $\{20, 40, \ldots, 380\}$, and the resulting averages were linearly interpolated. Note that the rewiring probability in the Watts–Strogatz model does not affect the number of vertices nor the edge density; hence, in each of our simulations, we examined Watts–Strogatz graphs with rewiring probabilities of 0.1, 0.3, 0.5, 0.7, and 0.9. In many of our simulations, Barabási–Albert graph and some of the Watts–Strogatz graphs had growth rates significantly larger than the others; thus, we have created additional plots to better examine the remaining graph models in all our plots.

Figure 4 contains our results for CI_1. Note that these plots are for the Barabási–Albert model where the parameter $m = m^*$ is chosen according to (5.1) in order to satisfy the edge density condition. In particular, this is in contrast to the standard Barabási–Albert model where m is a fixed positive integer. One would naturally obtain a distinct growth rate for the standard case. We give a more detailed relevant discussion later when we focus on the degree index.

Our expectation from our theoretical findings was that the growth of CI_1 for Erdős–Rényi graphs was sublinear in n, which is verified in our simulation. One would also expect that,

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

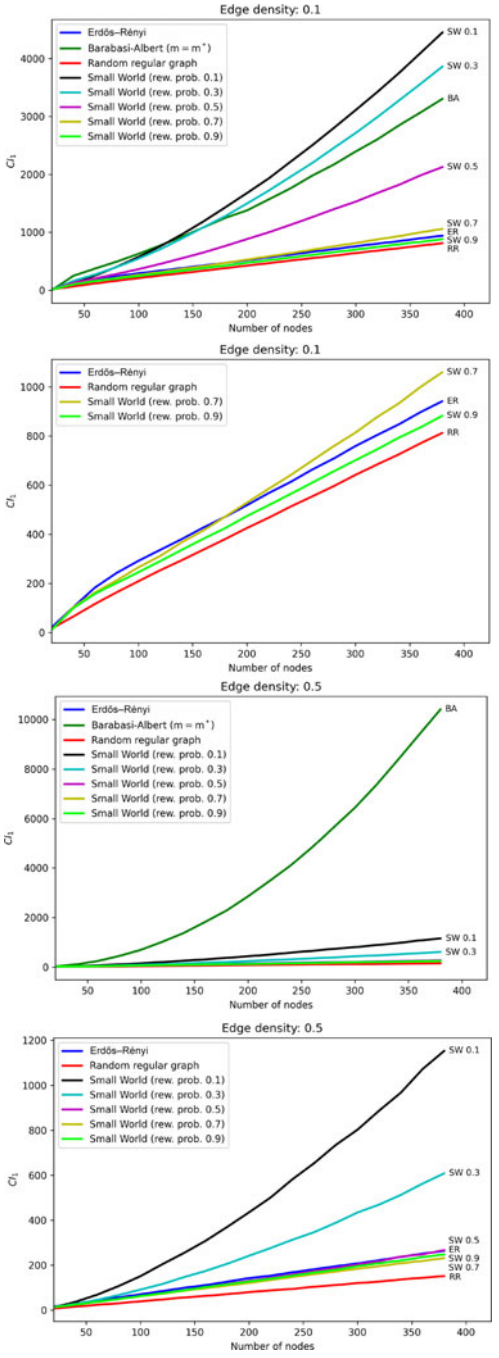


FIGURE 4. Average CI_1 values of the Erdős–Rényi graph, Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge densities 0.1 and 0.5.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

as the rewiring probability increases, the Watts–Strogatz graphs exhibit increasingly similar behavior to Erdős–Rényi graphs in terms of clustering, which seems to be the case.

In Figure 5, we include the comparison of CI_2 for the four models of interest. In the case of Erdős–Rényi, we showed that this index is upper bounded by a constant independent of the number of nodes and we conjecture that it is indeed also lower bounded by a constant. This is indeed the case according to our simulations, where CI_2 can be observed to approach $2(1-p)(1-p^2)/p$ as n increases. Watts–Strogatz graphs with low rewiring probability and the Barabási–Albert model again dominate the others in this case. In addition, as in the case of CI_1, we observe that as the rewiring probability increases, CI_2 of the Watts–Strogatz model approaches that of Erdős–Rényi model.

It is also interesting in Figure 5 to note that many of our random graphs seem to exhibit a peaking behavior around 40–60 nodes when the edge density is 0.1. We think that the appearance of these peaks is a result of an equilibrium of two opposing effects:

- as n increases, the number of terms appearing in the sum CI_2 also increases;
- as n increases, the variance of the distribution of local clustering coefficients decreases.

When the edge density is 0.5, the Barabási–Albert model with $m = m^*$ again dominates the others; it seems to exhibit $\Theta(n^2)$ growth while the others are sublinear. The Watts–Strogatz model with lower rewiring probabilities again follows the Barabási–Albert case in the ranking.

For the Barabási–Albert model with $m = m^*$, when the edge density is 0.5, both CI_1 and CI_2 seem to be of order n^2. This is caused by the vastly different clustering behavior of the leaves of the initial star graph and the other nodes. Since $m \approx n/2$ in this case, the initial star graph has $m - 1 \approx n/2 - 1$ leaves. Our simulation results show that, when n is large, the local clustering coefficients of these leaves are significantly higher compared with the clustering coefficients of the other nodes on average. This causes $\mathrm{E}[CI_1] \geq K_1 \binom{n}{2}$ and $\mathrm{E}[CI_2] \geq K_2 \binom{n}{2}$ for positive constants K_1, K_2.

5.3. Results for the degree index

In this subsection, we focus on the degree index and compare DI_1 and DI_2 for the same four random graph models again for varying values of the number of nodes.

Figure 6 is for DI_1 and the edge density 0.1. Note that again there are two versions corresponding to each edge density due to the domination of the Barabási–Albert model with $m = m^*$. In particular, those on the right do not contain the case for the Barabási–Albert model. Figure 7 is for DI_1 and the edge density 0.5.

We see that as in the case of the clustering index, the small world model behaves similarly to Erdős–Rényi as the rewiring probability increases. In order to examine the behaviors for different models in more detail, we also include the log–log plots for the graphs on the left in Figures 6 and 7.

The log–log plot clearly indicates that the growth of $\mathrm{E}[DI_1]$ for the preferential attachment model we consider is of higher order than the growth of other cases. The simulations we have and the heuristic arguments we use suggest that the growth rate in this case is of order $\Theta(n^3)$, which is strictly larger than that for Erdős–Rényi graphs (which is $\Theta(n^{5/2})$). However, note that the Barabási–Albert model we consider is different from the standard Barabási–Albert model in the following sense. Here, in order to have a fixed edge density for comparison among distinct models, the corresponding parameter m increases as a function of n as explained in

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

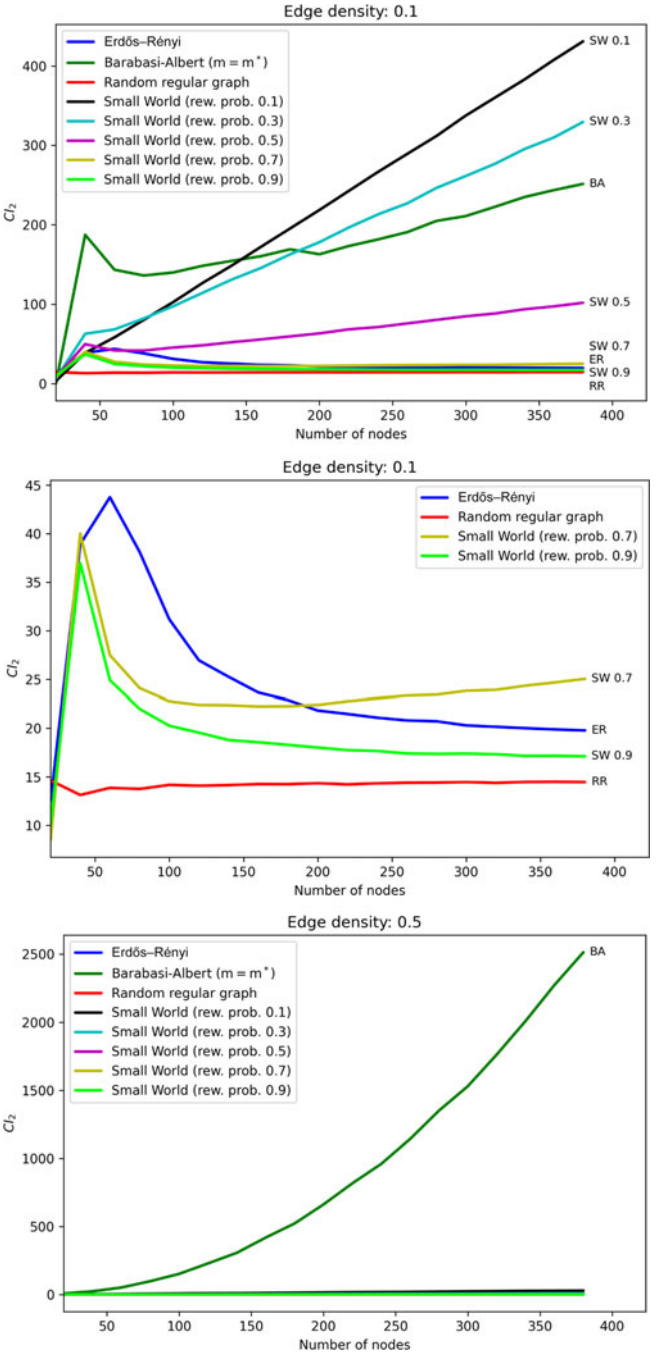


FIGURE 5. Average CI_2 values of the Erdős–Rényi graph, Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge densities 0.1 and 0.5.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

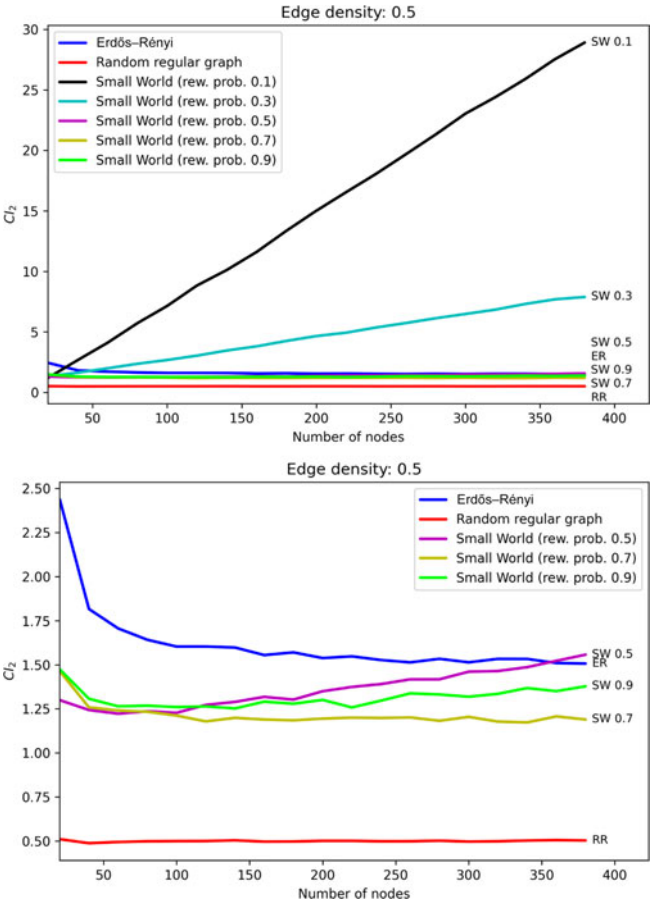


FIGURE 5. Continued.

Section 5.1; see (5.1). Thus, if we consider the standard Barabási–Albert model where m is fixed, then we would obtain a different behavior. Indeed, we were informed by an anonymous reviewer that the maximal degree in the standard Barabási–Albert model is of order $\sqrt{n}$; see, for example, [13, p. 280]. From this, one immediately obtains a trivial upper bound of order $n^{5/2}$ for DI_1 for fixed m. To have a better understanding, we also simulated the standard Barabási–Albert model. Figure 9 contains the corresponding log–log plot.

The plots in Figure 9 and the related slope calculations suggest that $\mathrm{E}[\mathrm{DI}_1] = \Theta(n^2)$ when m is fixed. This seems natural since in this case the degrees of most nodes tend to be small and close to each other, which then causes DI_1 to be small. On the other hand, in the Barabási–Albert model with $m = m^*$, (5.1) gives $m^* \sim Cn$ for some $C \in (0, 1)$. In turn, this causes a significant number of nodes to have degrees comparable to n and a remarkable number of nodes to have significantly smaller degrees, with the result that $\mathrm{E}[|d_i - d_j|]$ is of order n and $\mathrm{E}[\mathrm{DI}_1]$ is of order n^3.

For the DI_2 case, the simulation results are summarized in Figures 10 and 11. The observations here are similar to the DI_1 case, and we do not provide the details of these to avoid repetition.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

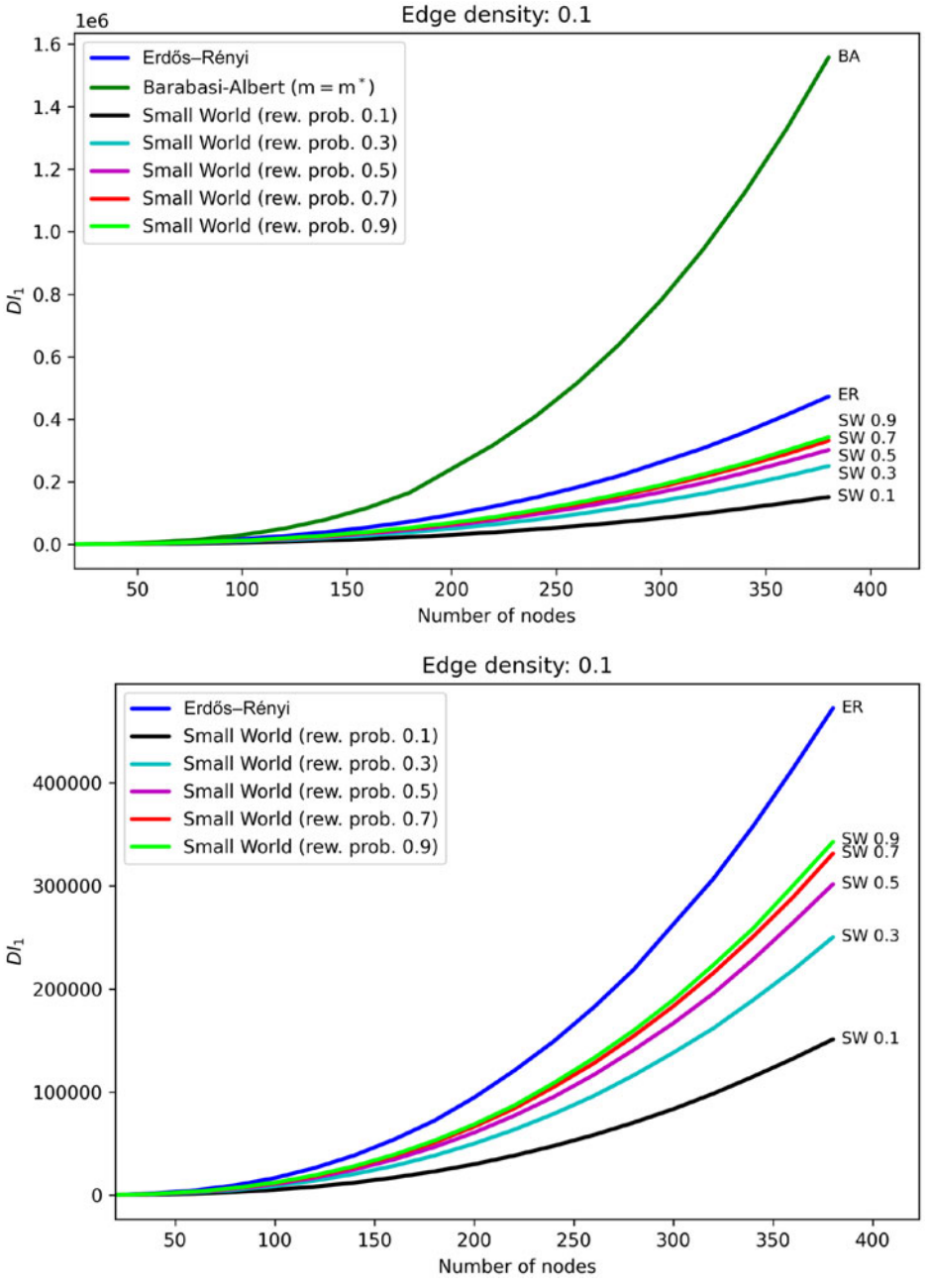


FIGURE 6. Average DI_1 values of the Erdős–Rényi graph, Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge density 0.1.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

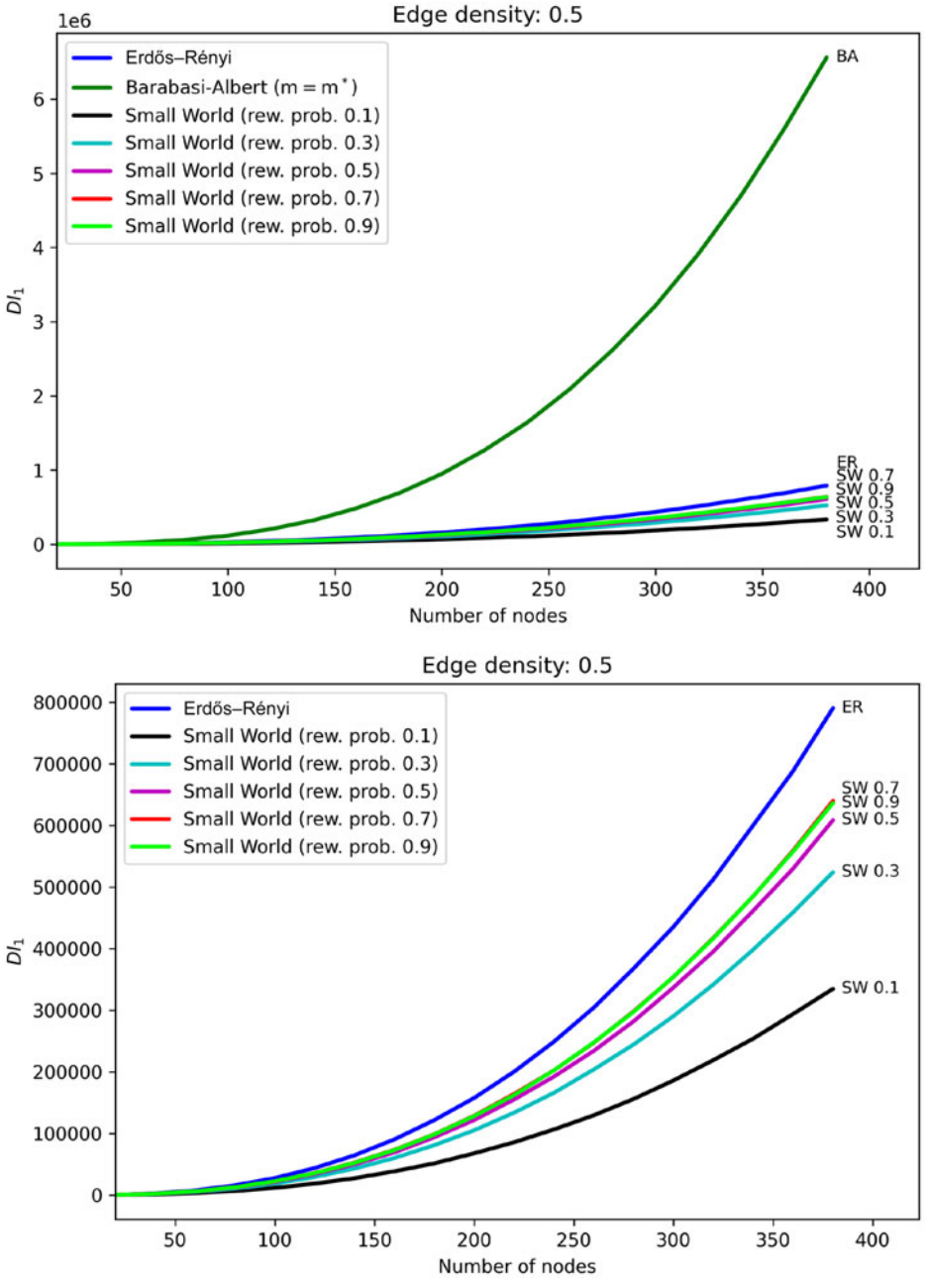


FIGURE 7. Average DI_1 values of the Erdős–Rényi graph, Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge density 0.5.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

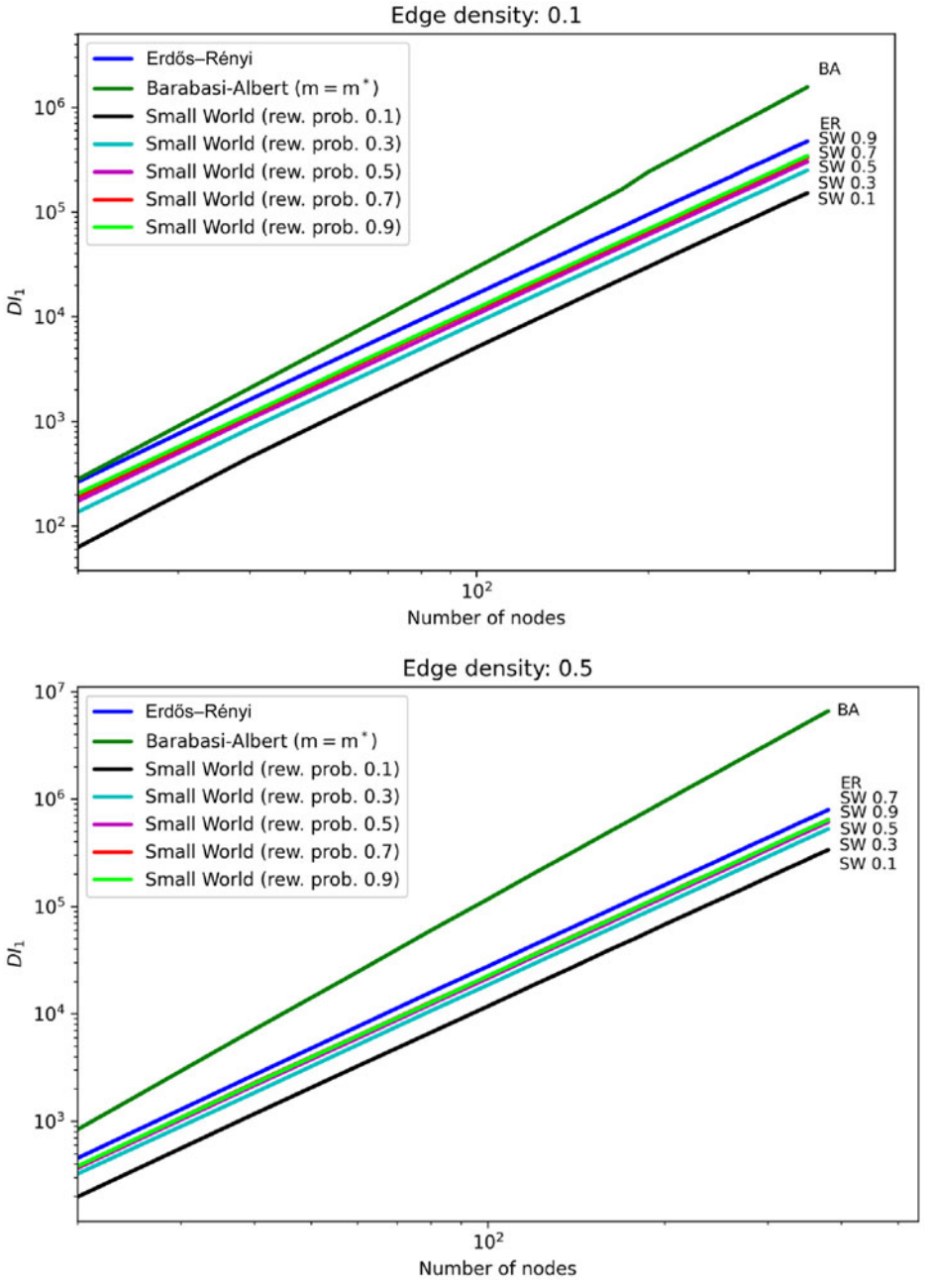


FIGURE 8. The log–log plots for the DI_1 values of the Erdős–Rényi graph, modified Barabási–Albert graph, and Watts–Strogatz graphs with edge densities 0.1 and 0.5.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

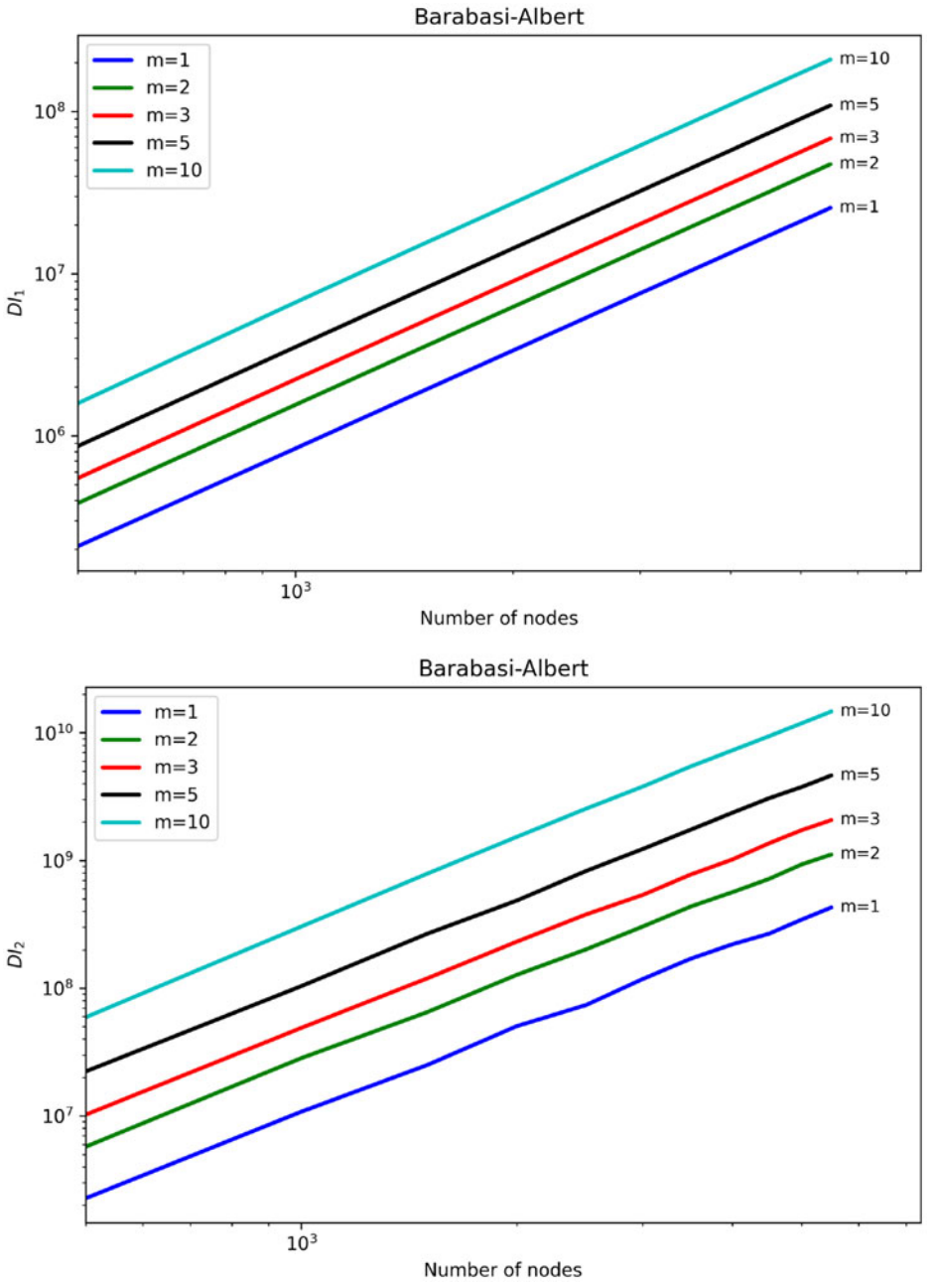


FIGURE 9. The log–log plots for the DI_1 and DI_2 values of the Barabási–Albert graphs with $m \in \{1, 2, 3, 5, 10\}$, where m denotes the number of edges to attach from a new node to existing nodes. The number of nodes is from 500 to 6000, with increments of 500.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

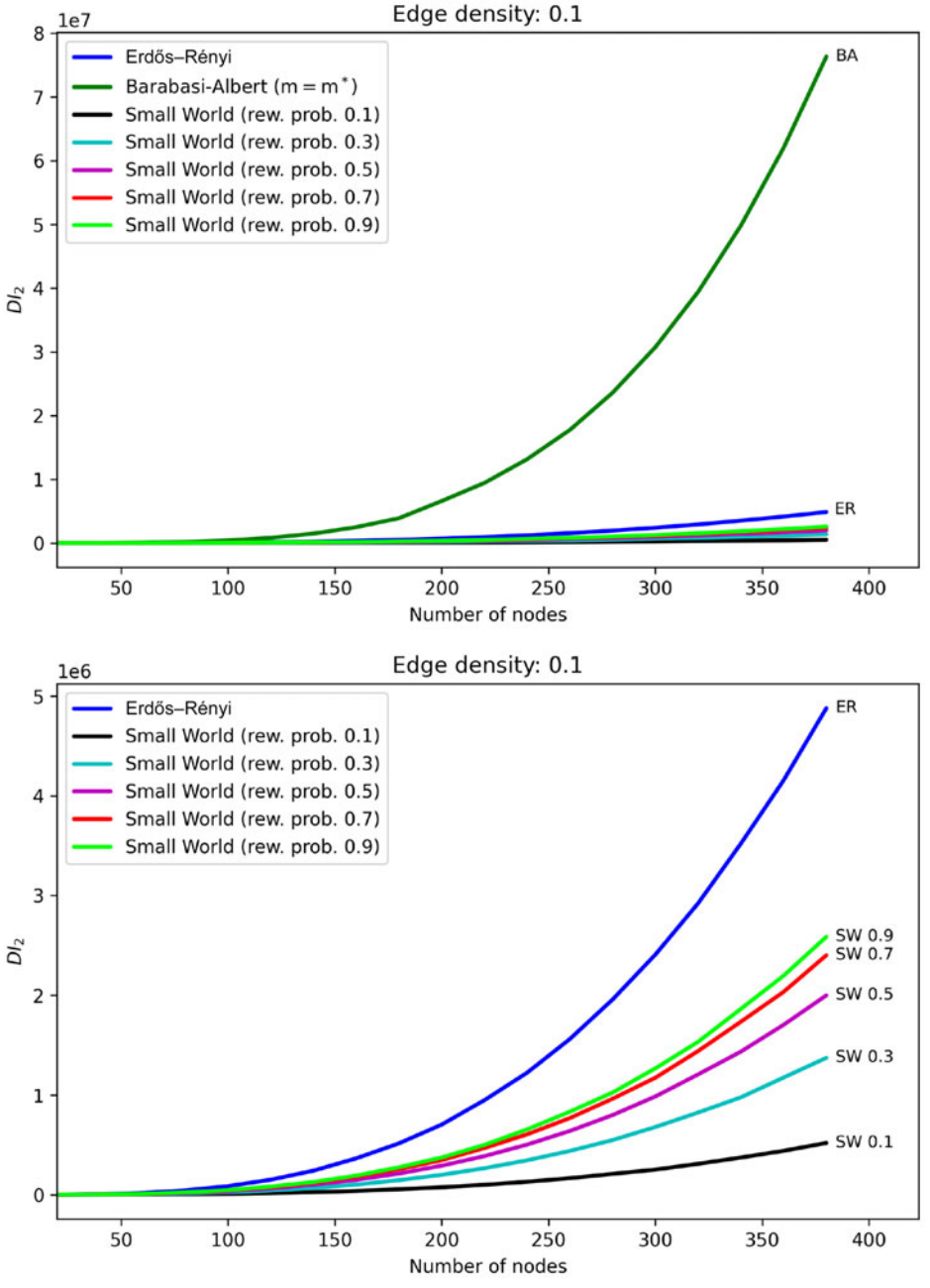


FIGURE 10. Average DI_2 values of the Erdős–Rényi graph, modified Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge density 0.1.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

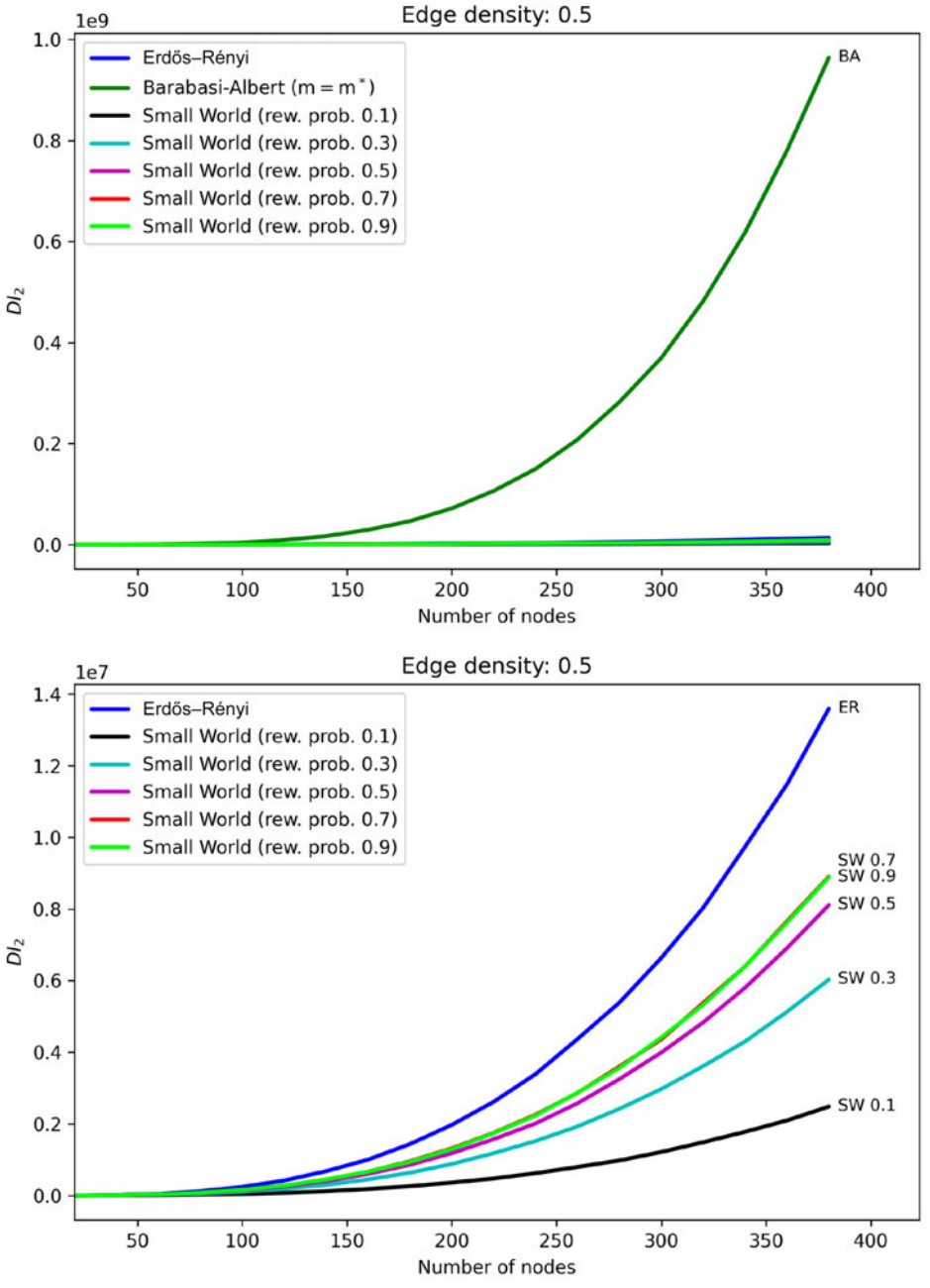


FIGURE 11. Average DI_2 values of the Erdős–Rényi graph, modified Barabási–Albert graph, random regular graph, and Watts–Strogatz graphs with edge density 0.5.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

We close this section by noting that we also analyzed the log–log plot for the case DI_2 for the Barabási–Albert model with $m = m^*$, and that the related observations suggest $E[DI_2] = \Theta(n^4)$.

6. Conclusion

In this paper, in addition to the well-known degree index (or irregularity), we have introduced the concept of the clustering index and studied the first moment of both in a random graph framework. Focusing on Erdős–Rényi graphs and the degree case, we were able to obtain either exact or asymptotic expressions for the expected degree indices of interest. Regarding the clustering index case, we have obtained a linear upper bound for $E[CI_1(\mathcal{G})]$, and also have shown that $E[CI_2(\mathcal{G})]$ is bounded above by a constant independent of n. Although we had a theoretical and simulation-based related analysis for random graphs, we have not done a detailed study of the indices of interest for real-life complex networks and we intend to do so in a subsequent work.

One main motivation for studying these indices in our case is their possible use in artificial intelligence classification algorithms as features. Characteristics such as average clustering and the average degree are already used in various such problems, and we intend to do an experimental study in future work to see whether the use of degree and clustering indices provides improvements in performance in classification problems. A second related path is in the field of finance, more particularly in the detection of financial crises. In a recent work [14], it was argued that a certain graph irregularity measure, namely the Laplace energy, can be used to detect economic crises. Recently, we did some preliminary experiments to see whether the indices introduced in this paper could also be considered for such purposes. The initial analysis indicated that both are useful in this direction, but that the results for degree index are more statistically significant. We plan to conduct a rigorous analysis on whether degree and clustering indices (as certain types of network irregularities) can also be considered as a precursor of financial crises in upcoming work.

The last direction we are willing to follow is in computing the exact asymptotics of CI_1 and CI_2 in the case of Erdős–Rényi graphs. For example, focusing on CI_2, we were able to obtain a constant upper bound, and simulations suggested that this index indeed behaves like a constant as the number of nodes increases. It would be interesting to verify this convergence theoretically.

Appendix: Proof of Lemma 4.1.

(i) Let X_n, Y_n be independent binomial random variables with parameters n and $1/2$. We claim that $E|X_n - Y_n| = n\binom{2n}{n}/2^{2n}$. For this purpose, letting Z_n be an independent binomial random variable with parameters n and $1/2$, observe that

$$|X_n - Y_n| =_d |X_n - n + Z_n| = |X_n + Z_n - n| = |T_n - n|,$$

where $T_n = X_n + Z_n$ is a binomial random variable with parameters $2n$ and $1/2$. Then,

$$P(|X_n - Y_n| = k) = P(|T_n - n| = k) = 2P(T_n = n + k) = \binom{2n}{n+k}\frac{1}{2^{2n-1}}.$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

We have

$$\begin{aligned}
\mathrm{E}|X_n - Y_n| = \sum_{k=0}^{n} k\binom{2n}{n+k}\frac{1}{2^{2n-1}} &= \frac{1}{2^{2n-1}}\sum_{k=1}^{n}(k-n+n)\binom{2n}{n+k} \\
&= \frac{1}{2^{2n-1}}\sum_{k=1}^{n}(n+k)\binom{2n}{n+k} - \frac{1}{2^{2n-1}}\sum_{k=1}^{n} n\binom{2n}{n+k} \\
&= \frac{2n}{2^{2n-1}}\sum_{k=1}^{n}\binom{2n-1}{n+k-1} - \frac{n}{2^{2n-1}}\sum_{k=1}^{n}\binom{2n}{n+k} \\
&= \frac{2n}{2^{2n-1}}\frac{2^{2n-1}}{2} - \frac{n}{2^{2n-1}}\left(\frac{2^n - \binom{2n}{n}}{2}\right) = \frac{n\binom{2n}{n}}{2^{2n}}.
\end{aligned}$$

The assertion that $\mathrm{E}|X_n - Y_n| \sim \sqrt{\frac{n}{\pi}}$, $n \to \infty$, now follows from Stirling's approximation.

(ii) Let Z_1, Z_2 be independent standard normal random variables. We denote the Wasserstein distance between the distributions by d_W. Now, letting $\mu_n = \mathrm{E}[X_n]$ and $\sigma_n^2 = \mathrm{var}(X_n)$, recall that

$$\begin{aligned}
d_W\left(\frac{X_n - \mu_n}{\sigma_n}, Z_1\right) := \sup\left\{\mathrm{E}\left|h\left(\frac{X_n - \mu_n}{\sigma_n}\right) - h(Z_1)\right| : h : \mathbb{R} \to \mathbb{R}, |h(x) - h(y)|\right. \\
\left. \leq |x - y|\right\} \leq \frac{D}{\sqrt{n}},
\end{aligned}$$

for some constant D independent of n (See, for example, [21]). We in particular have

$$\mathrm{E}\left|\frac{X_n - \mu_n}{\sigma_n} - Z_1\right| \leq \frac{D}{\sqrt{n}}.$$

With previous observations, and together with triangle inequality, this gives

$$\begin{aligned}
\mathrm{E}|X_n - Y_n| &= \sigma_n\left|\frac{X_n - \mu_n}{\sigma_n} - Z_1 + Z_1 - Z_2 + Z_2 - \frac{Y_n - \mu_n}{\sigma_n}\right| \\
&\leq 2\sigma_n\mathrm{E}\left|\frac{X_n - \mu_n}{\sigma_n} - Z_1\right| + \sigma_n\mathrm{E}|Z_1 - Z_2| \\
&\leq \frac{2C\sigma_n}{\sqrt{n}} + \sigma_n\mathrm{E}|Z_1 - Z_2| = \mathcal{O}(1) + \sigma_n\mathrm{E}|Z_1 - Z_2|.
\end{aligned}$$

In addition, using the triangle inequality in an appropriate way,

$$\begin{aligned}
\mathrm{E}|X_n - Y_n| &= \sigma_n\left|\left(\frac{X_n - \mu_n}{\sigma_n} - Z_1\right) + (Z_1 - Z_2) + \left(Z_2 - \frac{Y_n - \mu_n}{\sigma_n}\right)\right| \\
&= \sigma_n\mathrm{E}|a + b + c| \geq \sigma_n\mathrm{E}|b| - \sigma_n\mathrm{E}|a| - \sigma_n\mathrm{E}|c| \\
&= \sigma_n\mathrm{E}|Z_1 - Z_2| - 2\sigma_n\mathrm{E}\left|\frac{X_n - \mu_n}{\sigma_n} - Z_1\right| \\
&\geq \sigma_n\mathrm{E}|Z_1 - Z_2| - \frac{2\sigma_n D}{\sqrt{n}} = \mathcal{O}(1) + \sigma_n\mathrm{E}|Z_1 - Z_2|.
\end{aligned}$$

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

Thus, $(1/\sigma_n)\mathrm{E}|X_n - Y_n|$ converges to $\mathrm{E}|Z_1 - Z_2|$. The result now follows since $\mathrm{E}|Z_1 - Z_2| = 2/\sqrt{\pi}$.

Acknowledgements

We would like to thank the anonymous reviewers whose suggestions and corrections improved the paper significantly. Parts of this paper were completed at the Nesin Mathematics Village, the authors would like to thank Nesin Mathematics Village for their kind hospitality.

Funding information

The first author was supported by the BAGEP Award of the Science Academy, Turkey.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Abdo, H., Brandt, S. and Dimitrov, D. (2014). The total irregularity of a graph. *Discrete Mathematics & Theoretical Computer Science* **16**, 201–206.
[2] Abdo, H., Dimitrov, D. and Gutman, I. (2018). Graphs with maximal σ irregularity. *Discrete Applied Mathematics* **250**, 57–64.
[3] Abdo, H., Dimitrov, D. and Gutman, I. (2019). Graph irregularity and its measures. *Applied Mathematics and Computation* **357**, 317–324.
[4] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* 74.1, 47–97.
[5] Albertson, M. O. (1997). The irregularity of a graph. *Ars Combinatoria* **46**, 219–225.
[6] Ali, A., Chartrand, G. and Zhang, P. (2021). *Irregularity in Graphs*. Springer, Cham.
[7] Borgatti, S. P. and Everett, M. G. (2000). Models of core/periphery structures. *Social Networks* 21.4, 375–395.
[8] Chen, M., Li, N., Zheng, L., Huang, D. and Wu, B. (2022). Dynamic correlation of market connectivity, risk spillover and abnormal volatility in stock price. *Physica A: Statistical Mechanics and its Applications* **587**, 126506.
[9] Diaconis, P. and Zabell, S. L. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284–302.
[10] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publicationes Mathematicae (Debrecen)* **6**, 290–297.
[11] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–61.
[12] Gu, L., Huang, H. L. and Zhang, X. D. (2013). The clustering coefficient and the diameter of small-world networks. *Acta Mathematica Sinica, English Series* **29**, 199–208.
[13] van der Hofstad, R. (2024). *Random Graphs and Complex Networks*, Vol. 2. Cambridge University Press, Cambridge.
[14] Huang, C., Deng, Y., Yang, X., Yang, X. and Cao, J. (2023). Can financial crisis be detected? Laplacian energy measure. *The European Journal of Finance* **29**, 949–976.
[15] Kantarci, B. and Labatut, V. (2013). Classification of complex networks based on topological properties. In: *2013 International Conference on Cloud and Green Computing (CGC)*. IEEE, 297–304.
[16] Kartun-Giles, A. P. and Bianconi, G. (2019). Beyond the clustering coefficient: A topological analysis of node neighbourhoods in complex networks. *Chaos, Solitons & Fractals: X* **1**, 100004.
[17] Li, Y., Shang, Y. and Yang, Y. (2017). Clustering coefficients of large networks. *Information Sciences* **382**, 350–358.
[18] Soffer, S. N. and Vázquez, A. (2005). Network clustering coefficient without degree-correlation biases. *Physical Review E* **71**, 057101.
[19] McDiarmid, C. (1998). Concentration. In: Habib, M., McDiarmid, C., Ramírez-Alfonsín, J. L. and Reed, B. (eds.) Probabilistic Methods for Algorithmic Discrete Mathematics. Springer, Berlin, 195–248.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press

[20] RADICCHI, F., CASTELLANO, C., CECCONI, F., LORETO, V. AND PARISI, D. (2004). Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA* **101**, 2658–2663.
[21] ROSS, N. (2011). Fundamentals of Stein's method. *Probability Surveys* **8**, 210–293.
[22] SIEW, C. S. Q., WULFF, D. U., BECKAGE, N. M. AND KENETT, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, 2108423.
[23] SILVA, T. C. AND ZHAO, L. (2016). *Machine Learning in Complex Networks*. Springer, Cham.
[24] WEST, D. B. (2001). *Introduction to Graph Theory*, 2nd edn. Prentice Hall, Upper Saddle River, NJ.
[25] WATTS, D. J. AND STROGATZ, S. H. (1998). Collective dynamics of small-world networks. *Nature* **393**, 440–442.
[26] YIN, H., BENSON, A. R. AND LESKOVEC, J. (2018). Higher-order clustering in networks. *Physical Review E* **97**, 052306.
[27] YIN, H., BENSON, A. R., LESKOVEC, J. AND GLEICH, D. F. (2017). Local higher-order graph clustering. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17).* ACM, 555–564.

https://doi.org/10.1017/jpr.2025.10063 Published online by Cambridge University Press