

LIMIT THEOREMS FOR ASSORTATIVITY AND CLUSTERING IN NULL MODELS FOR SCALE-FREE NETWORKS

REMCO VAN DER HOFSTAD,* *Eindhoven University of Technology*

PIM VAN DER HOORN,* ** *Eindhoven University of Technology and Northeastern University*

NELLY LITVAK,* *** *Eindhoven University of Technology and University of Twente*

CLARA STEGEHUIS,*** *University of Twente*

Abstract

An important problem in modeling networks is how to generate a randomly sampled graph with given degrees. A popular model is the configuration model, a network with assigned degrees and random connections. The erased configuration model is obtained when self-loops and multiple edges in the configuration model are removed. We prove an upper bound for the number of such erased edges for regularly-varying degree distributions with infinite variance, and use this result to prove central limit theorems for Pearson's correlation coefficient and the clustering coefficient in the erased configuration model. Our results explain the structural correlations in the erased configuration model and show that removing edges leads to different scaling of the clustering coefficient. We prove that for the rank-1 inhomogeneous random graph, another null model that creates scale-free simple networks, the results for Pearson's correlation coefficient as well as for the clustering coefficient are similar to the results for the erased configuration model.

Keywords: Degree–degree correlations; clustering; configuration model; limit theorems

2010 Mathematics Subject Classification: Primary 05C80

Secondary 60F05

1. Introduction and results

1.1. Motivation

The configuration model (CM) [6, 43] is an important null model used to generate graphs with a given degree sequence, by assigning each node a number of half-edges equal to its degree and connecting stubs at random to form edges. Conditioned on the resulting graph being simple, its distribution is uniform over all graphs with the same degree sequence [13]. Because of this feature the CM is widely used to analyze the influence of degrees on other properties or processes on networks [11, 14, 15, 25, 32, 38].

An important property that many networks share is that their degree distributions are regularly varying, with the exponent γ of the degree distribution satisfying $\gamma \in (1, 2)$, so that the degrees have infinite variance. In this regime of degrees, the CM results in a simple graph with vanishing probability. To still be able to generate simple graphs with approximately the desired degree distribution, the erased configuration model (ECM) removes self-loops and

Received 14 June 2018; revision received 30 April 2020.

* Postal address: Department of Mathematics and Computer Science, Eindhoven University of Technology.

** Postal address: Department of Physics, Northeastern University, Boston. Email address: w.l.f.v.d.hoorn@tue.nl

*** Postal address: Department of Electrical Engineering, Mathematics and Computer Science, University of Twente.

multiple edges of the CM [7], while the empirical degree distribution still converges to the original one [13].

The degree distribution is a first-order characteristic of the network structure, since it is independent of the way nodes are connected. An important second-order network characteristic is the correlation between degrees of connected nodes, called degree–degree correlations or network assortativity. A classical measure for these correlations computes Pearson’s correlation coefficient on the vector of joint degrees of connected nodes [29, 30]. In the CM, Pearson’s correlation coefficient tends to zero in the large graph limit [33], so that the CM is only able to generate networks with neutral degree correlations.

The CM creates networks with asymptotic neutral degree correlations [33]. By this we mean that, as the size of the network tends to infinity, the joint distribution of degrees on both sides of a randomly sampled edge factorizes as the product of the size-biased distributions. As a result, the outcome of any degree–degree correlation measure converges to zero. Although one would expect fluctuations of such measures to be symmetric around zero, it has frequently been observed that constraining a network to be simple results in so-called structural negative correlations [8, 22, 40, 44], where the majority of measured degree–degree correlations are negative, while still converging to zero in the infinite graph limit. This is most prominent in the case where the variance of the degree distribution is infinite. To investigate the extent to which the edge removal procedure of the ECM results in structural negative correlations, we first characterize the scaling of the number of edges that have been removed. Such results are known when the degree distribution has finite variance [1, 26, 27]. However, for scale-free distributions with infinite variance only some preliminary upper bounds have been proven [23]. Here we prove a new upper bound and obtain several useful corollaries. Our result improves the one in [23] while strengthening [20, Theorem 8.13]. We then use this bound on the number of removed edges to investigate the consequences of the edge removal procedure on Pearson’s correlation coefficient in the ECM. We prove a central limit theorem, which shows that the correlation coefficient in the ECM converges to a random variable with negative support when properly rescaled. Thus, our result confirms the existence of structural correlations in simple networks theoretically.

We then investigate a ‘global’ clustering coefficient, which is the number of triangles divided by the number of triplets connected by two edges, eventually including multiple edges; see (3) and (4) for the precise definition. The clustering coefficient measures the tendency of sets of three vertices to form a triangle. In the CM, the clustering coefficient tends to zero whenever the exponent of the degree distribution satisfies $\gamma > 4/3$, whereas it tends to infinity for $\gamma < 4/3$ in the infinite graph limit [31]. In this paper, we obtain more detailed results on the behavior of the clustering coefficient in the CM in the form of a central limit theorem. We then investigate how the edge removal procedure of the ECM affects the clustering coefficient and obtain a central limit theorem for the clustering coefficient in the ECM.

Interestingly, it was shown in [34, 35] that in simple graphs with $\gamma \in (1, 2)$ the clustering coefficient converges to zero. This again shows that constraining a graph to be simple may significantly impact network statistics. We obtain a precise scaling for the clustering coefficient in ECM, which is sharper than the general upper bound in [34].

We further show that the results on Pearson’s correlation coefficient and the clustering coefficient for the ECM can easily be extended to another important random graph null model for simple scale-free networks: the rank-1 inhomogeneous random graph [5, 9]. In this model, every vertex is equipped with a weight w_i , and vertices are connected independently with some connection probability $p(w_i, w_j)$. We show that for a wide class of connection probabilities,

the rank-1 inhomogeneous random graph also has structurally negative degree correlations, satisfying the same central limit theorem as in the ECM. Furthermore, we show that for the particular choice $p(w_i, w_j) = 1 - e^{-w_i w_j / (\mu n)}$, where μ denotes the average weight, the clustering coefficient behaves asymptotically the same as in the ECM.

In the (erased) configuration model as well as the inhomogeneous random graph, Pearson's correlation coefficient for degree correlations and the global clustering coefficient naturally converge to zero. We would like to emphasize that this paper improves on the existing literature by establishing the *scaling laws* that govern the convergence of these statistics to zero. This is important because very commonly in the literature, various quantities measured in real-world networks are compared to null-models with same degrees but random rewiring. These rewired null-models are similar to a version of the inhomogeneous random graph [2, 12]. Without knowing the scaling of these quantities in the inhomogeneous random graph, it is not possible to assess how similar a small measured value on the real network is to that of the null model. Our results enable such analysis. In fact, we do even more, by also establishing exact limiting distributions of the rescaled Pearson's correlation coefficient and clustering coefficient, which are the most standard measures in statistical analysis of networks.

1.2. Outline of the paper

The remainder of the paper is structured as follows. In the next three sections we formally introduce the models, the measures of interest, and some additional notation. Then, in Section 1.7, we summarize our main results and discuss important insights obtained from them. We give a heuristic outline of our proof strategy in Section 2 and recall several results for regularly-varying degrees. Then we proceed with proving our result for Pearson's correlation coefficient in Section 3 and the clustering coefficient in Section 4. We then show in Section 5 how the proofs for Pearson's correlation coefficient and the clustering coefficient in the ECM can be adapted to prove the central limit theorems for the rank-1 inhomogeneous random graph. Finally, Appendix A contains the proof of Theorem 2.1 on the number of erased edges, as well as some additional technical results.

1.3. Configuration model with scale-free degrees

The first models of interest in this work are the configuration model and the erased configuration model. Given a vertex set $[n] := \{1, 2, \dots, n\}$ and a sequence $\mathbf{D}_n = \{D_1, D_2, \dots, D_n\}$ whose sum $\sum_{i \in [n]} D_i$ is even, the configuration model (CM) constructs a graph G_n with this degree sequence by assigning D_i stubs to each node i and then connecting stubs at random to form edges. This procedure will, in general, create a multi-graph with self-loops and multiple edges between two nodes. To make sure that the resulting graph is simple we can remove all self-loops and replace multiple edges between nodes by just one edge. This model is called the erased configuration model (ECM).

We will denote by $\text{CM}(\mathbf{D}_n)$ and $\text{ECM}(\mathbf{D}_n)$ graphs generated by, respectively, the standard and erased configuration model, starting from the degree sequence \mathbf{D}_n . We often couple the two constructions by first constructing a graph via the standard configuration model and then removing all the self-loops and multiple edges to create the erased configuration model. In this case we write G_n for the graph created by the CM and \widehat{G}_n for the ECM graph constructed from G_n . In addition we use the hats to distinguish between objects in the CM and the ECM. For example, D_i denotes the degree of node i in the graph $\text{CM}(\mathbf{D}_n)$, while \widehat{D}_i denotes its degree in $\text{ECM}(\mathbf{D}_n)$.

We consider degree sequences $\mathbf{D}_n = \{D_1, D_2, \dots, D_n\}$ where the degrees D_i are independent and identically distributed (i.i.d.) copies of an integer-valued random variable \mathcal{D} with regularly-varying distribution

$$\mathbb{P}(\mathcal{D} > t) = \mathcal{L}(t)t^{-\gamma}, \quad \gamma > 1. \tag{1}$$

Here, the function $\mathcal{L}(t)$ is slowly varying at infinity and γ is the exponent of the distribution.

As is common in the literature, D_n may include a correction term, equal to one, in order to make the sum $L_n = \sum_{i \in [n]} D_i$ even. We shall ignore this correction term, since it does not affect the asymptotic results. In the remainder of this paper \mathcal{D} always refers to a random variable with distribution (1).

1.4. Rank-1 inhomogeneous random graphs

Another model that generates networks with scale-free degrees is the rank-1 inhomogeneous random graph [5, 9]. In this model, every vertex is equipped with a weight. We assume these weights are an i.i.d. sample from the scale-free distribution (1). Then, vertices i and j with weights w_i and w_j are connected with some connection probability $p(w_i, w_j)$. Let the expected value of (1) be denoted by μ . We then assume the following conditions on the connection probabilities, similarly to [16].

Condition 1.1. (Class of connection probabilities.) Assume that

$$p(w_i, w_j) = \frac{w_i w_j}{\mu n} h\left(\frac{w_i w_j}{\mu n}\right),$$

for some continuous function $h : [0, \infty) \mapsto [0, 1]$ with the following properties:

- (i) $h(0) = 1$ and $h(u)$ decreases to 0 as $u \rightarrow \infty$.
- (ii) $q(u) = uh(u)$ increases to 1 as $u \rightarrow \infty$.
- (iii) There exists $u_1 > 0$ such that $h(u)$ is differentiable on $(0, u_1]$ and $h'_+(0) < \infty$.

This class includes the commonly used connection probabilities $q(u) = (u \wedge 1)$, where $(x \wedge y) := \min(x, y)$ (the Chung Lu setting) [9], $q(u) = 1 - e^{-u}$ (the Poisson random graph) [33], and $q(u) = u/(1 + u)$ (the maximal entropy random graph) [10, 21, 36]. Note that within the class of connection probabilities satisfying Condition 1.1, $q(u) \leq (u \wedge 1)$. Note that $p(w_i, w_j) = q\left(\frac{w_i w_j}{\mu n}\right)$.

1.5. Central quantities

Pearson’s correlation coefficient $r(G_n) \in [-1, 1]$ is a measure for degree–degree correlations. For an undirected multigraph G_n , this measure is defined as

$$r(G_n) = \frac{\sum_{i,j \in [n]} X_{ij} D_i D_j - \frac{1}{L_n} \left(\sum_{i \in [n]} D_i^2\right)^2}{\sum_{i \in [n]} D_i^3 - \frac{1}{L_n} \left(\sum_{i \in [n]} D_i^2\right)^2}, \tag{2}$$

where X_{ij} denotes the number of edges between nodes i and j in G_n , and self-loops are counted twice (see [33]). We write r_n for Pearson’s correlation coefficient on G_n generated by CM and \widehat{r}_n if G_n is generated by ECM.

The clustering coefficient of a graph G_n is defined as

$$C(G_n) = \frac{3\Delta_n}{\text{number of connected triples}}, \tag{3}$$

where Δ_n denotes the number of triangles in the graph. The clustering coefficient can be written as

$$C(G_n) = \frac{6\Delta_n}{\sum_{i \in [n]} D_i(D_i - 1)} = \frac{6 \sum_{1 \leq i < j < k \leq n} X_{ij}X_{jk}X_{ik}}{\sum_{i \in [n]} D_i(D_i - 1)}, \tag{4}$$

where X_{ij} again denotes the number of edges between vertices i and j in G_n . For simple graphs, $C(G_n) \in [0, 1]$. However, for multigraphs, $C(G_n)$ may exceed 1. As with Pearson’s correlation coefficient, we denote by C_n the clustering coefficient in G_n generated by CM, while \widehat{C}_n is the clustering coefficient in G_n generated by ECM.

1.6. Notation

We write \mathbb{P}_n and \mathbb{E}_n for, respectively, the conditional probability and expectation with respect to the sampled degree sequence \mathbf{D}_n . We use \xrightarrow{d} for convergence in distribution and $\xrightarrow{\mathbb{P}}$ for convergence in probability. We say that a sequence of events $(\mathcal{E}_n)_{n \geq 1}$ happens with high probability (w.h.p.) if $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_n) = 1$. Furthermore, we write $f(n) = o(g(n))$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ if $|f(n)|/g(n)$ is uniformly bounded, where $(g(n))_{n \geq 1}$ is nonnegative.

We say that $X_n = O_{\mathbb{P}}(g(n))$ for a sequence of random variables $(X_n)_{n \geq 1}$ if $|X_n|/g(n)$ is a tight sequence of random variables, and $X_n = o_{\mathbb{P}}(g(n))$ if $X_n/g(n) \xrightarrow{\mathbb{P}} 0$. Finally, we use $(x \wedge y)$ to denote the minimum of x and y and $(x \vee y)$ to denote the maximum of x and y .

1.7. Results

In this paper we study the interesting regime when $1 < \gamma < 2$, so that the degrees have finite mean but infinite variance. When $\gamma > 2$, the number of removed edges is constant in n , and hence asymptotically there will be no difference between the CM and ECM. We establish a new asymptotic upper bound for the number of erased edges in the ECM and prove new limit theorems for Pearson’s correlation coefficient and the clustering coefficient. We further show that the limit theorems for Pearson and clustering for the inhomogeneous random graph are very similar to the ones obtained for the ECM.

Our limit theorems involve random variables with joint stable distributions, which we define as follows. Let

$$\Gamma_i = \sum_{j=1}^i \xi_j, \quad i \geq 1, \tag{5}$$

with $(\xi_j)_{j \geq 1}$ being i.i.d. exponential random variables with mean 1. Then we define, for any integer $p \geq 2$,

$$\mathcal{S}_{\gamma/p} = \sum_{i=1}^{\infty} \Gamma_i^{-p/\gamma}. \tag{6}$$

We remark that for any $\alpha > 1$ we have that $\sum_{i=1}^{\infty} \Gamma_i^{-\alpha}$ has a stable distribution with stability index α (see [39, Theorem 1.4.5]).

In the remainder of this section we will present the theorems and highlight their most important aspects in view of the methods and current literature. We start with \widehat{r}_n .

Theorem 1.1. (Pearson in the ECM.) *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, if $G_n = \text{ECM}(\mathbf{D}_n)$, there exists a slowly-varying function \mathcal{L}_1 such that*

$$\mu \mathcal{L}_1(n) n^{1-\frac{1}{\gamma}} \widehat{r}_n \xrightarrow{d} -\frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}},$$

where $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ are given by (6).

The following theorem shows that the correlation coefficient for all rank-1 inhomogeneous random graphs satisfying Condition 1.1 behaves the same as in the ECM.

Theorem 1.2. (Pearson in the rank-1 inhomogeneous random graph.) *Let \mathbf{W}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, when G_n is a rank-1 inhomogeneous random graph with weights \mathbf{W}_n and connection probabilities satisfying Condition 1.1, there exists a slowly-varying function \mathcal{L}_1 such that*

$$\mu \mathcal{L}_1(n) n^{1-\frac{1}{\gamma}} r(G_n) \xrightarrow{d} -\frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}},$$

where $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ are given by (6).

Interestingly, the behavior of Pearson’s correlation coefficient in the rank-1 inhomogeneous random graph does not depend on the exact form of the connection probabilities, as long as these connection probabilities satisfy Condition 1.1.

Asymptotically vanishing correlation coefficient. It has been known for some time (c.f. [33, Theorem 3.1]) that when the degrees \mathbf{D}_n are sampled from a degree distribution with infinite third moment, any limit point of Pearson’s correlation coefficient is non-positive. Theorem 1.1 confirms this, showing that for the ECM, with infinite second moment, the limit is zero. Moreover, Theorem 1.1 gives the exact scaling in terms of the graph size n , which has not been available in the literature. Compare e.g. to [20, Theorem 5.1], where only the scaling of the negative part of \widehat{r}_n is given.

Structural negative correlations. It has also been observed many times that imposing the requirement of simplicity on graphs gives rise to so-called structural negative correlations; see e.g. [8, 22, 40, 44]. Our result is the first theoretical confirmation of the existence of structural negative correlations as a result of the simplicity constraint on the graph. To see this, note that the distributions of the random variables $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ have support on the positive real numbers. Therefore, Theorem 1.1 shows that when we properly rescale Pearson’s correlation coefficient in the ECM, the limit is a random variable whose distribution only has support on the negative real numbers. This result implies that when multiple instances of ECM graphs are generated and Pearson’s correlation coefficient is measured, the majority of the measurements will yield negative, although small, values. These small values have nothing to do with the network structure but are an artifact of the constraint that the resulting graph be simple. Interestingly, Theorem 1.2 shows that the same result holds for rank-1 inhomogeneous random graphs, indicating that structural negative correlations also exist in these models and thus further supporting the explanation that such negative correlations result from the constraint that the graphs be simple.

Pearson in ECM versus CM. Currently we only have a limit theorem for the erased model, in the scale-free regime $1 < \gamma < 2$. Interestingly, and also somewhat unexpectedly, proving

a limit theorem for CM, which is a simpler model, turns out to be more involved. The main reason for this is that in the ECM, the positive part of $r(G_n)$, determined by $\sum_{i,j} X_{ij}D_iD_j$, is negligible with respect to the other term, since the number of edges removed is polynomial in n (see Section 2.3 for more details). Therefore, the negative part determines the distribution of the limit. In the CM this is no longer true, and hence the distribution is determined by the tricky balance between the positive and the negative term and their fluctuations. Analyzing this requires more involved methods than we have been able to develop so far. Below, we state a conjecture about this case, as well as a partial result concerning the scaling of its variance that supports the conjecture.

Conjecture 1.1. (Scaling Pearson for CM.) *As $n \rightarrow \infty$, there exists some random σ^2 such that*

$$\sqrt{nr_n} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{7}$$

The intuition behind this conjecture is explained in Section 2.4. Although we do not have a proof of this scaling limit of r_n in the CM, the following result shows that at least $\sqrt{nr_n}$ is a tight sequence of random variables.

Lemma 1.1. (Convergence of $n\text{Var}_n(r_n)$ for CM.) *As $n \rightarrow \infty$, with Var_n denoting the conditional variance given the i.i.d. degrees,*

$$n\text{Var}_n(r_n) \xrightarrow{d} \frac{2 - \mathcal{S}_{\gamma/6}/\mathcal{S}_{\gamma/3}^2}{\mu}. \tag{8}$$

We now present our results for the clustering coefficient. The following theorem gives a central limit theorem for the clustering coefficient in the CM.

Theorem 1.3. (Clustering in the CM.) *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, if $G_n = \text{CM}(\mathbf{D}_n)$, there exists a slowly-varying function \mathcal{L}_2 such that*

$$\frac{C_n}{\mathcal{L}_2(n)n^{4/\gamma-3}} \xrightarrow{d} \frac{1}{\mu^3} \left(K_{\gamma/2}\mathcal{S}_{\gamma/2}^2 - 3K_{\gamma/4}\mathcal{S}_{\gamma/4} + \frac{2K_{\gamma/6}\mathcal{S}_{\gamma/6}}{K_{\gamma/2}\mathcal{S}_{\gamma/2}} \right), \tag{9}$$

where $\mathcal{S}_{\gamma/2}$, $\mathcal{S}_{\gamma/4}$ and $\mathcal{S}_{\gamma/6}$ are given by (6) and

$$K_\alpha = \left(\frac{1 - \alpha}{\Gamma(2 - \alpha) \cos(\pi\alpha/2)} \right)^\alpha,$$

with Γ denoting the gamma function.

Infinite clustering. For $\gamma < 4/3$, Theorem 1.3 shows that C_n tends to infinity. This observation shows that the global clustering coefficient may give nonphysical behavior when used on multi-graphs. In multi-graphs, several edges may close a triangle. In this case, the interpretation of the clustering coefficient as a fraction of connected triples does not hold. Rather, the clustering coefficient can be interpreted as the average number of closed triangles per wedge, where different wedges and triangles may involve the same nodes but have different edges between them. This interpretation shows that indeed in a multi-graph the clustering coefficient may go to infinity.

What is a triangle? The result in Theorem 1.3 depends on what we consider to be a triangle. In general, one can think of a triangle as a loop of length three. In the CM, however, self-loops

and multiple edges may be present. Then, for example, three self-loops at the same vertex also form a loop of length three. Similarly, a multiple edge between vertices v and w together with a self-loop at vertex w can also form a loop of length three. In Theorem 1.3, we do not consider these cases as triangles. Excluding these types of ‘triangles’ gives the terms $\mathcal{S}_{\gamma/4}$ and $\mathcal{S}_{\gamma/6}/\mathcal{S}_{\gamma/2}$ in Theorem 1.3.

To obtain the precise asymptotic behavior of the clustering coefficient in the ECM, we need an extra assumption on the degree distribution (1).

Assumption 1.1. *The degree distribution (1) satisfies, for all $x \in \{1, 2, \dots\}$ and for some $K > 0$,*

$$\mathbb{P}(D = x) \leq K\mathcal{L}(x)x^{-\gamma-1}.$$

Note that for all $t \geq 2$,

$$\mathbb{P}(D = t) = \mathbb{P}(D > t - 1) - \mathbb{P}(D > t) = \mathcal{L}(t - 1)(t - 1)^{-\gamma} - \mathcal{L}(t)t^{-\gamma}.$$

Hence, since $(t - 1)^{-\gamma} - t^{-\gamma} \sim \gamma t^{-\gamma-1}$ as $t \rightarrow \infty$, it follows that Assumption 1.1 is satisfied whenever the slowly-varying function $\mathcal{L}(t)$ is monotonically increasing for all t greater than some T .

Theorem 1.4. (Clustering in the ECM.) *Let \mathbf{D}_n be sampled from \mathcal{D} , satisfying Assumption 1.1, with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, if $G_n = \text{ECM}(\mathbf{D}_n)$, there exists a slowly-varying function \mathcal{L}_3 such that*

$$\frac{\mathcal{L}_3(n)\widehat{C}_n}{\mathcal{L}(\sqrt{\mu n})^3 n^{(-3\gamma^2+6\gamma-4)/(2\gamma)}} \xrightarrow{d} -\gamma^3 \mu^{-\frac{3}{2}\gamma} \frac{\Gamma\left(-\frac{\gamma}{2}\right)^3}{2\mathcal{S}_{\gamma/2}},$$

where $\mathcal{S}_{\gamma/2}$ is a stable random variable defined in (6), and Γ denotes the gamma function.

We now investigate the behavior of the clustering coefficient in rank-1 inhomogeneous random graphs.

Theorem 1.5. (Clustering in the rank-1 inhomogeneous random graph.) *Let \mathbf{W}_n be sampled from \mathcal{D} , satisfying Assumption 1.1, with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, if G_n is an inhomogeneous random graph with weights \mathbf{W}_n and connection probabilities satisfying Condition 1.1, there exists a slowly-varying function \mathcal{L}_3 such that*

$$\frac{\mathcal{L}_3(n)\mathcal{C}(G_n)}{\mathcal{L}(\sqrt{\mu n})^3 n^{(-3\gamma^2+6\gamma-4)/(2\gamma)}} \xrightarrow{d} \mu^{-\frac{3}{2}\gamma} \frac{1}{\mathcal{S}_{\gamma/2}} \int_0^\infty \int_0^\infty \int_0^\infty \frac{q(xy)q(xz)q(yz)}{(xyz)^{\gamma+1}} dx dy dz,$$

where q is as in Condition 1.1(ii), $\mathcal{S}_{\gamma/2}$ is a stable random variable defined in (6), and

$$\int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{(xyz)^{\gamma+1}} q(xy)q(xz)q(yz) dx dy dz < \infty.$$

Maximal clustering in the ECM and the inhomogeneous random graph. Figure 1 shows the exponents of n in the main multiplicative term of the clustering coefficient, in the CM and the ECM. The exponent in Theorem 1.4 is a quadratic expression in γ ; hence there may be a value of γ that maximizes the clustering coefficient. We set the derivative of the exponent equal to zero,

$$\frac{d}{d\gamma}(-3\gamma^2 + 6\gamma - 4)/(2\gamma) = -3/2 + 2\gamma^{-2} = 0,$$

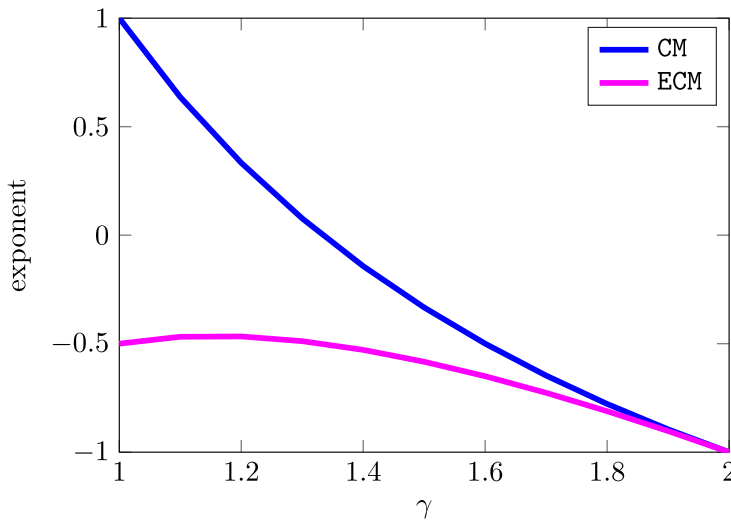


FIGURE 1: Exponents of the clustering coefficient of CM and ECM for $\gamma \in (1, 2)$.

which is solved by $\gamma = \sqrt{4/3} \approx 1.15$. Thus, the global clustering coefficient of an ECM with $\gamma \in (1, 2)$ is maximal for $\gamma \approx 1.15$, where the scaling exponent of the clustering coefficient equals $-2\sqrt{3} + 3 \approx -0.46$. This maximal value arises from the trade-off between the denominator and the numerator of the clustering coefficient in (3). When γ becomes close to 1, there will be some vertices with very high degrees. This makes the denominator of (3) very large. On the other hand, having more vertices of high degree also causes the graph to contain more triangles. Thus, the numerator of (3) also increases when γ decreases. The above computation shows that in the ECM, the optimal trade-off between the number of triangles and the number of connected triples is attained at $\gamma \approx 1.15$. Theorem 1.5 shows that the same phenomenon occurs in the rank-1 inhomogeneous random graph.

Mean clustering in CM versus ECM. In the CM, the normalized clustering coefficient converges to a constant times a stable random variable squared. This stable random variable has an infinite mean, and therefore its square also has an infinite mean. In the ECM as well as in the rank-1 inhomogeneous random graph, however, the normalized clustering coefficient converges to one divided by a stable random variable, which has a finite mean [37]. Thus, in the ECM and the rank-1 inhomogeneous random graph, the rescaled clustering coefficient converges to a random variable with finite mean. Formally,

$$\mathbb{E} \left[\frac{C_n}{n^{4/\gamma-3}} \right] = \infty$$

and

$$\mathbb{E} \left[\frac{\hat{C}_n}{n^{-3/2\gamma+3-2/\gamma}} \right] < \infty.$$

ECM and inhomogeneous random graphs. Theorems 1.4 and 1.5 show that the clustering coefficient in the ECM has the same scaling as the clustering coefficient in the rank-1 inhomogeneous random graph. In fact, choosing $q(u) = 1 - e^{-u}$ in Condition 1.1 even gives exactly

the same behavior for clustering in the ECM and in the inhomogeneous random graph. This shows that in terms of clustering, the ECM behaves similarly to an inhomogeneous random graph with connection probabilities $p(w_i, w_j) = 1 - e^{-w_i w_j / (\mu n)}$.

Vertices of degrees \sqrt{n} . In the proof of Theorem 1.4 we show that the main contribution to the number of triangles comes from vertices of degrees proportional to \sqrt{n} . Let us explain why this is the case. In the ECM, the probability that an edge exists between vertices i and j can be approximated by $1 - e^{-D_i D_j / L_n}$. Therefore, when $D_i D_j$ is proportional to n , the probability that an edge between i and j exists is bounded away from zero. Similarly, the probability that a triangle between vertices i, j , and k exists is bounded away from zero as soon as $D_i D_j$, $D_i D_k$, and $D_j D_k$ are all proportional to n . This is indeed achieved when all three vertices have degrees proportional to \sqrt{n} . If, for example, vertex i has degree of order larger than \sqrt{n} , this means that vertices j and k can have degrees of order smaller than \sqrt{n} while $D_i D_j$ and $D_i D_k$ are still of order n . However, $D_j D_k$ also has to be of size n for the probability of a triangle to be bounded away from zero. Now recall that the degrees follow a power-law distribution. Therefore, the probability that a vertex has degree much higher than \sqrt{n} is much smaller than the probability that a vertex has degree of order \sqrt{n} . Thus, the most likely way for all three contributions to be proportional to n is for D_i, D_j, D_k to be proportional to \sqrt{n} . Intuitively, this shows that the largest contribution to the number of triangles in the ECM comes from the vertices of degrees proportional to \sqrt{n} . This balancing of the number of vertices and the probability of forming a triangle also appears for other subgraphs [18].

Global and average clustering. Clustering can be measured by two different metrics: the global clustering coefficient and the average clustering coefficient [32, 41]. In this paper, we study the global clustering coefficient, as defined in (3). The average clustering coefficient is defined as the average over the local clustering coefficients of all the vertices, where the local clustering coefficient of a vertex is the number of triangles the vertex participates in divided by the number of pairs of neighbors of the vertex. For the CM, both the global clustering coefficient and the average clustering coefficient are known to scale as $n^{4/3-\gamma}$ [31]. In particular, this shows that both clustering coefficients in the CM diverge when $\gamma < 4/3$. Our main results, Theorems 1.3 and 1.4, provide the exact limiting behavior of the global clustering coefficients for CM and ECM, respectively.

The average clustering coefficient in the rank-1 inhomogeneous random graph has been shown to scale as $n^{1-\gamma} \log(n)$ [10, 16], which is very different from the scaling of the global clustering coefficient from Theorem 1.4. For example, the average clustering coefficient decreases in γ , whereas the global clustering coefficient first increases in γ and then decreases in γ (see Figure 1). Furthermore, the average clustering coefficient decays only very slowly in n as γ approaches 1. The global clustering coefficient, on the other hand, decays as $n^{-1/2}$ when γ approaches 1. This shows that the global clustering coefficient and the average clustering coefficient are two very different ways to characterize clustering.

Joint convergence. Before we proceed with the proofs, we remark that each of the three limit theorems uses a coupling between the sum of different powers of degrees and the limit distributions $\mathcal{S}_{\gamma/p}$. It follows from the proofs of our main results that these couplings hold simultaneously for all three measures. As a direct consequence, it follows that the rescaled measures converge jointly in distribution.

Theorem 1.6. (Joint convergence.) *Let \mathbf{D}_n be sampled from \mathcal{D} , satisfying Assumption 1.1, with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Let $G_n = \text{CM}(\mathbf{D}_n)$, $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$ and define $\alpha = (-3\gamma^2 + 6\gamma - 4)/2\gamma$. Then there exist slowly-varying functions $\mathcal{L}_1, \mathcal{L}_2$, and \mathcal{L}_3 such that as $n \rightarrow \infty$,*

$$\left(\mathcal{L}_1(n)n^{1-\frac{1}{\gamma}} \widehat{r}_n, \frac{C_n}{\mathcal{L}_2(n)n^{4/\gamma-3}}, \frac{\mathcal{L}_3(n)\widehat{C}_n}{\mathcal{L}(\sqrt{\mu n})^3 n^\alpha} \right) \xrightarrow{d} \left(-\frac{\mathcal{S}_{\gamma/2}^2}{\mu \mathcal{S}_{\gamma/3}}, \frac{\mathcal{S}_{\gamma/2}^2 - \mathcal{S}_{\gamma/4}}{\mu^3}, -\gamma^3 \mu^{-\frac{3}{2}\gamma} \frac{\Gamma(-\frac{\gamma}{2})^3}{2\mathcal{S}_{\gamma/2}} \right), \tag{10}$$

with $\mathcal{S}_{\gamma/2}, \mathcal{S}_{\gamma/3}$, and $\mathcal{S}_{\gamma/4}$ given by (6).

2. Overview of the proofs

Here we give an outline for the proofs of our main results for the CM and the ECM and explain the main ideas that lead to them. Since the goal is to convey the high-level ideas, we limit technical details in this section and often write $f(n) \approx g(n)$ to indicate that $f(n)$ behaves roughly as $g(n)$. The formal definition and exact details of these statements can be found in Sections 3 and 4 where the proofs are given. The proofs for rank-1 inhomogeneous random graphs follow very similar lines, and we show how the proofs for the ECM extend to rank-1 inhomogeneous random graphs satisfying Condition 1.1 in Section 5. We start with some results on the number of removed edges in the ECM.

2.1. The number of removed edges

The number of removed edges Z_n in the ECM is given by

$$Z_n = \sum_{i=1}^n X_{ii} + \sum_{1 \leq i < j \leq n} (X_{ij} - \mathbb{1}_{\{X_{ij} > 0\}}),$$

where X_{ij} again denotes the number of edges between vertices i and j . For the analysis of the ECM it is important to understand the behavior of this number. In particular we are interested in the scaling of Z_n with respect to n . Here we give an asymptotic upper bound, which implies that, up to sub-linear terms, Z_n scales no faster than $n^{2-\gamma}$. The proof can be found in Section A.1.

Theorem 2.1. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Then for any $\delta > 0$,*

$$\frac{Z_n}{n^{2-\gamma+\delta}} \xrightarrow{\mathbb{P}} 0.$$

The scaling $n^{2-\gamma}$ is believed to be sharp, up to some slowly-varying function. We therefore conjecture that for any $\delta > 0$,

$$\frac{Z_n}{n^{2-\gamma-\delta}} \xrightarrow{\mathbb{P}} \infty.$$

From Theorem 2.1 we obtain several useful results, summarized in the corollary below. The proof can be found in Section A.1. Let Z_{ij} be the number of edges between i and j that have been removed, and Y_i the number of removed stubs of node i . Then we have

$$Y_i = \sum_{j=1}^n Z_{ij} = X_{ii} + \sum_{j \neq i} (X_{ij} - \mathbb{1}_{\{X_{ij} > 0\}}).$$

Corollary 2.1. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $G_n = \text{ECM}(\mathbf{D}_n)$. Then, for any integer $p \geq 0$ and $\delta > 0$,*

$$\frac{\sum_{i=1}^n D_i^p Y_i}{n^{\frac{p}{\gamma}+2-\gamma+\delta}} \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{\sum_{1 \leq i < j \leq n} Z_{ij} D_i D_j}{n^{\frac{2}{\gamma}+2-\gamma+\delta}} \xrightarrow{\mathbb{P}} 0.$$

The first result of Corollary 2.1 gives the scaling of the difference of the sum of powers of degrees, between CM and ECM. To see why, note that since $\widehat{D}_i^q = (D_i - Y_i)^q$ and $Y_i \leq D_i$, by the mean value theorem we have, for any integer $q \geq 1$,

$$\left| \sum_{i=1}^n D_i^q - \sum_{i=1}^n \widehat{D}_i^q \right| \leq q \sum_{i=1}^n D_i^{q-1} Y_i.$$

Hence, for any $q \geq 1$ and $\delta > 0$,

$$\left| \sum_{i=1}^n D_i^q - \sum_{i=1}^n \widehat{D}_i^q \right| = o_{\mathbb{P}} \left(n^{\frac{q-1}{\gamma}+2-\gamma+\delta} \right).$$

2.2. Results for regularly-varying random variables

In addition to the number of edges, we shall make use of several results regarding the scaling of expressions with regularly-varying random variables. We summarize them here, starting with a concentration result on the sum of i.i.d. samples, which is a direct consequence of the Kolmogorov–Marcinkiewicz–Zygmund strong law of large numbers.

Lemma 2.1. *Let $(X_i)_{i \geq 1}$ be independent copies of a nonnegative regularly-varying random variable X with exponent $\gamma > 1$ and mean μ . Then, with $\kappa = (\gamma - 1)/(1 + \gamma)$,*

$$\frac{|\mu n - \sum_{i=1}^n X_i|}{n^{1-\kappa}} \xrightarrow{\mathbb{P}} 0.$$

In particular, since $L_n = \sum_{i=1}^n D_i$, with all D_i being i.i.d. with regularly-varying distribution (1) and mean μ , it holds that $n^{\kappa-1} |L_n - \mu n| \xrightarrow{\mathbb{P}} 0$. Therefore, the above lemma allows us to replace L_n with μn in our expressions.

The next proposition gives the scaling of sums of different powers of independent copies of a regularly-varying random variable. Recall that $(x \vee y)$ denotes the maximum of x and y .

Proposition 2.1. ([20, Proposition 2.4].) *Let $(X_i)_{i \geq 1}$ be independent copies of a nonnegative regularly-varying random variable X with exponent $\gamma > 1$. Then*

i) *for any integer $p \geq 1$ and $\delta > 0$,*

$$\frac{\sum_{i=1}^n X_i^p}{n^{\left(\frac{p}{\gamma} \vee 1\right) + \delta}} \xrightarrow{\mathbb{P}} 0;$$

ii) *for any integer $p \geq 1$ with $\gamma < p$ and $\delta > 0$,*

$$\frac{n^{\frac{p}{\gamma} - \delta}}{\sum_{i=1}^n X_i^p} \xrightarrow{\mathbb{P}} 0;$$

iii) for any integer $p \geq 1$ and $\delta > 0$,

$$\frac{\max_{1 \leq i \leq n} D_i^p}{n^{\frac{p}{\gamma} + \delta}} \xrightarrow{\mathbb{P}} 0.$$

Finally we have following lemma, where we write $f(t) \sim g(t)$ as $t \rightarrow \infty$ to denote that $\lim_{t \rightarrow \infty} f(t)/g(t) = 1$. Recall that $(x \wedge y)$ denotes the minimum of x and y .

Lemma 2.2. ([20, Lemma 2.6].) *Let X be a nonnegative regularly-varying random variable with exponent $1 < \gamma < 2$ and slowly-varying function \mathcal{L} . Then*

$$\mathbb{E} \left[X \left(1 \wedge \frac{X}{t} \right) \right] \sim \frac{\gamma}{3\gamma - 2 - \gamma^2} \mathcal{L}(t)t^{1-\gamma} \quad \text{as } t \rightarrow \infty.$$

2.3. Heuristics of the proof for Pearson in the ECM

Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, consider $G_n = \text{CM}(\mathbf{D}_n)$, and let us write $r_n = r_n^+ - r_n^-$, where r_n^\pm are positive functions given by

$$r_n^+ = \frac{\sum_{i,j=1}^n X_{ij} D_i D_j}{\sum_{i=1}^n D_i^3 - \frac{1}{L_n} (\sum_{i=1}^n D_i^2)^2}, \quad r_n^- = \frac{\frac{1}{L_n} (\sum_{i=1}^n D_i^2)^2}{\sum_{i=1}^n D_i^3 - \frac{1}{L_n} (\sum_{i=1}^n D_i^2)^2}. \tag{11}$$

First note that by the stable-law central limit theorem (see for instance [42]), there exist two slowly-varying functions \mathcal{L}_0 and \mathcal{L}'_0 such that

$$\frac{\sum_{i=1}^n D_i^2}{\mathcal{L}_0(n)n^{\frac{2}{\gamma}}} \xrightarrow{d} \mathcal{S}_{\gamma/2} \quad \text{and} \quad \frac{\sum_{i=1}^n D_i^3}{\mathcal{L}'_0(n)n^{\frac{3}{\gamma}}} \xrightarrow{d} \mathcal{S}_{\gamma/3} \quad \text{as } n \rightarrow \infty. \tag{12}$$

Applying this and using that $L_n \approx \mu n$,

$$\mathcal{L}_1(n)\mu n^{1-\frac{1}{\gamma}} r_n^- \approx \frac{\mathcal{L}'_0(n)n^{\frac{3}{\gamma}}}{(\mathcal{L}_0(n)n^{\frac{2}{\gamma}})^2} \frac{(\sum_{i=1}^n D_i^2)^2}{\sum_{i=1}^n D_i^3} \xrightarrow{d} \frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}}, \tag{13}$$

with $\mathcal{L}_1(n) = \mathcal{L}'_0(n)/(\mathcal{L}_0(n))^2$. Note that r_n^- scales roughly as $n^{\frac{1}{\gamma}-1}$ and thus tends to zero. This extends the results in the literature that r_n has a nonnegative limit [17].

Next, we need to show that this result also holds when we move to the erased model, i.e. when all degrees D_i are replaced by \widehat{D}_i . To understand how this works, consider the sum of the squares of the degrees in the erased model $\sum_{i=1}^n \widehat{D}_i^2$. Recall that Y_i is the number of removed stubs of node i . Then we have

$$\sum_{i=1}^n \widehat{D}_i^2 = \sum_{i=1}^n (D_i - Y_i)^2 = \sum_{i=1}^n D_i^2 + \sum_{i=1}^n (Y_i^2 - 2D_i Y_i),$$

and hence

$$\left| \sum_{i=1}^n \widehat{D}_i^2 - \sum_{i=1}^n D_i^2 \right| \leq 2 \sum_{i=1}^n Y_i D_i. \tag{14}$$

Therefore we only need to show that the error vanishes when we divide by $n^{\frac{2}{\gamma}}$. For this we can use Corollary 2.1 to get, heuristically,

$$n^{-\frac{2}{\gamma}} \left| \sum_{i=1}^n \widehat{D}_i^2 - \sum_{i=1}^n D_i^2 \right| \leq 2n^{-\frac{2}{\gamma}} \sum_{i=1}^n Y_i D_i \approx n^{-\frac{1}{\gamma}+2-\gamma} \rightarrow 0. \tag{15}$$

These results will be used to prove that when $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$,

$$n^{1-\frac{1}{\gamma}} |r_n^- - \widehat{r}_n^-| \xrightarrow{\mathbb{P}} 0,$$

so that by (13),

$$\mathcal{L}_1(n) \mu n^{1-\frac{1}{\gamma} \widehat{r}_n^-} \xrightarrow{d} \frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}} \text{ as } n \rightarrow \infty.$$

The final ingredient is Proposition 3.3, where we show that for some $\delta > 0$, as $n \rightarrow \infty$,

$$n^{1-\frac{1}{\gamma}+\delta} \widehat{r}_n^+ \xrightarrow{\mathbb{P}} 0. \tag{16}$$

The result then follows, since for n large enough $\mathcal{L}_1(n) \leq Cn^\delta$ and hence

$$\mu \mathcal{L}_1(n) n^{1-\frac{1}{\gamma} \widehat{r}_n^-} = \mu \mathcal{L}_1(n) n^{1-\frac{1}{\gamma} \widehat{r}_n^+} - \mu \mathcal{L}_1(n) n^{1-\frac{1}{\gamma} \widehat{r}_n^-} \xrightarrow{d} -\frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}}.$$

To establish (16), let \widehat{X}_{ij} denote the number of edges between i and j in the erased graph, and note that since we remove self-loops, $\widehat{X}_{ii} = 0$, while in the other cases $\widehat{X}_{ij} = \mathbb{1}_{\{X_{ij} > 0\}}$. We consider the numerator of \widehat{r}_n^+ ,

$$\sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j,$$

and will show that as $n \rightarrow \infty$,

$$n^{1-\frac{4}{\gamma}+\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j \xrightarrow{\mathbb{P}} 0,$$

by approximating $\mathbb{E}_n [\widehat{X}_{ij}]$ by $1 - e^{-D_i D_j / L_n}$; see Lemma 3.2. Since the denominator of \widehat{r}_n^+ scales as $n^{3/\gamma}$ we get that $n^{1-1/\gamma+\delta} \widehat{r}_n^+ \xrightarrow{\mathbb{P}} 0$.

2.4. Intuition behind Conjecture 1.1

We note that, by (2) and since $(\sum_{i \in [n]} D_i^2)^2 = o_{\mathbb{P}}(L_n \sum_{i \in [n]} D_i^3)$,

$$r_n = \frac{\sum_{i,j \in [n]} D_i D_j \left[X_{ij} - \frac{D_i D_j}{L_n} \right]}{\sum_{i \in [n]} D_i^3} (1 + o_{\mathbb{P}}(1)). \tag{17}$$

We rewrite this as

$$r_n = \frac{\sum_{i,j \in [n]} D_i D_j \left[X_{ij} - \frac{D_i D_j}{L_n - 1} \right]}{\sum_{i \in [n]} D_i^3} (1 + o_{\mathbb{P}}(1)) + \frac{\sum_{i,j \in [n]} \frac{D_i^2 D_j^2}{L_n(L_n - 1)}}{\sum_{i \in [n]} D_i^3} (1 + o_{\mathbb{P}}(1)). \tag{18}$$

The second term is

$$O_{\mathbb{P}}\left(n^{4\gamma-2-3\gamma}\right) = O_{\mathbb{P}}\left(n^{\gamma-2}\right) = o_{\mathbb{P}}\left(n^{-1/2}\right), \tag{19}$$

since $\gamma \in (\frac{1}{2}, 1)$, and can thus be ignored. We are thus left to study the first term.

Since $\mathbb{E}_n[X_{ij}] = D_i D_j / (L_n - 1)$, this term is centered. Furthermore, the probability that any half-edge incident to vertex i is connected to vertex j equals $D_j / (L_n - 1)$. These indicators are weakly dependent, so we assume that we can replace the conditional law of X_{ij} given the degrees by a binomial random variable with D_i experiments and success probability $D_j / (L_n - 1)$. We will also assume that these random variables are asymptotically independent. These are the two main assumptions made in this heuristic explanation.

Since a binomial random variable is close to a normal when the number of experiments tends to infinity, these assumptions then suggest that

$$r_n \approx \mathcal{N}(0, \sigma_n^2), \tag{20}$$

where, with Var_n denoting the conditional variance given the degrees,

$$\sigma_n^2 = \frac{\sum_{i,j \in [n]} D_i^2 D_j^2 \text{Var}_n(X_{ij})}{\left(\sum_{i \in [n]} D_i^3\right)^2}. \tag{21}$$

Furthermore, again using that X_{ij} is close to a binomial, $\text{Var}_n(X_{ij}) \approx D_i(D_j/(L_n - 1))(1 - D_j/(L_n - 1)) \approx D_i D_j / L_n$. This suggests that

$$\sigma_n^2 \approx \frac{\sum_{i < j \in [n]} D_i^3 D_j^3 / L_n}{\left(\sum_{i \in [n]} D_i^3\right)^2} = \frac{1}{L_n}, \tag{22}$$

which supports the conjecture in (7) but now with $\sigma^2 = 1/\mu$.

It turns out that the above analysis is not quite correct, as $X_{ij} = X_{ji}$ when $i \leq j$, which means that these terms are highly correlated. Since terms with $i < j$ also appear several times, whereas $i = j$ does not, this turns out to change the variance formula slightly, as we discuss in the proof of Lemma 1.1 in Section 3.5.

2.5. Proofs for clustering in CM and ECM

The general idea behind the proof for both Theorems 1.3 and 1.4 is that, conditioned on the degrees, the clustering coefficients are concentrated around their conditional mean. We then proceed by analyzing this term using stable laws for regularly-varying random variables to obtain the results.

2.5.1 Configuration model. To construct a triangle, six different half-edges at three distinct vertices need to be paired into a triangle. For a vertex with degree D_i , there are $D_i(D_i - 1)/2$ ways to choose two half-edges incident to it. The probability that any two half-edges are paired in the CM can be approximated by $1/L_n$. Thus, the probability that a given set of six half-edges forms a triangle can be approximated by $1/L_n^3$. We then investigate \mathcal{I} , the set of all sets of six half-edges that could possibly form a triangle together. The expected number of triangles can then be approximated by $|\mathcal{I}|/L_n^3$. By computing the size of the set \mathcal{I} , we obtain that the conditional expectation for the clustering coefficient can be written as

$$\mathbb{E}_n [C_n] \approx \frac{|\mathcal{I}|}{L_n^3 \sum_{i \in [n]} D_i(D_i - 1)} \approx \frac{1}{L_n^3} \left(\left(\sum_{i=1}^n D_i^2 \right)^3 - 3 \sum_{i=1}^n D_i^4 + 2 \frac{\sum_{i=1}^n D_i^6}{\sum_{i=1}^n D_i^2} \right).$$

The full details can be found in Section 4.1. Here the first term describes the expected number of times six half-edges are paired into a triangle. The last two terms exclude triangles containing either multi-edges or self-loops. Then by the stable-law central limit theorem [42] we have that there exists a slowly-varying function \mathcal{L}_2 such that

$$\frac{\sum_{i=1}^n D_i^2}{\mathcal{L}_2(n)n^{\frac{2}{\gamma}}} \xrightarrow{d} \mathcal{S}_{\gamma/2}, \quad \frac{\sum_{i=1}^n D_i^4}{\mathcal{L}_2(n)^2 n^{\frac{4}{\gamma}}} \xrightarrow{d} \mathcal{S}_{\gamma/4}, \quad \text{and} \quad \frac{\sum_{i=1}^n D_i^6}{\mathcal{L}_2(n)^6 n^{\frac{6}{\gamma}} \sum_{i=1}^n D_i^2} \xrightarrow{d} \frac{\mathcal{S}_{\gamma/6}}{\mathcal{S}_{\gamma/2}}.$$

Hence, using that $L_n \approx \mu n$ we obtain that

$$\frac{\mathbb{E}_n [C_n]}{\mathcal{L}_2(n)n^{\frac{4}{\gamma}-3}} \xrightarrow{d} \frac{1}{\mu^3} \left(\mathcal{S}_{\gamma/2}^2 - 3\mathcal{S}_{\gamma/4} + \frac{2\mathcal{S}_{\gamma/6}}{\mathcal{S}_{\gamma/2}} \right),$$

where $\mathcal{S}_{\gamma/2}$, $\mathcal{S}_{\gamma/4}$, and $\mathcal{S}_{\gamma/6}$ are given by (6). To complete the proof we establish a concentration result for C_n , conditioned on the degrees. To be more precise, we employ a careful counting argument, following the approach in the proof of [13, Proposition 7.13], to show (see Lemma 4.1) that there exists a $\delta > 0$ such that

$$\frac{n^\delta \text{Var}_n(C_n)}{n^{\frac{8}{\gamma}-6}} \xrightarrow{\mathbb{P}} 0,$$

where Var_n denotes the conditional variance given the degrees. Then it follows from Chebyshev’s inequality, conditioned on the degrees, that

$$\frac{|C_n - \mathbb{E}_n [C_n]|}{\mathcal{L}_2(n)n^{\frac{4}{\gamma}-3}} \xrightarrow{\mathbb{P}} 0,$$

and we conclude that

$$\frac{C_n}{\mathcal{L}_2(n)n^{\frac{4}{\gamma}-3}} \xrightarrow{d} \frac{1}{\mu^3} \left(\mathcal{S}_{\gamma/2}^2 - 3\mathcal{S}_{\gamma/4} + 2\mathcal{S}_{\gamma/6}/\mathcal{S}_{\gamma/2} \right).$$

2.5.2 Erased configuration model. The difficulty for clustering in ECM, compared to CM, is in showing that \widehat{C}_n behaves as its conditional expectation, as well as in establishing its scaling. To compute this we first fix an $\varepsilon > 0$ and show in Lemma 4.2 that the main contribution is given by triples of nodes with degrees $\varepsilon\sqrt{n} \leq D \leq \frac{\sqrt{n}}{\varepsilon}$; i.e.,

$$\begin{aligned} \sum_{1 \leq i < j < k \leq n} \widehat{X}_{ij} \widehat{X}_{jk} \widehat{X}_{ik} &= \sum_{1 \leq i < j < k \leq n} \widehat{X}_{ij} \widehat{X}_{jk} \widehat{X}_{ik} \mathbb{1}_{\{\varepsilon\sqrt{n} \leq D_i, D_j, D_k \leq \frac{\sqrt{n}}{\varepsilon}\}} \\ &+ O\left(\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}\right) \mathcal{E}_1(\varepsilon), \end{aligned}$$

where $\mathcal{E}_1(\varepsilon)$ is an error function, independent of n , with $\lim_{\varepsilon \rightarrow 0} \mathcal{E}_1(\varepsilon) = 0$. Then we use that approximately $\mathbb{E}_n [\widehat{X}_{ij}] \approx 1 - e^{-D_i D_j / L_n}$ to show that

$$\mathbb{E}_n [\widehat{C}_n] \approx \frac{6 \sum_{1 \leq i < j < k \leq n} g_{n,\varepsilon}(D_i, D_j, D_k) + O\left(\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}\right) \mathcal{E}_1(\varepsilon)}{\sum_{i=1}^n \widehat{D}_i (\widehat{D}_i - 1)},$$

where

$$g_{n,\varepsilon}(x, y, z) = \left(1 - e^{-\frac{xy}{L_n}}\right) \left(1 - e^{-\frac{yz}{L_n}}\right) \left(1 - e^{-\frac{zx}{L_n}}\right) \mathbb{1}_{\{\varepsilon\sqrt{n} \leq x, y, z \leq \frac{\sqrt{n}}{\varepsilon}\}}.$$

Here \mathbb{E}_n again denotes expectation conditioned on \mathbf{D}_n , hence conditional on the sampled degrees of the underlying CM. The precise statement can be found in Lemma 4.3. After that, we show in Lemma 4.6 that \widehat{C}_n concentrates around its expectation conditioned on the sampled degrees, so that conditioned on the sampled degree sequence, we can approximate $\widehat{C}_n \approx \mathbb{E}_n [\widehat{C}_n]$. We then replace L_n by μn in Lemma 4.4, so that, conditioned on the degree sequence,

$$\widehat{C}_n \approx \frac{6 \sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_i, D_j, D_k) + O\left(\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}\right) \mathcal{E}_1(\varepsilon)}{\sum_{i=1}^n \widehat{D}_i(\widehat{D}_i - 1)},$$

with

$$f_{n,\varepsilon}(x, y, z) = \left(1 - e^{-\frac{xy}{\mu n}}\right) \left(1 - e^{-\frac{yz}{\mu n}}\right) \left(1 - e^{-\frac{zx}{\mu n}}\right) \mathbb{1}_{\left\{\varepsilon \sqrt{n} \leq x, y, z \leq \frac{\sqrt{n}}{\varepsilon}\right\}}.$$

We then take the random degrees into account, by showing that

$$\frac{1}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}} \sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_1, D_2, D_3) \xrightarrow{\mathbb{P}} \frac{1}{6} \mu^{-\frac{3\gamma}{2}} A_\gamma(\varepsilon) + \mathcal{E}_2(\varepsilon),$$

where

$$A_\gamma(\varepsilon) = \int_\varepsilon^{1/\varepsilon} \int_\varepsilon^{1/\varepsilon} \int_\varepsilon^{1/\varepsilon} \frac{1}{(xyz)^{\gamma+1}} (1 - e^{-xy})(1 - e^{-xy})(1 - e^{-xy}) \, dx \, dy \, dz,$$

and $\mathcal{E}_2(\varepsilon)$ is a deterministic error function, with $\lim_{\varepsilon \rightarrow 0} \mathcal{E}_2(\varepsilon) = 0$. Finally, we again replace the \widehat{D}_i with D_i and use the stable-law central limit theorem to obtain a slowly-varying function \mathcal{L}_3 such that

$$\frac{\mathcal{L}_3(n) n^{\frac{2}{\gamma}}}{\sum_{i=1}^n \widehat{D}_i(\widehat{D}_i - 1)} \xrightarrow{d} \frac{1}{S_{\gamma/2}}.$$

Combining all these results implies that, for any $\varepsilon > 0$,

$$\frac{\mathcal{L}_3(n) \widehat{C}_n}{\mathcal{L}(\sqrt{\mu n})^3 n^{-\frac{3\gamma}{2} + 3 - \frac{2}{\gamma}}} \xrightarrow{d} \mu^{-\frac{3\gamma}{2}} \frac{A_\gamma(\varepsilon)}{S_{\gamma/2}^2} + \frac{\mathcal{E}_1(\varepsilon) + \mathcal{E}_2(\varepsilon)}{S_{\gamma/2}^2},$$

from which the result follows by taking $\varepsilon \downarrow 0$.

3. Pearson’s correlation coefficient

In this section we first give the proof of Theorem 1.1, where we follow the approach described in Section 2.3. We then prove Lemma 1.1, which supports Conjecture 1.1 on the behavior of Pearson in the CM.

3.1. Limit theorem for r_n^-

We first prove a limit theorem for r_n^- , when $G_n = \text{CM}(\mathbf{D}_n)$. Recall that

$$r_n^- = \frac{\frac{1}{L_n} \left(\sum_{i=1}^n D_i^2\right)^2}{\sum_{i=1}^n D_i^3 - \frac{1}{L_n} \left(\sum_{i=1}^n D_i^2\right)^2}.$$

Proposition 3.1. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Then, if $G_n = \text{CM}(\mathbf{D}_n)$, there exists a slowly-varying function \mathcal{L}_0 such that*

$$\mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} r_n^- \xrightarrow{d} \frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}}$$

as $n \rightarrow \infty$. Here $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ are given by (6).

Proof. We will first show that there exists a slowly-varying function \mathcal{L}_0 such that

$$\mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} \frac{(\sum_{i=1}^n D_i^2)^2}{\mu n \sum_{i=1}^n D_i^3} \xrightarrow{d} \frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}} \tag{23}$$

as $n \rightarrow \infty$, with $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ defined by (6).

Let $D_{(i)}$ denote the i th largest degree in \mathbf{D}_n , i.e. $D_{(1)} \geq D_{(2)} \geq \dots \geq D_{(n)}$, and let Γ_i be defined as in (5). Then, since $\sum_{i=1}^n D_{(i)}^p = \sum_{i=1}^n D_i^p$ for any $p \geq 0$, it follows from [13, Theorem 2.33] that for some slowly-varying functions $\mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4$, and \mathcal{L}_6 , as $n \rightarrow \infty$,

$$\begin{aligned} & \left(\frac{n^{-\frac{4}{\gamma}}}{\mathcal{L}_2(n)^2} \left(\sum_{i=1}^n D_i^2 \right)^2, \frac{n^{-\frac{3}{\gamma}}}{\mathcal{L}_3(n)} \sum_{i=1}^n D_i^3, \frac{n^{-\frac{4}{\gamma}}}{\mathcal{L}_4(n)} \sum_{i=1}^n D_i^4, \frac{n^{-\frac{6}{\gamma}}}{\mathcal{L}_6(n)} \sum_{i=1}^n D_i^6 \right) \\ & \xrightarrow{d} \left(\left(\sum_{i=1}^{\infty} \Gamma_i^{-\frac{2}{\gamma}} \right)^2, \sum_{i=1}^{\infty} \Gamma_i^{-\frac{3}{\gamma}}, \sum_{i=1}^{\infty} \Gamma_i^{-\frac{4}{\gamma}}, \sum_{i=1}^{\infty} \Gamma_i^{-\frac{6}{\gamma}} \right). \end{aligned} \tag{24}$$

Here we include the fourth and sixth moments, since these will be needed later for proving Theorem 1.3.

Note that $\mathcal{L}_0(n) := \mathcal{L}_2(n)/\mathcal{L}_1(n)^2$ is slowly varying and $\mathbb{P}\left(\sum_{i=1}^{\infty} \Gamma_i^{-t/\gamma} \leq 0\right) = 0$ for any $t \geq 2$. Hence (23) follows from (24) and the continuous mapping theorem. Hence, to prove the main result, it is enough to show that

$$\mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} \left| r_n^- - \frac{(\sum_{i=1}^n D_i^2)^2}{\mu n \sum_{i=1}^n D_i^3} \right| \xrightarrow{\mathbb{P}} 0.$$

We will prove the stronger statement

$$n^{1-\frac{1}{\gamma}+\frac{\kappa}{2}} \left| r_n^- - \frac{(\sum_{i=1}^n D_i^2)^2}{\mu n \sum_{i=1}^n D_i^3} \right| \xrightarrow{\mathbb{P}} 0, \tag{25}$$

where $\kappa = (\gamma - 1)/(\gamma + 1) > 0$ is the same as in Lemma 2.1.

Note that by Lemma 2.1 we have that $\mu n/L_n \xrightarrow{\mathbb{P}} 1$. Hence, by (23), we have that for any $\delta > 0$,

$$n^{1-\frac{1}{\gamma}-\delta} \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} = n^{1-\frac{1}{\gamma}-\delta} \frac{(\sum_{i=1}^n D_i^2)^2}{\mu n \sum_{i=1}^n D_i^3} \left(\frac{\mu n}{L_n} \right) \xrightarrow{\mathbb{P}} 0,$$

from which we conclude that

$$n^{1-\frac{1}{\gamma}-\delta} \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \xrightarrow{\mathbb{P}} 0. \tag{26}$$

To show (25), we write

$$\begin{aligned} \left| \frac{(\sum_{i=1}^n D_i^2)^2}{\mu n \sum_{i=1}^n D_i^3} - r_n^- \right| &= \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} \left| \frac{(L_n - \mu n) \sum_{i=1}^n D_i^3 + \mu n (\sum_{i=1}^n D_i^2)^2}{\mu n (L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2)} \right| \\ &\leq \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \frac{|\mu n - L_n|}{\mu n} \\ &\quad + \left(\frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \right)^2. \end{aligned}$$

For the first term we have, using (26) and Lemma 2.1,

$$\begin{aligned} &n^{1-\frac{1}{\gamma}+\frac{\kappa}{2}} \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \frac{|\mu n - L_n|}{\mu n} \\ &= n^{1-\frac{1}{\gamma}-\frac{\kappa}{2}} \left(\frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \right) \left(\frac{|\mu n - L_n|}{\mu n^{1-\kappa}} \right) \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Now, let $\delta = 1 - 1/\gamma - \kappa/2 > 0$. Then, since $1 - 1/\gamma + \kappa/2 = 2 - 2/\gamma - \delta$, it follows from (26) that

$$n^{1-\frac{1}{\gamma}+\frac{\kappa}{2}} \left(\frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \right)^2 = \left(\frac{n^{1-\frac{1}{\gamma}-\frac{\delta}{2}} (\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \right)^2 \xrightarrow{\mathbb{P}} 0,$$

which finishes the proof of (25). □

3.2. Limit theorem for \widehat{r}_n^-

We now turn to the ECM. Observe that for $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$, (14) and Corollary 2.1 imply that, for any $\delta > 0$,

$$\frac{|\sum_{i=1}^n D_i^2 - \sum_{i=1}^n \widehat{D}_i^2|}{n^{\frac{1}{\gamma}+2-\gamma+\delta}} \leq \frac{2 \sum_{i=1}^n D_i Y_i}{n^{\frac{1}{\gamma}+2-\gamma+\delta}} \xrightarrow{\mathbb{P}} 0.$$

Since $\frac{2}{\gamma} > \frac{1}{\gamma} + 2 - \gamma$ for all $\gamma > 1$ this result implies that for any $\delta > 0$,

$$\frac{\sum_{i=1}^n \widehat{D}_i^2}{n^{\frac{2}{\gamma}+\delta}} \xrightarrow{\mathbb{P}} 0. \tag{27}$$

This line of reasoning can be extended to sums of \widehat{D}_i^p for any $p > 0$, proving that the degrees \widehat{D}_i in the ECM satisfy the same scaling results as those for D_i . In particular we have the following extension of (26) to the ECM. Recall that \mathcal{D} denotes an integer-valued random variable with a regularly-varying distribution defined by

$$\mathbb{P}(\mathcal{D} > t) = \mathcal{L}(t)t^{-\gamma}$$

(see (1)).

Lemma 3.1. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 3$ and $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Then, for any $\delta > 0$,*

$$n^{1-\frac{1}{\gamma}-\delta} \frac{(\sum_{i=1}^n \widehat{D}_i^2)^2}{L_n \sum_{i=1}^n \widehat{D}_i^3 - (\sum_{i=1}^n \widehat{D}_i^2)^2} \xrightarrow{\mathbb{P}} 0.$$

Now recall that \widehat{r}_n^- denotes the negative part of Pearson’s correlation coefficient for the CM, i.e.

$$\widehat{r}_n^- = \frac{\frac{1}{L_n} (\sum_{i=1}^n \widehat{D}_i^2)^2}{\sum_{i=1}^n \widehat{D}_i^3 - \frac{1}{L_n} (\sum_{i=1}^n \widehat{D}_i^2)^2}.$$

The next proposition shows that $|r_n^- - \widehat{r}_n^-| = o_{\mathbb{P}}(n^{\frac{1}{\gamma}-1})$.

Proposition 3.2. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$. Let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$, denote by G_n the graph before the removal of edges, and recall that r_n^- and \widehat{r}_n^- denote the negative parts of Pearson’s correlation coefficient in G_n and \widehat{G}_n , respectively. Then*

$$n^{1-\frac{1}{\gamma}+\frac{(\gamma-1)^2}{4\gamma}} |r_n^- - \widehat{r}_n^-| \xrightarrow{\mathbb{P}} 0.$$

Proof. The proof consists of splitting the main term into separate terms, which can be expressed in terms of erased stubs or edges, and showing that each of these terms converges to zero. Throughout the proof we let

$$\delta = \frac{(\gamma - 1)^2}{4\gamma}.$$

We start by splitting the main term as

$$|r_n^- - \widehat{r}_n^-| \leq \frac{|\left(\sum_{i=1}^n D_i^2\right)^2 - \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2|}{L_n \sum_{i=1}^n D_i^3 - \left(\sum_{i=1}^n D_i^2\right)^2} \tag{28}$$

$$+ \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2 \left| \frac{1}{L_n \sum_{i=1}^n D_i^3 - \left(\sum_{i=1}^n D_i^2\right)^2} - \frac{1}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3 - \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2} \right|. \tag{29}$$

For (28) we use that

$$\begin{aligned} \left| \left(\sum_{i=1}^n D_i^2\right)^2 - \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2 \right| &= \left(\sum_{i=1}^n D_i^2 - \widehat{D}_i^2\right) \left(\sum_{i=1}^n D_i^2 + \widehat{D}_i^2\right) \\ &\leq \left(\sum_{i=1}^n 2D_i Y_i\right) \left(\sum_{i=1}^n 2D_i^2 + Y_i^2\right) \leq 6 \sum_{i=1}^n D_i^2 \sum_{j=1}^n Y_j D_j \end{aligned}$$

to obtain

$$\begin{aligned} \frac{|\left(\sum_{i=1}^n D_i^2\right)^2 - \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2|}{L_n \sum_{i=1}^n D_i^3 - \left(\sum_{i=1}^n D_i^2\right)^2} &\leq \frac{6 \sum_{i=1}^n D_i^2 \sum_{i=1}^n Y_i D_i}{L_n \sum_{i=1}^n D_i^3 - \left(\sum_{i=1}^n D_i^2\right)^2} \\ &= \frac{6 \sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i^2} \frac{\left(\sum_{i=1}^n D_i^2\right)^2}{L_n \sum_{i=1}^n D_i^3 - \left(\sum_{i=1}^n D_i^2\right)^2}. \end{aligned} \tag{30}$$

Now observe that

$$2 - \frac{1}{\gamma} - \gamma = -(\gamma - 1)^2/\gamma = -4\delta, \tag{31}$$

and hence

$$1 - \frac{1}{\gamma} + \delta = \gamma - 1 - 3\delta = -\left(\frac{1}{\gamma} + 2 - \gamma + \delta\right) + \left(\frac{2}{\gamma} - \frac{\delta}{2}\right) + \left(1 - \frac{1}{\gamma} - \frac{\delta}{2}\right) - \delta,$$

with all three terms inside the brackets positive. Therefore, it follows from (30), together with Corollary 2.1, Proposition 2.1, and (26), that

$$\begin{aligned} & n^{1-\frac{1}{\gamma}+\delta} \frac{\left|(\sum_{i=1}^n D_i^2)^2 - (\sum_{i=1}^n \widehat{D}_i^2)^2\right|}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} \\ & \leq n^{-\delta} \left(\frac{6 \sum_{i=1}^n D_i Y_i}{n^{\frac{1}{\gamma}+2-\gamma+\delta}}\right) \left(\frac{n^{\frac{2}{\gamma}-\frac{\delta}{2}}}{\sum_{i=1}^n D_i^2}\right) \left(\frac{n^{1-\frac{1}{\gamma}-\frac{\delta}{2}} (\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2}\right) \xrightarrow{\mathbb{P}} 0. \end{aligned} \tag{32}$$

The second term, (29), requires more work. Let us first write

$$\begin{aligned} & \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2 \left| \frac{1}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} - \frac{1}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3 - (\sum_{i=1}^n \widehat{D}_i^2)^2} \right| \\ & := \frac{(\sum_{i=1}^n \widehat{D}_i^2)^2}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3 - (\sum_{i=1}^n \widehat{D}_i^2)^2} \left(I_n^{(1)} + I_n^{(2)} + I_n^{(3)}\right), \end{aligned}$$

with

$$\begin{aligned} I_n^{(1)} & := \frac{(\sum_{i=1}^n D_i^2)^2 - (\sum_{i=1}^n \widehat{D}_i^2)^2}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2}, \\ I_n^{(2)} & := \frac{Z_n \sum_{i=1}^n \widehat{D}_i^3}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2}, \\ I_n^{(3)} & := \frac{L_n \left| \sum_{i=1}^n D_i^3 - \sum_{i=1}^n \widehat{D}_i^3 \right|}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2}, \end{aligned}$$

and we recall that $Z_n = L_n - \widehat{L}_n$ denotes the total number of removed edges. Note that

$$n^{1-\frac{1}{\gamma}-\frac{\delta}{2}} \frac{(\sum_{i=1}^n \widehat{D}_i^2)^2}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3 - (\sum_{i=1}^n \widehat{D}_i^2)^2} \xrightarrow{\mathbb{P}} 0.$$

Therefore, in order to complete the proof, it suffices to show that

$$n^{\frac{3\delta}{2}} I_n^{(t)} \xrightarrow{\mathbb{P}} 0 \quad \text{for } t = 1, 2, 3.$$

For $t = 1$ this follows from (32), since

$$\frac{1}{\gamma} - 1 + \frac{\delta}{2} = \frac{\gamma^2 - 10\gamma + 9}{8\gamma} = \frac{(\gamma - 1)(\gamma - 9)}{8\gamma} < 0,$$

and hence

$$\frac{3\delta}{2} < 1 - \frac{1}{\gamma} + \delta.$$

For $t = 2$ we use that $\widehat{D}_i \leq D_i$ to obtain

$$I_n^{(2)} \leq \frac{E_n \sum_{i=1}^n D_i^3}{L_n \sum_{i=1}^n D_i^3 - (\sum_{i=1}^n D_i^2)^2} = \frac{E_n}{L_n} \left(1 - \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} \right)^{-1}.$$

By Proposition 3.1 it follows that

$$\left(1 - \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} \right)^{-1} \xrightarrow{\mathbb{P}} 1. \tag{33}$$

In addition we have that $\varepsilon := \gamma - 1 - 3\delta/2 > 0$ and hence, by Theorem 2.1 and the strong law of large numbers,

$$n^{\frac{3\delta}{2}} I_n^{(2)} \leq \left(n^{\gamma-2-\varepsilon} \frac{E_n}{\mu} \right) \frac{\mu n}{L_n} \left(1 - \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} \right)^{-1} \xrightarrow{\mathbb{P}} 0.$$

Finally, for $I_n^{(3)}$ we first compute

$$\left| \sum_{i=1}^n D_i^3 - \sum_{i=1}^n \widehat{D}_i^3 \right| = \sum_{i=1}^n Y_i^3 + 3D_i^2 Y_i - 3D_i Y_i^2 \leq 4 \sum_{i=1}^n Y_i D_i^2,$$

and hence,

$$I_n^{(3)} \leq \frac{4 \sum_{i=1}^n Y_i D_i^2}{\sum_{i=1}^n D_i^3} \left(1 - \frac{(\sum_{i=1}^n D_i^2)^2}{L_n \sum_{i=1}^n D_i^3} \right)^{-1}.$$

By (33) the last term converges in probability to one. Finally, by (31),

$$\frac{3\delta}{2} \leq 2\delta = 4\delta - 2\delta = \left(\gamma - 2 - \frac{2}{\gamma} - \delta \right) + \left(\frac{3}{\gamma} - \delta \right),$$

and hence, by Corollary 2.1 and Proposition 2.1,

$$n^{\frac{3\delta}{2}} I_n^{(3)} \leq \left(4n^{\gamma-2-\frac{2}{\gamma}-\delta} \sum_{i=1}^n Y_i D_i^2 \right) \left(\frac{n^{\frac{3}{\gamma}-\delta}}{\sum_{i=1}^n D_i^3} \right) \xrightarrow{\mathbb{P}} 0,$$

which finishes the proof. □

3.3. Convergence of \widehat{r}_n^+

The next step towards the proof of Theorem 1.1 is to show that, for some $\delta > 0$,

$$n^{1-\frac{1}{\gamma}+\delta}\widehat{r}_n^+ \xrightarrow{\mathbb{P}} 0,$$

where

$$\widehat{r}_n^+ = \frac{\sum_{i,j=1}^n \widehat{X}_{ij}\widehat{D}_i\widehat{D}_j}{\sum_{i=1}^n \widehat{D}_i^3 - \frac{1}{L_n} \left(\sum_{i=1}^n \widehat{D}_i^2\right)^2}$$

denotes the positive part of Pearson’s correlation coefficient in the ECM. Here $\widehat{X}_{ij} = \mathbb{1}_{\{X_{ij}>0\}}$ denotes the event that i and j are connected by at least one edge in the CM graph G_n . The main ingredient of this result is the following lemma, which gives an approximation for $\sum_{1\leq i<j\leq n} \mathbb{P}_n(X_{ij}=0) D_i D_j$.

Lemma 3.2. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\mathbb{E}[\mathcal{D}] = \mu$. Consider graphs $G_n = \text{CM}(\mathbf{D}_n)$ and define*

$$M_n = \sum_{1\leq i<j\leq n} \left| \mathbb{P}_n(X_{ij}=0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\} \right| D_i D_j.$$

Then, for any $K > 0$ and $0 < \delta < \left(\frac{2-\gamma}{\gamma} \wedge \frac{\gamma-1}{\gamma}\right)$,

$$n^{1-\frac{4}{\gamma}+\delta} Z_n \xrightarrow{\mathbb{P}} 0.$$

In our proofs M_n will be divided by

$$\sum_{i=1}^n D_i^3 - \frac{1}{L_n} \left(\sum_{i=1}^n D_i^2\right)^2,$$

which is of the order $n^{3/\gamma}$. Hence the final expression will be of the order $n^{\frac{1}{\gamma}-1-\delta} = o(n^{\frac{1}{\gamma}-1})$, which is enough to prove the final result.

To prove Lemma 3.2, we will use the following technical result.

Lemma 3.3. ([24, Lemma 6.7].) *For any nonnegative $x, x_0 > 0, y_i, z_i \geq 0$ with $z_i < x$ for all i , and any $m \geq 1$, we have*

$$-\frac{x_0}{x^2}(x_0 - x)^+ - \frac{x_0}{2} \max_{1\leq i\leq m} \frac{z_i}{(x - z_i)^2} \leq \prod_{i=1}^m \left(1 - \frac{z_i}{x}\right)^{y_i} - \exp\left\{-\frac{1}{x_0} \sum_{i=1}^m y_i z_i\right\} \leq \frac{|x - x_0|}{(x \wedge x_0)}.$$

Proof of Lemma 3.2. We will first consider the term $\left| \mathbb{P}_n(X_{ij}=0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\} \right|$. It follows from computations done in [14] that

$$0 \leq \mathbb{P}_n(X_{ij}=0) - \prod_{t=0}^{D_i-1} \left(1 - \frac{D_j}{L_n - 2t - 1}\right) \leq \frac{D_i^2 D_j}{(L_n - 2D_i)^2}. \tag{34}$$

For the product term in (34) we have the bounds

$$\left(1 - \frac{D_j}{L_n - 2D_i + 1}\right)^{D_i} \leq \prod_{t=0}^{D_i-1} \left(1 - \frac{D_j}{L_n - 2t - 1}\right) \leq \left(1 - \frac{D_j}{L_n}\right)^{D_i},$$

and therefore, using Lemma 3.3 with $m = 1$, we can bound the difference between $\mathbb{P}_n(X_{ij} = 0)$ and $\exp\{-D_i D_j / L_n\}$. For the lower bound we take $x = L_n$, $x_0 = L_n + 1 - 2D_i$, $y = D_i$, and $z = D_j$ to get

$$-\frac{L_n(2D_i - 1)}{(L_n - 2D_i + 1)^2} - \frac{D_j}{L_n + 1 - 2D_1 - D_j} \leq \mathbb{P}_n(X_{ij} = 0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\}, \tag{35}$$

while changing x_0 to L_n yields

$$\mathbb{P}_n(X_{ij} = 0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\} \leq \frac{D_i^2 D_j}{(L_n - 2D_i)^2}. \tag{36}$$

Combining (35) and (36) gives

$$\begin{aligned} & \sum_{1 \leq i < j \leq n} \left| \mathbb{P}_n(X_{ij} = 0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\} \right| D_i D_j \\ & \leq \sum_{1 \leq i < j \leq n} \frac{2L_n D_i^2 D_j}{(L_n - 2D_i + 1)^2} + \sum_{1 \leq i < j \leq n} \frac{D_i D_j^2}{L_n + 1 - 2D_1 - D_j} + \sum_{1 \leq i < j \leq n} \frac{D_i^3 D_j^2}{(L_n - 2D_i)^2} \\ & := I_n^{(1)} + I_n^{(2)} + I_n^{(3)}. \end{aligned}$$

We will now show that

$$n^{1 - \frac{4}{\gamma} + \delta} I_n^{(t)} \xrightarrow{\mathbb{P}} 0 \quad \text{for } t = 1, 2, 3, \tag{37}$$

which proves the result.

For the remainder of the proof we define

$$D_n^{\max} := \max_{1 \leq i \leq n} D_i$$

and observe that by our choice of δ ,

$$\varepsilon_1 := \frac{2}{\gamma} - 1 - \delta = \frac{2 - \gamma}{\gamma} - \delta > 0 \quad \text{and} \quad \varepsilon_2 := 1 - \frac{1}{\gamma} - \delta = \frac{\gamma - 1}{\gamma} - \delta > 0.$$

For $t = 1$, we have

$$I_n^{(1)} = \sum_{1 \leq i < j \leq n} \frac{2L_n D_i^2 D_j}{(L_n - 2D_i + 1)^2} \leq \frac{2L_n^2 \sum_{i=1}^n D_i^2}{(L_n^2 - 4L_n D_n^{\max})} = \left(2 \sum_{i=1}^n D_i^2\right) \left(1 - \frac{4D_n^{\max}}{L_n}\right)^{-1}.$$

By the strong law of large numbers and Proposition 2.1, it follows that

$$\frac{M_n}{L_n} = \frac{\mu n D_n^{\max}}{L_n \mu n} \xrightarrow{\mathbb{P}} 0,$$

and hence

$$\left(1 - \frac{4D_n^{\max}}{L_n}\right)^{-1} \xrightarrow{\mathbb{P}} 1. \tag{38}$$

Proposition 2.1 then implies

$$n^{1 + \delta - \frac{4}{\gamma}} I_n^{(1)} \leq \left(2n^{-\frac{2}{\gamma} - \varepsilon_1} \sum_{i=1}^n D_i^2\right) \left(1 - \frac{4D_n^{\max}}{L_n}\right)^{-1} \xrightarrow{\mathbb{P}} 0.$$

The analysis for $I_n^{(2)}$ is similar, so that we are left with $I_n^{(3)}$. Here, we have

$$\begin{aligned} I_n^{(3)} &= \sum_{1 \leq i < j \leq n} \frac{D_i^3 D_j^2}{(L_n - 2D_i)^2} \leq \frac{\sum_{i=1}^n D_i^2 \sum_{j=1}^n D_j^3}{(L_n^2 - 4L_n D_n^{\max})} \\ &= \frac{1}{L_n^2} \left(\sum_{i=1}^n D_i^2 \right) \left(\sum_{j=1}^n D_j^3 \right) \left(1 - \frac{4D_n^{\max}}{L_n} \right)^{-1}. \end{aligned}$$

The last term again converges in probability to one, by (38). For the remaining terms we use the definition of ε_2 and Proposition 2.1 to obtain

$$\begin{aligned} n^{1+\delta-\frac{4}{\gamma}} I_n^{(3)} &\leq n^{2-\frac{5}{\gamma}+\varepsilon_2} \frac{1}{L_n^2} \left(\sum_{i=1}^n D_i^2 \right) \left(\sum_{j=1}^n D_j^3 \right) \left(1 - \frac{4D_n^{\max}}{L_n} \right)^{-1} \\ &= \frac{n^2}{L_n^2} \left(n^{-\frac{2}{\gamma}-\frac{\varepsilon_2}{2}} \sum_{i=1}^n D_i^2 \right) \left(n^{-\frac{3}{\gamma}-\frac{\varepsilon_2}{2}} \sum_{j=1}^n D_j^3 \right) \left(1 - \frac{4D_n^{\max}}{L_n} \right)^{-1} \xrightarrow{\mathbb{P}} 0, \end{aligned}$$

which finishes the proof of (37). □

We proceed to prove the convergence of \widehat{r}_n^+ .

Proposition 3.3. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, $\mathbb{E}[\mathcal{D}] = \mu$, and let $\widehat{\mathbf{G}}_n = \text{ECM}(\mathbf{D}_n)$. Then, for any slowly-varying function \mathcal{L} ,*

$$\mathcal{L}(n)n^{1-\frac{1}{\gamma}}\widehat{r}_n^+ \xrightarrow{\mathbb{P}} 0. \tag{39}$$

Proof. Let

$$\delta = \left(\frac{2-\gamma}{4\gamma} \wedge \frac{\gamma-1}{4\gamma} \right).$$

We will show that

$$n^{1-\frac{1}{\gamma}+\delta}\widehat{r}_n^+ \xrightarrow{\mathbb{P}} 0, \tag{40}$$

which then implies (39), since by Potter’s theorem $\mathcal{L}(n)n^{-\delta} \rightarrow 0$ for any $\delta > 0$.

The main part of the proof of (40) will be to show that

$$n^{1-\frac{4}{\gamma}+2\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j \xrightarrow{\mathbb{P}} 0. \tag{41}$$

To see that (41) implies (40), we write

$$\begin{aligned} \widehat{r}_n^+ &\leq \frac{\sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j}{\sum_{i=1}^n \widehat{D}_i^3 - \frac{1}{L_n} \left(\sum_{i=1}^n \widehat{D}_i^2 \right)^2} \\ &= \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j \left(\frac{1}{\sum_{i=1}^n \widehat{D}_i^3} \right) \left(1 - \frac{\left(\sum_{i=1}^n \widehat{D}_i^2 \right)^2}{L_n \sum_{i=1}^n \widehat{D}_i^3} \right)^{-1}. \end{aligned}$$

By Lemma 3.1,

$$\left(1 - \frac{(\sum_{i=1}^n \widehat{D}_i^2)^2}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3}\right)^{-1} \xrightarrow{\mathbb{P}} 1,$$

and hence, using (41) and Proposition 2.1,

$$n^{1-\frac{1}{\gamma}+\delta} \widehat{r}_n^+ \leq \left(n^{1-\frac{4}{\gamma}+2\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j\right) \left(\frac{n^{-\frac{3}{\gamma}-\delta}}{\sum_{i=1}^n \widehat{D}_i^3}\right) \left(1 - \frac{(\sum_{i=1}^n \widehat{D}_i^2)^2}{\widehat{L}_n \sum_{i=1}^n \widehat{D}_i^3}\right)^{-1} \xrightarrow{\mathbb{P}} 0.$$

To prove (41) let $\kappa = (\gamma - 1)/(\gamma + 1)$ and define the events

$$\begin{aligned} \mathcal{A}_n &= \{|L_n - \mu n| \leq n^{1-\kappa}\}, \\ \mathcal{B}_n &= \left\{ \sum_{1 \leq i < j \leq n} \left| \mathbb{P}_n(X_{ij}=0) - \exp\left\{-\frac{D_i D_j}{L_n}\right\} \right| D_i D_j \leq n^{\frac{4}{\gamma}-1-3\delta} \right\}. \end{aligned} \tag{42}$$

Then, if we set $\Lambda_n = \mathcal{A}_n \cap \mathcal{B}_n$, it follows from Lemmas 2.1 and 3.2 that $\mathbb{P}(\Lambda_n) \rightarrow 1$, and hence it is enough to prove that for any $K > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(n^{1-\frac{4}{\gamma}+2\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j > K, \Lambda_n\right) = 0. \tag{43}$$

First, since $\mathbb{E}_n[\widehat{X}_{ij}] = \mathbb{P}_n(X_{ij} > 0)$ and Λ_n is completely determined by the degree sequence,

$$\begin{aligned} \mathbb{E}_n \left[\sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j \mathbb{1}_{\Lambda_n} \right] &= \sum_{1 \leq i < j \leq n} \mathbb{P}_n(X_{ij} > 0) D_i D_j \mathbb{1}_{\Lambda_n} \\ &= \sum_{1 \leq i < j \leq n} (1 - \mathbb{P}_n(X_{ij} = 0)) D_i D_j \mathbb{1}_{\Lambda_n} \\ &\leq \sum_{1 \leq i < j \leq n} \left(1 - \exp\left\{-\frac{D_i D_j}{L_n}\right\}\right) D_i D_j \mathbb{1}_{\Lambda_n} + n^{\frac{4}{\gamma}-1-3\delta} \\ &\leq \sum_{1 \leq i < j \leq n} \left(1 - \exp\left\{-\frac{D_i D_j}{\mu n - n^{1-\kappa}}\right\}\right) D_i D_j + n^{\frac{4}{\gamma}-1-3\delta}. \end{aligned}$$

From this we obtain, using Markov’s inequality, that

$$\begin{aligned} &\mathbb{P}\left(n^{1-\frac{4}{\gamma}+2\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j > K, \Lambda_n\right) \\ &\leq \frac{n^{3-\frac{4}{\gamma}+2\delta}}{K} \mathbb{E} \left[\left(1 - \exp\left\{-\frac{\mathcal{D}_1 \mathcal{D}_2}{\mu n - n^{1-\kappa}}\right\}\right) \mathcal{D}_1 \mathcal{D}_2 \right] + O(n^{-\delta}), \end{aligned} \tag{44}$$

where \mathcal{D}_1 and \mathcal{D}_2 are two independent copies of \mathcal{D} . It follows that $\mathcal{D}_1\mathcal{D}_2$ is again regularly varying with exponent $1 < \gamma < 2$. Therefore, since $1 - e^{-x} \leq (1 \wedge x)$, and using Lemma 2.2,

$$\begin{aligned} & \frac{n^{3-\frac{4}{\gamma}+2\delta}}{K} \mathbb{E} \left[\left(1 - \exp \left\{ -\frac{\mathcal{D}_1\mathcal{D}_2}{\mu n - n^{1-\kappa}} \right\} \right) \mathcal{D}_1\mathcal{D}_2 \right] \\ & \leq \frac{n^{3-\frac{4}{\gamma}+2\delta}}{K} \mathbb{E} \left[\left(1 \wedge \frac{\mathcal{D}_1\mathcal{D}_2}{\mu n - n^{1-\kappa}} \right) \mathcal{D}_1\mathcal{D}_2 \right] \\ & = O \left(n^{4-\frac{4}{\gamma}-\gamma+2\delta} \right). \end{aligned} \tag{45}$$

Now observe that by our choice of $\delta > 0$ and since $2 - \gamma < 1$, we get

$$4 - \frac{4}{\gamma} - \gamma + 2\delta = \frac{-(\gamma - 2)^2}{\gamma} + 2\delta \leq \frac{-(\gamma - 2)^2}{\gamma} + \frac{2 - \gamma}{2\gamma} \leq \frac{-(\gamma - 2)^2}{2\gamma} < 0.$$

Plugging this into (45), it follows from (44) that

$$\mathbb{P} \left(n^{1-\frac{4}{\gamma}+2\delta} \sum_{1 \leq i < j \leq n} \widehat{X}_{ij} D_i D_j > K, \Lambda_n \right) = O \left(n^{-\frac{(\gamma-2)^2}{2\gamma}} + n^{-\delta} \right),$$

and hence (43) follows. □

3.4. Proving Theorem 1.1

We are now ready to prove the central limit theorem for the ECM.

Proof of Theorem 1.1. Let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$, let $\mathcal{S}_{\gamma/2}$ and $\mathcal{S}_{\gamma/3}$ be defined as in (6), and let \mathcal{L}_0 be given by Proposition 3.1. Now we write

$$\mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} r(\widehat{G}_n) = \mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} \widehat{r}_n^+ - \mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} \widehat{r}_n^-.$$

By Proposition 3.3 it follows that the first part converges to zero in probability, as $n \rightarrow \infty$. For the second part, let $\delta = (\gamma - 1)^2 / (4\gamma)$ and note that by Potter’s theorem [4, Theorem 1.5.6] we have that $\mathcal{L}_0(n) \leq n^\delta$ for all large enough n . Then, if we denote by G_n the graph before the removal of edges, it follows by Proposition 3.2 that

$$\mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} |r_n^- - \widehat{r}_n^-| \xrightarrow{\mathbb{P}} 0.$$

Finally, we remark that the graph G_n is generated by the CM, so that the above and Proposition 3.1 now imply

$$\mu \mathcal{L}_0(n) n^{1-\frac{1}{\gamma}} \widehat{r}_n^- \xrightarrow{d} \frac{\mathcal{S}_{\gamma/2}^2}{\mathcal{S}_{\gamma/3}}$$

as $n \rightarrow \infty$, from which the result follows. □

3.5. Pearson in the CM: proof of Lemma 1.1

In this section, we prove Lemma 1.1 on the tightness of the conditional variance of Pearson in the CM.

Proof of Lemma 1.1. Since, conditionally on the degrees, the only randomness in r_n is in $(X_{ij})_{1 \leq i < j \leq n}$, we use the covariance formula for sums of random variables to compute that

$$\begin{aligned} \text{Var}_n(r_n) &= \frac{\sum_{i,j,k,l \in [n]} D_i D_j D_k D_l \text{Cov}_n(X_{ij}, X_{kl})}{\left[\sum_{i \in [n]} D_i^3 - \left(\sum_{i \in [n]} D_i^2 \right)^2 / L_n \right]^2} \tag{46} \\ &= \frac{\sum_{i,j,k,l \in [n]} D_i D_j D_k D_l \text{Cov}_n(X_{ij}, X_{kl})}{\left(\sum_{i \in [n]} D_i^3 \right)^2} (1 + o_{\mathbb{P}}(1)), \end{aligned}$$

since $\sum_{i \in [n]} D_i^3 \gg \left(\sum_{i \in [n]} D_i^2 \right)^2 / L_n$, where we write Cov_n for the conditional variance in the CM given the i.i.d. degrees. We next compute $\text{Cov}_n(X_{ij}, X_{kl})$, depending on the precise form of $\{i, j, k, l\}$. For this, we note that

$$X_{ij} = \sum_{s=1}^{D_i} \sum_{t=1}^{D_j} I_{st}, \tag{47}$$

with I_{st} the indicator that the s th half-edge incident to i pairs to the t th half-edge incident to j .

Case: $(i, j) = (k, l)$ with $i < j$. We compute that

$$\begin{aligned} \text{Var}_n(X_{ij}) &= \sum_{(s,t)} \sum_{(s',t')} \text{Cov}_n(I_{st}, I_{s't'}) \tag{48} \\ &= \frac{D_i(D_i - 1)D_j(D_j - 1)}{(L_n - 1)(L_n - 3)} + \frac{D_i D_j}{(L_n - 1)} - \frac{D_i^2 D_j^2}{(L_n - 1)^2} \\ &= \frac{D_i D_j}{(L_n - 1)} + 2 \frac{D_i^2 D_j^2}{(L_n - 1)^2 (L_n - 3)} - \frac{D_i D_j (D_j - 1)}{(L_n - 1)(L_n - 3)} - \frac{D_i^2 D_j}{(L_n - 1)(L_n - 3)} \\ &= \frac{D_i D_j}{L_n} (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

Thus,

$$\begin{aligned} n \frac{\sum_{i < j \in [n]} D_i^2 D_j^2 \text{Var}_n(X_{ij})}{\left(\sum_{i \in [n]} D_i^3 \right)^2} &= \frac{n}{L_n} (1 + o_{\mathbb{P}}(1)) \frac{\sum_{i < j \in [n]} D_i^3 D_j^3}{\left(\sum_{i \in [n]} D_i^3 \right)^2} \tag{49} \\ &= \frac{1}{\mu} (1 + o_{\mathbb{P}}(1)) \frac{\sum_{i < j \in [n]} D_i^3 D_j^3}{\left(\sum_{i \in [n]} D_i^3 \right)^2}. \end{aligned}$$

Since we sum over all $i, j \in [n]$ and not just $i < j$, this term appears four times.

Case: $|\{i, j, k, l\}| = 1$. In this case, we obtain

$$\begin{aligned} \text{Var}_n(X_{ii}) &= \frac{D_i(D_i - 1)}{(L_n - 1)} + \frac{D_i(D_i - 1)(D_i - 2)(D_i - 3)}{(L_n - 1)(L_n - 3)} - \frac{D_i^4}{(L_n - 1)^2} \\ &= \frac{D_i^2}{L_n}(1 + o_{\mathbb{P}}(1)) + \frac{D_i^4}{L_n^3}(1 + o_{\mathbb{P}}(1)) - 6\frac{D_i^3}{L_n^2}(1 + o_{\mathbb{P}}(1)) \\ &= \frac{D_i^2}{L_n}(1 + o_{\mathbb{P}}(1)), \end{aligned} \tag{50}$$

since $D_i = O_{\mathbb{P}}(n^\gamma)$ with $\gamma \in (\frac{1}{2}, 1)$. Therefore,

$$n \frac{\sum_{i \in [n]} D_i^4 \text{Var}_n(X_{ii})}{\left(\sum_{i \in [n]} D_i^3\right)^2} = \frac{n(1 + o_{\mathbb{P}}(1))}{L_n} \frac{\sum_{i \in [n]} D_i^6}{\left(\sum_{i \in [n]} D_i^3\right)^2}.$$

The above two computations show that these contributions sum up to

$$\frac{1}{\mu}(1 + o_{\mathbb{P}}(1)) \frac{4 \sum_{i < j \in [n]} D_i^3 D_j^3}{\left(\sum_{i \in [n]} D_i^3\right)^2} + \frac{1}{\mu}(1 + o_{\mathbb{P}}(1)) \frac{\sum_{i \in [n]} D_i^6}{\left(\sum_{i \in [n]} D_i^3\right)^2} \xrightarrow{d} \frac{2 - \mathcal{S}_{\gamma/6}/\mathcal{S}_{\gamma/3}^2}{\mu}. \tag{51}$$

We are left to show that all other terms constitute error terms.

Case: $|\{i, j, k, l\}| = 4$. When $|\{i, j, k, l\}| = 4$, we compute that

$$\begin{aligned} \text{Cov}_n(X_{ij}, X_{kl}) &= D_i D_j D_l D_k \left(\frac{1}{(L_n - 1)(L_n - 3)} - \frac{1}{(L_n - 1)^2} \right) \\ &= \frac{2D_i D_j D_l D_k}{(L_n - 1)^2(L_n - 3)}. \end{aligned} \tag{52}$$

Thus, the contribution to the variance of r_n of this case equals

$$\frac{\left(\sum_{i \in [n]} D_i^2\right)^4}{L_n^3 \left(\sum_{i \in [n]} D_i^3\right)^2} (1 + o_{\mathbb{P}}(1)) = O_{\mathbb{P}}\left(n^{2\gamma - 3}\right) = o_{\mathbb{P}}(n^{-1}), \tag{53}$$

since $\gamma \in (\frac{1}{2}, 1)$.

Case: $|\{i, j, k, l\}| = 3$. When $|\{i, j, k, l\}| = 3$, we compute that

$$\begin{aligned} \text{Cov}_n(X_{ij}, X_{il}) &= \frac{D_i(D_i - 1)D_j D_l}{(L_n - 1)(L_n - 3)} - \frac{D_i^2 D_j D_l}{(L_n - 1)^2} \\ &= 2 \frac{D_i^2 D_j D_l}{(L_n - 1)^2(L_n - 3)} - \frac{D_i D_j D_l}{(L_n - 1)(L_n - 3)}. \end{aligned} \tag{54}$$

Thus, the contribution of this case to the variance of r_n equals

$$\begin{aligned}
 & 2 \frac{\left(\sum_{i \in [n]} D_i^2\right)^2}{L_n^3 \sum_{i \in [n]} D_i^3} (1 + o_{\mathbb{P}}(1)) - \frac{\left(\sum_{i \in [n]} D_i^2\right)^3}{L_n^2 \left(\sum_{i \in [n]} D_i^3\right)^2} (1 + o_{\mathbb{P}}(1)) \\
 & = O_{\mathbb{P}}\left(n^{\gamma-3}\right) + O_{\mathbb{P}}\left(n^{-2}\right) = o_{\mathbb{P}}\left(n^{-1}\right),
 \end{aligned} \tag{55}$$

since $\gamma \in (\frac{1}{2}, 1)$.

This completes the proof. □

4. Clustering coefficient

In this section, we prove Theorems 1.3 and 1.4 on the clustering coefficient in the CM as well as the ECM. In both models, we first study the clustering coefficient when the degree sequence is fixed. We show that the clustering coefficient concentrates around its expected value when the degrees are given. Then we analyze how the random degrees influence the clustering coefficient.

4.1. Clustering in the configuration model

In this section, we compute the clustering coefficient for a CM with a power-law degree distribution with $\gamma \in (1, 2)$. To prove Theorem 1.3, we first use a second moment method to show that the number of triangles Δ_n concentrates on its expected value conditioned on the degrees. Then we take the random degrees into account and show that the rescaled clustering coefficient converges to the stable distributions from Theorem 1.3.

4.1.1 *Concentration for the number of triangles.* The concentration result is formally stated and proved in the next lemma.

Lemma 4.1. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, and $G_n = \text{CM}(\mathbf{D}_n)$. Let Δ_n denote the number of triangles in G_n . Then, for any $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}_n (|\Delta_n - \mathbb{E}_n [\Delta_n]| > \varepsilon \mathbb{E}_n [\Delta_n]) = 0.$$

Proof. Fix $0 < \delta < (\frac{2}{\gamma} - 1)/6$. Define the event

$$\mathcal{B}_n = \left\{ \sum_{i=1}^n D_i^2 \geq Kn^{2/\gamma-\delta}, \sum_{i=1}^n D_i^3 \leq Kn^{3/\gamma+\delta} \right\}. \tag{56}$$

Let \mathcal{A}_n be the event defined in (42), and let $\Lambda_n = \mathcal{B}_n \cap \mathcal{A}_n$. We have $\lim_{n \rightarrow \infty} \mathbb{P}_n (\Lambda_n) = 1$; thus, we only need to prove the result on the event Λ_n . The proof is similar to the proof of [13, Proposition 7.13].

Define

$$\begin{aligned}
 \mathcal{I} = \{ & (s_1 t_1, s_2 u_1, u_2 t_2, i, j, k) : 1 \leq i < j < k \leq n, 1 \leq s_1 \neq s_2 \leq D_i, 1 \leq t_1 \neq t_2 \leq D_j, \\
 & 1 \leq u_1 \neq u_2 \leq D_k \}.
 \end{aligned}$$

Thus, \mathcal{I} contains all combinations of six labeled half-edges that could possibly form a triangle on three distinct vertices. Then

$$|\mathcal{I}| = \sum_{1 \leq i < j < k \leq n} D_i(D_i - 1)D_j(D_j - 1)D_k(D_k - 1) \tag{57}$$

denotes the number of ways six half-edges could form a triangle. For $m \in \mathcal{I}$, let $\mathbb{1}_m$ denote the indicator variable that the six half-edges of m form a triangle in the way specified by m . Then

$$\Delta_n = \sum_{m \in \mathcal{I}} \mathbb{1}_m.$$

The probability that the half-edges in m form a triangle can be written as

$$\mathbb{P}_n(\mathbb{1}_m = 1) = \prod_{j=1}^3 (L_n + 1 - 2j)^{-1}.$$

This results in

$$\mathbb{E}_n[\Delta_n] = \sum_{m \in \mathcal{I}} \mathbb{P}_n(\mathbb{1}_m = 1) = \frac{|\mathcal{I}|}{\prod_{j=1}^3 (L_n + 1 - 2j)}. \tag{58}$$

Furthermore, by [13, Theorem 2.5],

$$\mathbb{E}_n[\Delta_n(\Delta_n - 1)] = \sum_{m_1 \neq m_2 \in \mathcal{I}} \mathbb{P}_n(\mathbb{1}_{m_1} = \mathbb{1}_{m_2} = 1).$$

When all six pairs of half-edges involved in m_1 and m_2 are distinct, the probability that these pairs of half-edges that form m_1 and m_2 are paired in the correct way is

$$\mathbb{P}_n(\mathbb{1}_{m_1} = \mathbb{1}_{m_2} = 1) = \prod_{j=1}^6 (L_n + 1 - 2j)^{-1}.$$

If m_1 and m_2 contain one pair of half-edges that is the same (so that m_1 and m_2 form two triangles merged on one edge), then

$$\mathbb{P}_n(\mathbb{1}_{m_1} = \mathbb{1}_{m_2} = 1) = \prod_{j=1}^5 (L_n + 1 - 2j)^{-1}.$$

Let \mathcal{I}_2 denote the set of combinations of 10 half-edges that could possibly form two triangles merged by one edge. Then, similarly to (57),

$$\begin{aligned} |\mathcal{I}_2| &= \sum_{1 \leq i < j \leq n} D_i(D_i - 1)(D_i - 2)D_j(D_j - 1)(D_j - 2) \sum_{1 \leq k < l \leq n} D_k(D_k - 1)D_l(D_l - 1) \\ &\leq \left(\sum_{i \in [n]} D_i^3 \right)^2 \left(\sum_{i \in [n]} D_i^2 \right)^2. \end{aligned} \tag{59}$$

Similarly, if m_1 and m_2 overlap at two pairs of half-edges (so that m_1 and m_2 form two triangles merged by two edges), then

$$\mathbb{P}_n(\mathbb{1}_{m_1} = \mathbb{1}_{m_2} = 1) = \prod_{j=1}^4 (L_n + 1 - 2j)^{-1}.$$

Let \mathcal{I}_3 denote the set of combinations of eight half-edges that could possibly form two triangles merged by two edges. Then, similarly to (57),

$$\begin{aligned}
 |\mathcal{I}_3| &= \sum_{1 \leq i < j \leq n} D_i(D_i - 1)(D_i - 2)D_j(D_j - 1)(D_j - 2) \sum_{1 \leq k \leq n} D_k(D_k - 1) \\
 &\leq \left(\sum_{i \in [n]} D_i^3 \right)^2 \sum_{i \in [n]} D_i^2.
 \end{aligned}
 \tag{60}$$

In all other cases the probability of the event $\mathbb{1}_{m_1} = \mathbb{1}_{m_2} = 1$ then equals zero. These are cases where m_1 prescribes some half-edge to be merged to half-edge j_1 , whereas m_2 prescribes it to be merged to some other half-edge j_2 . Therefore,

$$\mathbb{E}_n [\Delta_n(\Delta_n - 1)] \leq \frac{|\mathcal{I}|^2}{\prod_{j=1}^6 (L_n + 1 - 2j)} + \frac{|\mathcal{I}_2|}{\prod_{j=1}^5 (L_n + 1 - 2j)} + \frac{|\mathcal{I}_3|}{\prod_{j=1}^4 (L_n + 1 - 2j)}.
 \tag{61}$$

On the event \mathcal{B}_n defined in (56),

$$\begin{aligned}
 \frac{|\mathcal{I}_2|}{\prod_{j=1}^5 (L_n + 1 - 2j)} / \frac{|\mathcal{I}|^2}{\prod_{j=1}^6 (L_n + 1 - 2j)} &= O\left(\frac{\left(\sum_{i \in [n]} D_i^3 \right)^2 / \left(\sum_{i \in [n]} D_i^2 \right)^4}{(L_n + 1 - 12)^{-1}} \right) \\
 &= O\left(n \cdot n^{6/\gamma + 2\delta - 8/\gamma + 4\delta} \right) = O\left(n^{1 - 2/\gamma + 6\delta} \right) \\
 &= o(1),
 \end{aligned}$$

by the choice of δ . In a similar way, we can show that the third term is small compared to the first term of (61). Therefore,

$$\mathbb{E}_n [\Delta_n(\Delta_n - 1)] \leq \frac{|\mathcal{I}|^2}{\prod_{j=1}^6 (L_n + 1 - 2j)} (1 + o(1)).$$

Finally, on the event \mathcal{B}_n , we have

$$|\mathcal{I}| = \Theta\left(\left(\sum_i D_i^2 \right)^3 \right) = \Omega(n^{6/\gamma - 3\delta}).$$

Using that $L_n = \mu n(1 + o(1))$ on the event \mathcal{A}_n results in

$$\begin{aligned}
 \frac{\text{Var}_n(\Delta_n)}{(\mathbb{E}_n[\Delta_n])^2} &= \frac{\mathbb{E}_n[\Delta_n(\Delta_n - 1)]}{(\mathbb{E}_n[\Delta_n])^2} - 1 + \frac{\mathbb{E}_n[\Delta_n]}{(\mathbb{E}_n[\Delta_n])^2} \\
 &\leq \frac{(L_n - 1)(L_n - 3)(L_n - 5)}{(L_n - 7)(L_n - 9)(L_n - 11)} (1 + o(1)) - 1 + \frac{\prod_{j=1}^3 (L_n + 1 - 2j)}{|\mathcal{I}|} \\
 &= 1 + o(1) - 1 + O(n^{3 + 3\delta - \frac{6}{\gamma}}) = o(1),
 \end{aligned}$$

for $\gamma \in (1, 2)$. Then by Chebyshev’s inequality, on the event Δ_n ,

$$\mathbb{P}_n (|\Delta_n - \mathbb{E}_n[\Delta_n]| > \varepsilon \mathbb{E}_n[\Delta_n]) \leq \frac{\text{Var}_n(\Delta_n)}{(\mathbb{E}_n[\Delta_n])^2 \varepsilon^2} = o(1),$$

which gives the result. □

4.1.2 *Proof of Theorem 1.3.* We again prove the result under the event $\Lambda_n = \mathcal{B}_n \cap \mathcal{A}_n$, where \mathcal{B}_n and \mathcal{A}_n are given in (56) and (42), respectively.

By (57) and (58),

$$\begin{aligned} \mathbb{E}_n [\Delta_n] &= \frac{1}{\mu^3 n^3} \sum_{1 \leq i < j < k \leq n} D_i(D_i - 1)D_j(D_j - 1)D_k(D_k - 1)(1 + o_{\mathbb{P}}(1)) \\ &= \frac{1}{\mu^3 n^3} \left(\frac{1}{6} \left(\sum_{i=1}^n D_i(D_i - 1) \right)^3 - \frac{1}{2} \sum_{i=1}^n D_i(D_i - 1) \sum_{j=1}^n D_j^2(D_j - 1)^2 \right. \\ &\quad \left. + \frac{1}{3} \sum_{i=1}^n D_i^3(D_i - 1)^3 \right) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

Then, the definition of the clustering coefficient in (4) yields

$$\begin{aligned} \mathbb{E}_n [C_n] &= \frac{1}{\mu^3 n^3} \left(\left(\sum_{i=1}^n D_i(D_i - 1) \right)^2 - 3 \sum_{i=1}^n D_i^2(D_i - 1)^2 \right. \\ &\quad \left. + 2 \frac{\sum_{i=1}^n D_i^3(D_i - 1)^3}{\sum_{i=1}^n D_i(D_i - 1)} \right) (1 + o_{\mathbb{P}}(1)) \\ &= \frac{1}{\mu^3 n^3} \left(\left(\sum_{i=1}^n D_i^2 \right)^2 - 3 \sum_{i=1}^n D_i^4 + 2 \sum_{i=1}^n D_i^6 / \sum_{i=1}^n D_i^2 \right) (1 + o_{\mathbb{P}}(1)). \end{aligned} \tag{62}$$

Lemma 4.1 then gives that conditioned on the degree sequence,

$$C_n = \frac{1}{\mu^3 n^3} \left(\left(\sum_{i=1}^n D_i^2 \right)^2 - 3 \sum_{i=1}^n D_i^4 + 2 \sum_{i=1}^n D_i^6 / \sum_{i=1}^n D_i^2 \right) (1 + o_{\mathbb{P}}(1)).$$

By (24), there exist slowly-varying functions $\mathcal{L}_1(n)$, $\mathcal{L}_2(n)$, and $\mathcal{L}_3(n)$ such that

$$\left(n^{-2/\gamma} \frac{\sum_{i=1}^n D_i^2}{\mathcal{L}_1(n)}, n^{-4/\gamma} \frac{\sum_{i=1}^n D_i^4}{\mathcal{L}_2(n)}, n^{-6/\gamma} \frac{\sum_{i=1}^n D_i^6}{\mathcal{L}_3(n)} \right) \xrightarrow{d} (\mathcal{S}_{\gamma/2}, \mathcal{S}_{\gamma/4}, \mathcal{S}_{\gamma/6}), \tag{63}$$

where

$$\mathcal{S}_{\gamma/2} = \sum_{i=1}^{\infty} \Gamma_i^{-2/\gamma}, \quad \mathcal{S}_{\gamma/4} = \sum_{i=1}^{\infty} \Gamma_i^{-4/\gamma}, \quad \mathcal{S}_{\gamma/6} = \sum_{i=1}^{\infty} \Gamma_i^{-6/\gamma},$$

for the same random variables Γ_i . Furthermore, by [42, Eq. (5.23)], the slowly-varying functions in (63) satisfy, for some slowly-varying function $\mathcal{L}_0(n)$,

$$\mathcal{L}_1(n) = \sqrt{\mathcal{L}_0(n)}(K_{\gamma/2})^{2/\gamma}, \quad \mathcal{L}_2(n) = \mathcal{L}_0(n)(K_{\gamma/4})^{4/\gamma}, \quad \mathcal{L}_3(n) = \mathcal{L}_0(n)^{3/2}(K_{\gamma/6})^{6/\gamma},$$

where

$$K_{\alpha} = \frac{1 - \alpha}{\Gamma(2 - \alpha) \cos(\pi\alpha/2)},$$

with Γ the gamma function. Therefore,

$$\frac{n^{-4/\gamma}}{\mathcal{L}_0(n)} \left(\left(\sum_{i=1}^n D_i^2 \right)^2, \sum_{i=1}^n D_i^4, \frac{\sum_{i=1}^n D_i^6}{\sum_{i=1}^n D_i^2} \right) \xrightarrow{d} \left((K_{\gamma/2})^{\gamma/4} \mathcal{S}_{\gamma/2}^2, (K_{\gamma/4})^{\gamma/4} \mathcal{S}_{\gamma/4}, \frac{(K_{\gamma/6})^{\gamma/6} \mathcal{S}_{\gamma/6}}{(K_{\gamma/2})^{\gamma/2} \mathcal{S}_{\gamma/2}} \right).$$

Combining this with (62) results in (9). □

4.2. Clustering coefficient in the erased configuration model

In this section, we study the clustering coefficient in the ECM. Again, we start with the expectation and the variance of the clustering coefficient conditioned on the sampled degree sequence, i.e. the sequence $\mathbf{D}_n = \{D_1, D_2, \dots, D_n\}$ sampled from the distribution (1). Note that this is not the eventual degree sequence of the graph constructed by the ECM.

Structure of the proof of Theorem 1.4. We prove Theorem 1.4 in four steps:

Step 1. We show in Lemma 4.2 that the expected contribution to the number of triangles from vertices with sampled degrees larger than \sqrt{n}/ε and smaller than $\varepsilon\sqrt{n}$ is small for fixed $0 < \varepsilon < 1$. Therefore, in the rest of the proof we focus only on counting triangles between vertices of degrees in $[\varepsilon\sqrt{n}, \sqrt{n}/\varepsilon]$.

Step 2. We calculate the expected number of triangles between vertices of sampled degrees proportional to \sqrt{n} , conditioned on the degree sequence. In Lemma 4.3, we show that this expectation can be written as the sum of a function of the degrees.

Step 3. We show that the variance of the number of triangles between vertices of sampled degree proportional to \sqrt{n} is small in Lemma 4.6. Thus, we can replace the number of triangles conditioned on the degrees by its expected value, which we computed in Step 2.

Step 4. We show that the expected number of triangles conditioned on the sampled degrees converges to the value given in Theorem 1.4, when taking the random degrees into account.

We will start by proving the three lemmas described above. Let $B_n(\varepsilon)$ denote the interval $[\varepsilon\sqrt{\mu n}, \sqrt{\mu n}/\varepsilon]$ for some $\varepsilon > 0$. Furthermore, let \widehat{X}_{ij} denote the number of edges between vertices i and j in the ECM. Then we can write the number of triangles as

$$\begin{aligned} \Delta_n &= \sum_{1 \leq i < j < k \leq n} \widehat{X}_{ij} \widehat{X}_{jk} \widehat{X}_{ik} \mathbb{1}_{\{D_i, D_j, D_k \in B_n(\varepsilon)\}} + \sum_{1 \leq i < j < k \leq n} \widehat{X}_{ij} \widehat{X}_{jk} \widehat{X}_{ik} \mathbb{1}_{\{D_i, D_j \text{ or } D_k \notin B_n(\varepsilon)\}} \\ &=: \Delta_n(B_n(\varepsilon)) + \Delta_n(\bar{B}_n(\varepsilon)). \end{aligned} \tag{64}$$

We want to show that the major contribution to Δ_n comes from $\Delta_n(B_n(\varepsilon))$. The following lemma shows that the expected contribution of $\Delta_n(\bar{B}_n(\varepsilon))$ to the number of triangles is small.

Lemma 4.2. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, and let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Let $\Delta_n(\bar{B}_n(\varepsilon))$ denote the number of triangles in \widehat{G}_n with at least one of the sampled degrees not in $B_n(\varepsilon)$. Then*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[\Delta_n(\bar{B}_n(\varepsilon))]}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}} = O(\mathcal{E}_1(\varepsilon))$$

for some function $\mathcal{E}_1(\varepsilon)$ not depending on n such that $\mathcal{E}_1(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Proof. Let $\Delta_{i,j,k}$ denote the event that a triangle is present on vertices $i, j,$ and k . By [18, Lemma 4.1],

$$\begin{aligned} \mathbb{P}_n(\Delta_{i,j,k}) &= \Theta\left(\prod_{(u,v)\in\{(i,j),(j,k),(i,k)\}} (1 - e^{-D_u D_v / L_n}) \mathbb{1}_{\{D_u D_v < L_n\}}\right) \\ &= \Theta\left(\left(\frac{D_i D_j}{L_n} \wedge 1\right) \left(\frac{D_i D_k}{L_n} \wedge 1\right) \left(\frac{D_j D_k}{L_n} \wedge 1\right)\right). \end{aligned}$$

Therefore, for some $\tilde{K} > 0,$

$$\begin{aligned} \mathbb{E}_n[\Delta_n(\bar{B}_n(\varepsilon))] &\leq \tilde{K} \sum_{1 \leq i < j < k \leq n} \left(\frac{D_i D_j}{L_n} \wedge 1\right) \left(\frac{D_i D_k}{L_n} \wedge 1\right) \left(\frac{D_j D_k}{L_n} \wedge 1\right) \mathbb{1}_{\{D_i, D_j, \text{ or } D_k \in \bar{B}_n(\varepsilon)\}}. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}[\Delta_n(\bar{B}_n(\varepsilon))] &\leq \tilde{K} \frac{1}{2} n^3 \mathbb{E}\left[\left(\frac{D_1 D_2}{\mu n} \wedge 1\right) \left(\frac{D_1 D_3}{\mu n} \wedge 1\right) \left(\frac{D_2 D_3}{\mu n} \wedge 1\right) \mathbb{1}_{\{D_1 \in \bar{B}_n(\varepsilon)\}}\right]. \end{aligned} \tag{65}$$

We now show that the contribution of (65) when $D_1 < \varepsilon \sqrt{\mu n}$ is small. We write

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{D_1 D_2}{\mu n} \wedge 1\right) \left(\frac{D_1 D_3}{\mu n} \wedge 1\right) \left(\frac{D_2 D_3}{\mu n} \wedge 1\right) \mathbb{1}_{\{D_1 < \varepsilon \sqrt{\mu n}\}}\right] \\ &\leq \tilde{K} \sum_{t_1=1}^{\varepsilon \sqrt{\mu n}} \sum_{t_2=1}^n \sum_{t_3=1}^n \mathbb{P}(D_1 = t_1, D_2 = t_2, D_3 = t_3) \left(\frac{t_1 t_2}{\mu n} \wedge 1\right) \left(\frac{t_1 t_3}{\mu n} \wedge 1\right) \left(\frac{t_2 t_3}{\mu n} \wedge 1\right) \\ &\leq \tilde{K} K^3 \sum_{t_1=1}^{\varepsilon \sqrt{\mu n}} \sum_{t_2, t_3=1}^n \mathcal{L}(t_1) \mathcal{L}(t_2) \mathcal{L}(t_3) (t_1 t_2 t_3)^{-\gamma-1} \left(\frac{t_1 t_2}{\mu n} \wedge 1\right) \left(\frac{t_1 t_3}{\mu n} \wedge 1\right) \left(\frac{t_2 t_3}{\mu n} \wedge 1\right), \end{aligned} \tag{66}$$

where we have used Assumption 1.1 and the fact that $D_1, D_2,$ and D_3 are independent. As

$$\int_0^\infty \int_0^\infty \int_0^\infty (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) \, dx \, dy \, dz < \infty, \tag{67}$$

for all $\gamma \in (1, 2),$ by [28, Theorem 2],

$$\begin{aligned} &\int_0^\varepsilon \int_0^\varepsilon \int_0^\varepsilon \mathcal{L}(t_1) \mathcal{L}(t_2) \mathcal{L}(t_3) (t_1 t_2 t_3)^{-\gamma-1} \left(\frac{t_1 t_2}{\mu n} \wedge 1\right) \left(\frac{t_1 t_3}{\mu n} \wedge 1\right) \left(\frac{t_2 t_3}{\mu n} \wedge 1\right) dt_3 \, dt_2 \, dt_1 \\ &= (\mu n)^{-\frac{3}{2}\gamma} \int_0^\varepsilon \int_0^\varepsilon \int_0^\varepsilon \mathcal{L}(x \sqrt{\mu n}) \mathcal{L}(y \sqrt{\mu n}) \mathcal{L}(z \sqrt{\mu n}) (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) \, dx \, dy \, dz \\ &= (1 + o(1)) \mathcal{L}(\sqrt{\mu n})^3 \int_0^\varepsilon \int_0^\varepsilon \int_0^\varepsilon (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) \, dx \, dy \, dz. \end{aligned}$$

We then bound the sum in (66) as

$$\begin{aligned} & \sum_{t_1=1}^{\varepsilon\sqrt{\mu n}} \sum_{t_2=1}^n \sum_{t_3=1}^n \mathcal{L}(t_1)\mathcal{L}(t_2)\mathcal{L}(t_3)(t_1t_2t_3)^{-\gamma-1} \left(\frac{t_1t_2}{\mu n} \wedge 1\right) \left(\frac{t_1t_3}{\mu n} \wedge 1\right) \left(\frac{t_2t_3}{\mu n} \wedge 1\right) \\ & \leq 2 \int_0^{\varepsilon\sqrt{\mu n}+1} \int_0^\infty \int_0^\infty \mathcal{L}(t_1)\mathcal{L}(t_2)\mathcal{L}(t_3)(t_1t_2t_3)^{-\gamma-1} \\ & \quad \times \left(\frac{t_1t_2}{\mu n} \wedge 1\right) \left(\frac{t_1t_3}{\mu n} \wedge 1\right) \left(\frac{t_2t_3}{\mu n} \wedge 1\right) dt_3 dt_2 dt_1 \\ & = 2(1 + o(1))(\mu n)^{-\frac{3}{2}\gamma} \mathcal{L}(\sqrt{\mu n})^3 \int_0^\varepsilon \int_0^\infty \int_0^\infty (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) dx dy dz. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{D_1D_2}{\mu n} \wedge 1\right) \left(\frac{D_1D_3}{\mu n} \wedge 1\right) \left(\frac{D_2D_3}{\mu n} \wedge 1\right) \mathbb{1}_{\{D_1 < \varepsilon\sqrt{\mu n}\}} \right] \\ & \leq K_2 n^{\frac{3}{2}(1-\gamma)} \mathcal{L}(\sqrt{\mu n})^3 \int_0^\varepsilon \int_0^\infty \int_0^\infty (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) dx dy dz, \quad (68) \end{aligned}$$

for some constant K_2 . Thus, we only need to prove that the last triple integral in (68) tends to zero as $\varepsilon \rightarrow 0$. Using (67), we obtain

$$\int_0^\varepsilon \int_0^\infty \int_0^\infty (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) dx dy dz := \mathcal{E}_0(\varepsilon),$$

where $\mathcal{E}_0(\varepsilon)$ is such that $\mathcal{E}_0(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Thus, by (65) and (66), the contribution to the expectation where one of the degrees is smaller than $\varepsilon\sqrt{\mu n}$ is bounded by $O_{\mathbb{P}}\left(\mathcal{L}(\sqrt{\mu n})^3 n^{3-\frac{3}{2}\gamma} \mathcal{E}_0(\varepsilon)\right)$. Similarly,

$$\int_{1/\varepsilon}^\infty \int_0^\infty \int_0^\infty (xyz)^{-\gamma-1} (xy \wedge 1) (xz \wedge 1) (yz \wedge 1) dx dy dz := \mathcal{E}'_0(\varepsilon),$$

where again $\mathcal{E}'_0(\varepsilon)$ satisfies $\mathcal{E}'_0(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Therefore, we can show in a similar way that the contribution to the expected number of triangles where one of the degrees is larger than $\sqrt{\mu n}/\varepsilon$ is $O_{\mathbb{P}}\left(\mathcal{L}(\sqrt{\mu n})^3 n^{3-\frac{3}{2}\gamma} \mathcal{E}'_0(\varepsilon)\right)$. Then, taking $\mathcal{E}_1(\varepsilon) = \max(\mathcal{E}_0(\varepsilon), \mathcal{E}'_0(\varepsilon))$ proves the lemma. \square

The next lemma computes the expected contribution of the vertices of sampled degree proportional to \sqrt{n} to the number of triangles. Define

$$g_{n,\varepsilon}(D_i, D_j, D_k) := \left(1 - e^{-\frac{D_iD_j}{Ln}}\right) \left(1 - e^{-\frac{D_iD_k}{Ln}}\right) \left(1 - e^{-\frac{D_kD_j}{Ln}}\right) \mathbb{1}_{\{(D_i, D_j, D_k) \in B_n(\varepsilon)\}}, \quad (69)$$

and let \sum' denote a sum over distinct indices, such that $i < j < k$.

Lemma 4.3. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, and let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Let $\Delta_n(B_n(\varepsilon))$ denote the number of triangles in \widehat{G}_n with sampled degrees in $B_n(\varepsilon)$. Then, on the event \mathcal{A}_n as defined in (42),*

$$\mathbb{E}_n [\Delta_n(B_n(\varepsilon))] = (1 + o(1)) \sum'_{i,j,k} g_{n,\varepsilon}(D_i, D_j, D_k).$$

Proof. We can write the expectation as

$$\mathbb{E}_n [\Delta_n(B_n(\varepsilon))] = \sum'_{i,j,k} \mathbb{P}_n (\Delta_{i,j,k}) \mathbb{1}_{\{(D_i, D_j, D_k) \in B_n(\varepsilon)\}}, \tag{70}$$

where $\mathbb{P}_n (\Delta_{i,j,k})$ denotes the probability of a triangle between vertices i, j, k being present. This probability can be written as

$$\begin{aligned} \mathbb{P}_n (\Delta_{i,j,k}) &= 1 - \mathbb{P}_n (X_{ij} = 0) - \mathbb{P}_n (X_{ik} = 0) - \mathbb{P}_n (X_{jk} = 0) + \mathbb{P}_n (X_{ij} = X_{ik} = 0) \\ &\quad + \mathbb{P}_n (X_{ij} = X_{jk} = 0) + \mathbb{P}_n (X_{ik} = X_{jk} = 0) \\ &\quad - \mathbb{P}_n (X_{ij} = X_{ik} = X_{jk} = 0). \end{aligned} \tag{71}$$

Because $D_i, D_j, D_k \leq \sqrt{n}/\varepsilon$ and $L_n = \mu n(1 + o(1))$ under the event \mathcal{A}_n , we can use [18, Lemma 3.1], which calculates the probability that an edge is present conditionally on the presence of other edges in CMs with arbitrary degree distributions. This results in

$$\begin{aligned} \mathbb{P}_n (X_{ij} = X_{ik} = X_{jk} = 0) &= \mathbb{P}_n (X_{ij} = 0) \mathbb{P}_n (X_{jk} = 0 \mid X_{ij} = 0) \mathbb{P}_n (X_{ik} = 0 \mid X_{ij} = X_{jk} = 0) \\ &= e^{-\frac{D_i D_j}{L_n}} e^{-\frac{D_j D_k}{L_n}} e^{-\frac{D_i D_k}{L_n}} (1 + o(1)), \end{aligned}$$

and similarly

$$\mathbb{P}_n (X_{ij} = X_{jk} = 0) = e^{-\frac{D_i D_j}{L_n}} e^{-\frac{D_j D_k}{L_n}} (1 + o(1)).$$

Combining this with (71) yields

$$\begin{aligned} \mathbb{P}_n (\Delta_{i,j,k}) &= 1 - \left(e^{-\frac{D_i D_j}{L_n}} + e^{-\frac{D_i D_k}{L_n}} + e^{-\frac{D_j D_k}{L_n}} \right) (1 + o(1)) \\ &\quad + \left(e^{-\frac{D_i D_j}{L_n}} e^{-\frac{D_j D_k}{L_n}} + e^{-\frac{D_i D_j}{L_n}} e^{-\frac{D_i D_k}{L_n}} + e^{-\frac{D_i D_k}{L_n}} e^{-\frac{D_j D_k}{L_n}} \right) (1 + o(1)) \\ &\quad - e^{-\frac{D_i D_j}{L_n}} e^{-\frac{D_j D_k}{L_n}} e^{-\frac{D_i D_k}{L_n}} (1 + o_{\mathbb{P}}(1)) \\ &= (1 + o(1)) \left(1 - e^{-\frac{D_i D_j}{L_n}} \right) \left(1 - e^{-\frac{D_j D_k}{L_n}} \right) \left(1 - e^{-\frac{D_i D_k}{L_n}} \right), \end{aligned} \tag{72}$$

where the second equality follows because $\varepsilon\sqrt{n} < D_i, D_j, D_k < \sqrt{n}/\varepsilon$ and $L_n = \mu n(1 + o(1))$ under \mathcal{A}_n . For $D_i, D_j, D_k \in [\varepsilon\sqrt{n}, 1/\varepsilon\sqrt{n}]$, the main term in (72) can be uniformly bounded from above and from below by some functions $f_1(\varepsilon)$ and $f_2(\varepsilon)$ not depending on n . Combining this with (70) shows that

$$\mathbb{E}_n [\Delta_n(B_n(\varepsilon))] = (1 + o(1)) \sum'_{i,j,k} g_{n,\varepsilon}(D_i, D_j, D_k), \tag{73}$$

on the event \mathcal{A}_n , which proves the lemma. □

We now replace the L_n inside the definition of $g_{n,\varepsilon}$ by μn . In particular, define

$$f_{n,\varepsilon}(x, y, z) = \left(1 - e^{-\frac{xy}{\mu n}} \right) \left(1 - e^{-\frac{yz}{\mu n}} \right) \left(1 - e^{-\frac{zx}{\mu n}} \right) \mathbb{1}_{\{x,y,z \in B_n(\varepsilon)\}}; \tag{74}$$

then we have the following result.

Lemma 4.4. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$ and $\kappa = (\gamma - 1)/(1 + \gamma) > 0$. Then, for all $\varepsilon > 0$ and $\delta < \kappa$,*

$$n^{\frac{3\gamma}{2}-3+\delta} \left| \sum_{1 \leq i < j < k \leq n} g_{n,\varepsilon}(D_i, D_j, D_k) - f_{n,\varepsilon}(D_i, D_j, D_k) \right| \xrightarrow{\mathbb{P}} 0.$$

Proof. Let \mathcal{A}_n be as in (42) and note that κ is the same as in the statement of the lemma. Then, since $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$, it is enough to prove the result conditioned on \mathcal{A}_n . Next, note that $1 - e^{-a} \leq 1$ for all $a \geq 0$. Hence, by symmetry it suffices to prove

$$n^{\frac{3\gamma}{2}-3+\delta} \sum_{1 \leq i < j < k \leq n} \left| e^{-\frac{D_i D_j}{L_n}} - e^{-\frac{D_i D_j}{\mu n}} \right| \mathbb{1}_{\{D_i, D_j, D_k \in B_n(\varepsilon)\}} \xrightarrow{\mathbb{P}} 0.$$

For this we compute that, on the event \mathcal{A}_n ,

$$\left| e^{-\frac{D_i D_j}{L_n}} - e^{-\frac{D_i D_j}{\mu n}} \right| \leq D_i D_j \frac{|L_n - \mu n|}{\mu n L_n} \leq D_i D_j \frac{n^{-1-\kappa}}{(\mu^2 - n^{-\kappa})} = D_i D_j O\left(n^{-1-\kappa}\right).$$

We recall that by Karamata’s theorem [4, Theorem 1.5.11] it follows that, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{E}\left[D_1 \mathbb{1}_{\{D_1 \in B_n(\varepsilon)\}}\right] &\leq \mathbb{E}\left[D_1 \mathbb{1}_{\{D_1 > \varepsilon \sqrt{\mu n}\}}\right] \sim \frac{\mathcal{L}\left(\varepsilon \sqrt{\mu n}\right) \varepsilon^{1-\gamma} (\mu n)^{\frac{1-\gamma}{2}}}{\gamma - 1} \\ &= O\left(\mathcal{L}\left(\varepsilon \sqrt{\mu n}\right) n^{\frac{1-\gamma}{2}}\right), \end{aligned}$$

where $a_n \sim b_n$ as $n \rightarrow \infty$ means $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

Therefore, using Lemma A.1,

$$\begin{aligned} &\mathbb{E}\left[\left| e^{-\frac{D_1 D_2}{L_n}} - e^{-\frac{D_1 D_2}{\mu n}} \right| \mathbb{1}_{\{D_1, D_2, D_3 \in B_n(\varepsilon)\}}\right] \\ &\leq O\left(n^{-1-\kappa}\right) \mathbb{E}\left[D_1 \mathbb{1}_{\{D_1 \in B_n(\varepsilon)\}}\right]^2 \mathbb{P}\left(D_3 \in B_n(\varepsilon)\right) \\ &= O\left(\mathcal{L}\left(\varepsilon \sqrt{\mu n}\right)^2 \mathcal{L}\left(\sqrt{\mu n}\right) n^{-(\kappa + \frac{3\gamma}{2})}\right), \end{aligned}$$

so that by Markov’s inequality, we obtain that for any $K > 0$ and $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}\left(n^{\frac{3\gamma}{2}-3+\delta} \sum_{1 \leq i < j < k \leq n} \left| e^{-\frac{D_i D_j}{L_n}} - e^{-\frac{D_i D_j}{\mu n}} \right| \mathbb{1}_{\{D_i, D_j, D_k \in B_n(\varepsilon)\}} > K, \mathcal{A}_n\right) \\ &\leq \frac{\binom{n}{3} n^{\frac{3\gamma}{2}-3+\delta}}{K} \mathbb{E}\left[\left| e^{-\frac{D_1 D_2}{L_n}} - e^{-\frac{D_1 D_2}{\mu n}} \right| \mathbb{1}_{\mathcal{A}_n} \mathbb{1}_{\{D_1, D_2, D_3 \in B_n(\varepsilon)\}}\right] \\ &= O\left(\mathcal{L}\left(\varepsilon \sqrt{\mu n}\right)^2 \mathcal{L}\left(\sqrt{\mu n}\right) n^{\delta-\kappa}\right) = o(1). \quad \square \end{aligned}$$

We now again study the expected number of triangles of vertices with sampled degrees in $B_n(\varepsilon)$, and investigate the behavior of the expression of Lemma 4.3.

Lemma 4.5. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, and let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Let $\Delta_n(B_n(\varepsilon))$ denote the number of triangles in \widehat{G}_n with sampled degrees in $B_n(\varepsilon)$. Then, as $n \rightarrow \infty$,*

$$\begin{aligned} & \frac{\mathbb{E}_n [\Delta_n(B_n(\varepsilon))]}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma}} \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{6} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} h(t_1, t_2, t_3) dt_1 dt_2 dt_3 + O_{\mathbb{P}}(\varepsilon), \end{aligned} \tag{75}$$

where $h(x, y, z) := (1 - e^{-xy})(1 - e^{-xz})(1 - e^{-yz})$.

Proof. Combining Lemmas 4.3 and 4.4 yields that conditionally on the sampled degrees $(D_i)_{i \in [n]}$,

$$\mathbb{E}_n [\Delta_n(B_n(\varepsilon))] = (1 + o_{\mathbb{P}}(1)) \sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_i, D_j, D_k) + o_{\mathbb{P}}\left(\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}\right), \tag{76}$$

where $f_{n,\varepsilon}(D_i, D_j, D_k)$ is as in (74). To investigate the convergence of (81) when taking the random degrees into account, define for $b > a \geq 0$ the random measure

$$N_1^{(n)}([a, b]) = \frac{(\mu n)^{\frac{\gamma}{2}}}{\mathcal{L}(\sqrt{\mu n})} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i \in [a\sqrt{\mu n}, b\sqrt{\mu n}]\}},$$

so that $N_1^{(n)}$ counts the number of vertices with degrees in the interval $[a\sqrt{\mu n}, b\sqrt{\mu n}]$. Since every D_i is drawn i.i.d. from (1), the number of vertices with degrees between $a\sqrt{\mu n}$ and $b\sqrt{\mu n}$ is binomially distributed, and therefore this number concentrates around its mean value, which is large. Combining this with Lemma A.1 yields

$$\begin{aligned} N_1^{(n)}([a, b]) &= \frac{(\mu n)^{\frac{\gamma}{2}}}{\mathcal{L}(\sqrt{\mu n})} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i \in [a\sqrt{\mu n}, b\sqrt{\mu n}]\}} \\ &= \frac{(\mu n)^{\frac{\gamma}{2}}}{\mathcal{L}(\sqrt{\mu n})} \mathbb{P}(D_i \in [a\sqrt{\mu n}, b\sqrt{\mu n}]) (1 + o_{\mathbb{P}}(1)) \\ &= (1 + o_{\mathbb{P}}(1)) \gamma \int_a^b t^{-\gamma-1} dt = (1 + o_{\mathbb{P}}(1))(a^{-\gamma} - b^{-\gamma}). \end{aligned}$$

Therefore,

$$N_1^{(n)}([a, b]) \xrightarrow{\mathbb{P}} \int_a^b t^{-\gamma-1} dt = (a^{-\gamma} - b^{-\gamma}) =: \lambda([a, b]). \tag{77}$$

Let $N^{(n)}$ be the product measure $N_1^{(n)} \times N_1^{(n)} \times N_1^{(n)}$ and let $F = [\varepsilon, 1/\varepsilon]^3$. Thus, $N^{(n)}$ counts the number of triples with all three degrees proportional to $\sqrt{\mu n}$.

By (74),

$$\begin{aligned} & \sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_i, D_j, D_k) \\ &= \sum_{1 \leq i < j < k \leq n} \left(1 - e^{-\frac{D_i D_j}{\mu n}}\right) \left(1 - e^{-\frac{D_i D_k}{\mu n}}\right) \left(1 - e^{-\frac{D_j D_k}{\mu n}}\right) \mathbb{1}_{\{D_i, D_j, D_k \in B_n(\varepsilon)\}} \\ &= \frac{1}{6} \mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma} \int_F (1 - e^{-t_1 t_2})(1 - e^{-t_1 t_3})(1 - e^{-t_2 t_3}) dN^{(n)}(t_1, t_2, t_3) \\ &\quad - \frac{1}{3} \mathcal{L}(\sqrt{\mu n})^2 n^{(2-\gamma)} \mu^{-\gamma} \int_F (1 - e^{-t_1 t_2})^2 (1 - e^{-t_1^2}) dN^{(n)}(t_1, t_2) \\ &= \frac{(1 + o(1))}{6} \mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma} \int_F (1 - e^{-t_1 t_2})(1 - e^{-t_1 t_3})(1 - e^{-t_2 t_3}) dN^{(n)}(t_1, t_2, t_3), \end{aligned}$$

where the second term in the second equality excludes the summation terms where at least two indices are equal.

The function $h(x, y, z)$ is bounded and continuous on $F = [\varepsilon, 1/\varepsilon]^3$. Therefore, writing the Taylor expansion of e^{-x} yields

$$(1 - e^{-xy}) = \sum_{i=1}^k \frac{(xy)^i}{i!} (-1)^i + O\left(\frac{\varepsilon^{-2k}}{(k+1)!}\right)$$

on $x, y \in [\varepsilon, 1/\varepsilon]$, where the error term goes to zero as $k \rightarrow \infty$. Therefore, for any $\delta > 0$, we can choose k_1, k_2, k_3 such that

$$\begin{aligned} h(x, y, z) &= \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \left(\frac{(xy)^{l_1}}{l_1!} (-1)^{l_1} \frac{(yz)^{l_2}}{l_2!} (-1)^{l_2} \frac{(xz)^{l_3}}{l_3!} (-1)^{l_3} \right) + O(\delta) \\ &= \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \left(\frac{x^{l_1+l_3} y^{l_1+l_2} z^{l_2+l_3}}{l_1! l_2! l_3!} (-1)^{l_1+l_2+l_3} \right) + O(\delta). \end{aligned}$$

Choosing $\delta = \varepsilon^{3\gamma+1}$ gives

$$\begin{aligned} & \frac{\sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_i, D_j, D_k)}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma}} = \frac{1}{6} \int_F h(t_1, t_2, t_3) dN^{(n)}(t) \\ &= \frac{1}{6} \int_F \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \left(\frac{x^{l_1+l_3} y^{l_1+l_2} z^{l_2+l_3}}{l_1! l_2! l_3!} (-1)^{l_1+l_2+l_3} \right) + O(\varepsilon^{3\gamma+1}) dN^{(n)}(t) \\ &= \frac{1}{6} \int_F \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \frac{x^{l_1+l_3} y^{l_1+l_2} z^{l_2+l_3}}{l_1! l_2! l_3!} (-1)^{l_1+l_2+l_3} dN^{(n)}(t) + O_{\mathbb{P}}(\varepsilon) \\ &= \frac{1}{6} \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \frac{(-1)^{l_1+l_2+l_3}}{l_1! l_2! l_3!} \int_{\varepsilon}^{1/\varepsilon} x^{l_1+l_3} dN_1^{(n)}(x) \int_{\varepsilon}^{1/\varepsilon} y^{l_1+l_2} dN_1^{(n)}(y) \\ &\quad \cdot \int_{\varepsilon}^{1/\varepsilon} z^{l_2+l_3} dN_1^{(n)}(z) + O_{\mathbb{P}}(\varepsilon). \end{aligned}$$

Here we used that because $N_1^{(n)}[a, b] \xrightarrow{\mathbb{P}} \lambda[a, b]$,

$$\begin{aligned} \int_F \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \varepsilon^4 dN^{(n)}(t) &= k_1 k_2 k_3 \varepsilon^{3\gamma+1} (N_1^{(n)}([\varepsilon, 1/\varepsilon]))^3 \\ &= \varepsilon^{3\gamma+1} O_{\mathbb{P}} \left(\lambda([\varepsilon, 1/\varepsilon])^3 \right) = \varepsilon^{3\gamma+1} O_{\mathbb{P}} \left((\varepsilon^{-\gamma} - \varepsilon^{\gamma})^3 \right) = O_{\mathbb{P}}(\varepsilon). \end{aligned}$$

Since x^t is bounded and continuous on $[\varepsilon, 1/\varepsilon]$, we may use [19, Lemma 5] to conclude that

$$\int_{\varepsilon}^{1/\varepsilon} x^t dN_1^{(n)}(x) \xrightarrow{\mathbb{P}} \int_{\varepsilon}^{1/\varepsilon} x^t d\lambda(x) := \varphi_{\varepsilon}(t).$$

Thus,

$$\begin{aligned} &\frac{\sum_{1 \leq i < j < k \leq n} f_{n,\varepsilon}(D_i, D_j, D_k)}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma}} \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{6} \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \frac{(-1)^{l_1+l_2+l_3}}{l_1! l_2! l_3!} \varphi_{\varepsilon}(l_1 + l_3) \varphi_{\varepsilon}(l_1 + l_2) \varphi_{\varepsilon}(l_2 + l_3) + O_{\mathbb{P}}(\varepsilon) \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{6} \int_F \sum_{l_1=1}^{k_1} \sum_{l_2=1}^{k_2} \sum_{l_3=1}^{k_3} \frac{x^{l_1+l_3} y^{l_1+l_2} z^{l_2+l_3} (-1)^{l_1+l_2+l_3}}{l_1! l_2! l_3!} d\lambda(x) d\lambda(y) d\lambda(z) + O_{\mathbb{P}}(\varepsilon) \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{6} \int_F h(x, y, z) d\lambda(x) d\lambda(y) d\lambda(z) + O_{\mathbb{P}}(\varepsilon), \end{aligned}$$

which concludes the proof together with (76). □

The following lemma bounds the variance of the number of triangles between vertices of sampled degrees proportional to \sqrt{n} .

Lemma 4.6. *Let \mathbf{D}_n be sampled from \mathcal{D} with $1 < \gamma < 2$, and let $\widehat{G}_n = \text{ECM}(\mathbf{D}_n)$. Let $\Delta_n(B_n(\varepsilon))$ denote the number of triangles in \widehat{G}_n with sampled degrees in $B_n(\varepsilon)$. Then, as $n \rightarrow \infty$,*

$$\frac{\text{Var}_n(\Delta_n(B_n(\varepsilon)))}{\mathbb{E}_n[\Delta_n(B_n(\varepsilon))]^2} \xrightarrow{\mathbb{P}} 0.$$

Proof. Choose $0 < \delta < \frac{1}{5}(1 - \frac{1}{2}\gamma)$. Define

$$\mathcal{B}_n = \left\{ \sum_{i=1}^n \mathbb{1}_{\{D_i > \varepsilon \sqrt{\mu n}\}} < n^{\frac{1}{2}(2-\gamma)+\delta} \right\},$$

let \mathcal{A}_n be as defined in (42), and set $\Lambda_n = \mathcal{A}_n \cap \mathcal{B}_n$. Because the sampled degrees are i.i.d. samples from (1), $\mathbb{P}(\Lambda_n) \rightarrow 1$. Therefore, we work on the event Λ_n in the rest of the proof. By Lemma 4.5, $\mathbb{E}_n[\Delta_n(B_n(\varepsilon))] = O_{\mathbb{P}}(n^{3-\frac{3}{2}\gamma} \mathcal{L}(\sqrt{n}))^3$. Thus, we need to prove that

$$\text{Var}_n(\Delta_n(B_n(\varepsilon))) = o_{\mathbb{P}} \left(n^{6-3\gamma} \mathcal{L}(\sqrt{n})^6 \right).$$

We can write the variance $\text{Var}_n(\Delta_n(B_n(\varepsilon)))$ as

$$\sum_{(i,j,k),(s,t,u)} \left(\mathbb{P}_n(\Delta_{i,j,k}, \Delta_{s,t,u}) - \mathbb{P}_n(\Delta_{i,j,k}) \mathbb{P}_n(\Delta_{s,t,u}) \right) \mathbb{1}_{\{D_i, D_j, D_j, D_s, D_t, D_u \in B_n(\varepsilon)\}}, \tag{78}$$

where $\Delta_{i,j,k}$ denotes the event that there is a triangle between vertices i, j , and k . This splits into several cases, depending on $|\{i, j, k, s, t, u\}|$. Let the part of the variance where $|\{i, j, k, s, t, u\}| = m$ be denoted by $V^{(m)}$, so that

$$\text{Var}_n (\Delta_n(B_n(\varepsilon))) = V^{(6)} + V^{(5)} + V^{(4)} + V^{(3)}.$$

Let $M_n(\varepsilon) = \sum_i \mathbb{1}_{\{D_i \in B_n(\varepsilon)\}}$. Then

$$V^{(m)} \leq \sum_{i,j,k,s,t: |\{i,j,k,s,t\}|=m} \mathbb{1}_{\{D_i, D_j, D_j, D_s, D_t \in B_n(\varepsilon)\}} = M_n(\varepsilon)^m, \quad m = 3, 4, 5.$$

Under B_n ,

$$M_n(\varepsilon) \leq \sum_{i=1}^n \mathbb{1}_{\{D_i > \varepsilon \sqrt{\mu n}\}} < n^{\frac{1}{2}(2-\gamma)+\delta}.$$

Thus, by the choice of δ ,

$$V^{(5)} \leq M_n(\varepsilon)^5 = O\left(n^{5-5\gamma/2+5\delta}\right) = o(n^{6-3\gamma}),$$

as required. Similar bounds show that $V^{(4)}$ and $V^{(3)}$ are $o(n^{6-3\gamma})$. Thus, the contributions of $V^{(m)}$, $m = 3, 4, 5$, to the variance are sufficiently small.

Now we investigate the case where six different indices are involved and show that $V^{(6)} = o_{\mathbb{P}}(n^{6-3\gamma} \mathcal{L}(\sqrt{n})^6)$. Equation (73) computes the second term inside the brackets in (78). To compute the first term inside the brackets, we make a very similar computation that leads to (73). A similar computation as in (72) yields that on the event \mathcal{A}_n ,

$$\mathbb{P}_n (\Delta_{i,j,k}, \Delta_{s,t,u}) = (1 + o_{\mathbb{P}}(1))g_{n,\varepsilon}(D_i, D_j, D_k)g_{n,\varepsilon}(D_s, D_t, D_u).$$

Hence,

$$\mathbb{P}_n (\Delta_{i,j,k}, \Delta_{s,t,u}) - \mathbb{P}_n (\Delta_{i,j,k}) \mathbb{P}_n (\Delta_{s,t,u}) = o_{\mathbb{P}}(1)g_{n,\varepsilon}(D_i, D_j, D_k)g_{n,\varepsilon}(D_s, D_t, D_u).$$

When $D_i, D_j, D_k \in [\varepsilon, 1/\varepsilon]\sqrt{n}$, we have $g_{n,\varepsilon}(D_i, D_j, D_k) \in [f_1(\varepsilon), f_2(\varepsilon)]$, uniformly in i, j, k . Therefore, by Lemma 4.3,

$$\begin{aligned} V^{(6)} &= \sum_{(i,j,k),(s,t,u)} o(1)g_{n,\varepsilon}(D_i, D_j, D_k)g_{n,\varepsilon}(D_s, D_t, D_u) \\ &= o_{\mathbb{P}}\left(\mathbb{E}_n [\Delta_n(B_n(\varepsilon))]^2\right) = o_{\mathbb{P}}\left(n^{6-3\gamma} \mathcal{L}(\sqrt{n})^6\right), \end{aligned}$$

and therefore also the contribution to the variance where $|\{i, j, k, s, t, u\}| = 6$ is small enough. □

Proof of Theorem 1.4. First, we look at the denominator of the clustering coefficient in (4). By (15) and the stable-law central limit theorem [42, Theorem 4.5.2], there exists a slowly-varying function \mathcal{L}_0 such that

$$\frac{\sum_{i=1}^n \widehat{D}_i(\widehat{D}_i - 1)}{\mathcal{L}_0(n)n^{2/\gamma}} = \frac{\sum_{i=1}^n D_i^2}{\mathcal{L}_0(n)n^{2/\gamma}}(1 + o_{\mathbb{P}}(1)) \xrightarrow{d} \mathcal{S}_{\gamma/2}, \tag{79}$$

where $\mathcal{S}_{\gamma/2}$ is a stable distribution.

Now we consider the numerator of the clustering coefficient. We prove the convergence of the number of triangles in several steps. First, we show that the major contribution to the number of triangles comes from the triangles between vertices with degrees proportional to \sqrt{n} . Fix $\varepsilon > 0$. We use (64), where we want to show that the contribution of $\Delta_n(\bar{B}_n(\varepsilon))$ is negligible. Applying Lemma 4.2 with the Markov inequality yields, for every $\delta > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\Delta_n(\bar{B}_n(\varepsilon)) > \delta \mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \right) = O \left(\frac{\mathcal{E}_1(\varepsilon)}{\delta} \right).$$

Therefore,

$$\Delta_n(\bar{B}_n(\varepsilon)) = O_{\mathbb{P}} \left(\mathcal{E}_1(\varepsilon) \mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \right). \tag{80}$$

Because $\mathcal{E}_1(\varepsilon)$ tends to zero as $\varepsilon \rightarrow 0$, we now focus on $\Delta_n(B_n(\varepsilon))$. The number of triangles consists of two sources of randomness: the random pairing of the edges, and the random degrees. First we show that $\Delta_n(B_n(\varepsilon))$ concentrates around its mean when conditioned on the degrees. By Lemma 4.6 and Chebyshev’s inequality,

$$\frac{\Delta_n(B_n(\varepsilon))}{\mathbb{E}_n[\Delta_n(B_n(\varepsilon))]} \xrightarrow{\mathbb{P}} 1,$$

conditionally on the degree sequence $(D_i)_{i \in [n]}$. Combining this with Lemma 4.5 yields that conditionally on $(D_i)_{i \in [n]}$,

$$\begin{aligned} \Delta_n(B_n(\varepsilon)) &= \mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma} \\ &\times \left((1 + o_{\mathbb{P}}(1)) \frac{1}{6} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} h(t_1, t_2, t_3) dt_1 dt_2 dt_3 + o_{\mathbb{P}}(\varepsilon) \right), \end{aligned} \tag{81}$$

where $h(x, y, z) := (1 - e^{-xy})(1 - e^{-xz})(1 - e^{-yz})$ and ε is the same as in (80).

Combining (80) and (81) gives

$$\begin{aligned} \frac{\Delta_n}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}} &= \frac{\Delta_n(B_n(\varepsilon)) + \Delta_n(\bar{B}_n(\varepsilon))}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}} \\ &= (1 + o_{\mathbb{P}}(1)) \mu^{-\frac{3}{2}\gamma} \frac{1}{6} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \int_{\varepsilon}^{1/\varepsilon} \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} h(t_1, t_2, t_3) dt_1 dt_2 dt_3 \\ &\quad + O_{\mathbb{P}}(\mathcal{E}_1(\varepsilon)) + O_{\mathbb{P}}(\varepsilon). \end{aligned}$$

Taking the limit as $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$ combined with (79) and the definition of the clustering coefficient in (4) then results in

$$\frac{\hat{C}_n \mathcal{L}_0(n)}{\mathcal{L}(\sqrt{\mu n})^3 n^{(-3\gamma^2+6\gamma-4)/(2\gamma)}} = 6 \frac{\mathcal{L}_0(n) n^{2/\gamma}}{\sum_{i \in [n]} \hat{D}_i (\hat{D}_i - 1)} \frac{\Delta_n}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)}} \tag{82}$$

$$\xrightarrow{d} \mu^{-\frac{3}{2}\gamma} \frac{\int_0^\infty \int_0^\infty \int_0^\infty \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} h(t_1, t_2, t_3) dt_1 dt_2 dt_3}{\mathcal{S}_{\gamma/2}}. \tag{83}$$

We then compute the triple integral by making the change of variables $t_1 = \sqrt{ac/b}$, $t_2 = \sqrt{ab/c}$, $t_3 = \sqrt{bc/a}$. The Jacobian $J(a, b, c)$ of this change of variables equals

$$J(a, b, c) = \frac{1}{2\sqrt{abc}}.$$

Then

$$\begin{aligned} & \int_0^\infty \int_0^\infty \int_0^\infty \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} h(t_1, t_2, t_3) dt_1 dt_2 dt_3 \\ &= \int_0^\infty \int_0^\infty \int_0^\infty \frac{\gamma^3}{(t_1 t_2 t_3)^{\gamma+1}} (1 - e^{-t_1 t_2})(1 - e^{-t_1 t_3})(1 - e^{-t_2 t_3}) dt_1 dt_2 dt_3 \\ &= \gamma^3 \int_0^\infty \int_0^\infty \int_0^\infty (\sqrt{abc})^{-1-\gamma} (1 - e^{-a})(1 - e^{-b})(1 - e^{-c}) \frac{1}{2\sqrt{abc}} da db dc \\ &= \frac{\gamma^3}{2} \left(\int_0^\infty a^{-1-\gamma/2} (1 - e^{-a}) da \right)^3 = \frac{\gamma^3}{2} \left(-\Gamma \left(-\frac{\gamma}{2} \right) \right)^3, \end{aligned} \tag{84}$$

where we have used that

$$\begin{aligned} \int_0^\infty a^{-1-\gamma/2} (1 - e^{-a}) da &= \int_0^\infty \int_0^a a^{-1-\gamma/2} e^{-y} dy da = \int_0^\infty \int_y^\infty a^{-1-\gamma/2} e^{-y} da dy \\ &= -\frac{2}{\gamma} \int_0^\infty y^{-\gamma/2} e^{-y} dy = -\frac{\Gamma(\gamma/2 + 1)}{\gamma/2} = -\Gamma(\gamma/2). \end{aligned}$$

Combining this with (82) proves Theorem 1.4. □

5. Proofs of Theorems 1.2 and 1.5

We now show how the proofs of Theorems 1.1 and 1.4 can be adapted to prove similar results for the class of rank-1 inhomogeneous random graphs. We remind the reader that in this setting the degrees D_i are no longer an i.i.d. sample from (1).

Let $\kappa \leq (\gamma - 1)/(1 + \gamma)$, $0 < \delta < 1 - 1/\gamma$ and define the events

$$\mathcal{A}_n = \left\{ \left| \sum_{i \in [n]} w_i - \mu n \right| \leq n^{1-\kappa} \right\}, \quad \mathcal{B}_n = \left\{ \left| \sum_{i=1}^n w_i \mathbb{1}_{\{w_i < n^\delta\}} - \mu n \right| \leq \frac{n}{\log(n)} \right\}.$$

Let $\Lambda_n = \mathcal{A}_n \cap \mathcal{B}_n$. Because $\mathbb{P}(\Lambda_n) \rightarrow 0$, we condition on the event Λ_n . For the ECM, we used that the degrees and the erased degrees are close to prove Theorems 1.1 and 1.4. Similarly, for the inhomogeneous random graphs we now show that the weights and the degrees are close. By Condition 1.1(i),

$$\mathbb{E}_n [D_i] = \sum_{j \neq i} \frac{w_i w_j}{\mu n} h \left(\frac{w_i w_j}{\mu n} \right) \leq \sum_{j \in [n]} \frac{w_i w_j}{\mu n} = w_i (1 + o(1)), \tag{85}$$

where \mathbb{E}_n now denotes expectation conditioned on the weight sequence. Furthermore, for $w_i = O(n^{1/\gamma})$, on the event \mathcal{B}_n , by Condition 1.1(ii)–(iii),

$$\begin{aligned} \mathbb{E}_n [D_i] &\geq \sum_{j: w_j < n^{\frac{\gamma-1}{2\gamma}}} \frac{w_j w_i}{\mu n} h \left(\frac{w_i w_j}{\mu n} \right) = \sum_{j: w_j < n^{\frac{\gamma-1}{2\gamma}}} \frac{w_j w_i}{\mu n} \left(1 + o \left(\frac{w_i w_j}{\mu n} \right) \right) \\ &= (1 + o(1)) \sum_{j: w_j < n^{\frac{\gamma-1}{2\gamma}}} \frac{w_j w_i}{\mu n} = w_i (1 + o(1)), \end{aligned} \tag{86}$$

where the first equality follows from a first-order Taylor expansion of $h(x)$. Combining (85) and (86) yields $\mathbb{E}_n [D_i] = w_i (1 + o(1))$.

Let X_{ij} again denote the indicator that edge $\{i, j\}$ is present. Note that $\text{Var}_n(X_{ij}) = p(w_i, w_j)(1 - p(w_i, w_j)) \leq p(w_i, w_j)$. Because conditioned on the weights, the degree of a vertex $D_i = \sum_{i \neq j} X_{ij}$ is the sum of independent indicators with success probability $p(w_i, w_j)$, $\text{Var}_n(D_i) \leq \sum_{i \neq j} p(w_i, w_j) \leq 2w_i$ for n large enough. Then, Bernstein's inequality yields that for $t > 0$,

$$\mathbb{P}_n(|D_i - w_i| > t) \leq \exp\left(-\frac{t^2/2}{w_i + t/3}\right).$$

Thus, for $w_i > \log(n)$,

$$D_i = w_i(1 + o_{\mathbb{P}}(1)).$$

Therefore,

$$\begin{aligned} \sum_{i \in [n]} |D_i^2 - w_i^2| &= \sum_{i: w_i \leq \log(n)} |D_i^2 - w_i^2| + \sum_{i: w_i > \log(n)} |D_i^2 - w_i^2| \\ &= O_{\mathbb{P}}\left(n \log^2(n)\right) + o_{\mathbb{P}}\left(\sum_{i: w_i > \log(n)} w_i^2\right) = o_{\mathbb{P}}\left(\sum_{i \in [n]} w_i^2\right), \end{aligned}$$

and a similar result holds for the third moment of the degrees. In particular, this implies that (27) and (12) also hold for the inhomogeneous random graph under Condition 1.1.

5.1 Pearson in the rank-1 inhomogeneous random graph

The analysis of the term r_n^- in (11) is the same as in the ECM, since it only depends on the degrees, and (27) also holds for the inhomogeneous random graph. We therefore only need to show that Proposition 3.3 also holds for the rank-1 inhomogeneous random graph. This means that we need to show that (41) also holds for the rank-1 inhomogeneous random graph. For all models satisfying Condition 1.1, $p(w_i, w_j) \leq (w_i w_j / (\mu n) \wedge 1)$. Because $\mathbb{E}_n[D_i] = w_i(1 + o(1))$, $D_i = O_{\mathbb{P}}(w_i)$, by Markov's inequality. Thus,

$$\sum_{1 \leq i < j \leq n} D_i D_j X_{ij} = O_{\mathbb{P}}\left(\sum_{1 \leq i < j \leq n} w_i w_j X_{ij}\right) = O_{\mathbb{P}}\left(\sum_{1 \leq i < j \leq n} w_i w_j \left(\frac{w_i w_j}{\mu n} \wedge 1\right)\right).$$

Because the weights are sampled from (1), this is exactly the same bound as in (45), so that from there we can follow the same lines as the proof of Proposition 3.3. Thus, Proposition 3.3 also holds for rank-1 inhomogeneous random graphs satisfying Condition 1.1. Then we can follow the same lines as the proof of Theorem 1.1 to prove Theorem 1.2.

5.2. Clustering in the rank-1 inhomogeneous random graph

For the clustering coefficient, note that conditioned on the weights, $\mathbb{P}_n(\Delta_{i,j,k}) = p(w_i, w_j)p(w_j, w_k)p(w_i, w_k)$. Furthermore, Lemma 4.2 only requires the bound $p(w_i, w_j) \leq (w_i w_j / (\mu n) \wedge 1)$, which also holds for all rank-1 inhomogeneous random graphs satisfying Condition 1.1, so that Lemma 4.2 also holds for these rank-1 inhomogeneous random graphs. Furthermore, conditioned on the weights, the probabilities of distinct edges being present are independent, so that

$$\mathbb{E}_n[\Delta_n(B_n(\varepsilon))] = \sum_{1 \leq i < j < k \leq n} q\left(\frac{w_i w_j}{\mu n}\right) q\left(\frac{w_j w_k}{\mu n}\right) q\left(\frac{w_i w_k}{\mu n}\right) \mathbb{1}_{\{(w_i, w_j, w_k) \in B_n(\varepsilon)\}},$$

similarly to Lemma 4.3, with q as defined in Condition 1.1. Furthermore, the bound on the variance of the number of triangles in the ECM in Lemma 4.6 for three, four, or five contributing vertices only depends on the degrees, so that it also holds for the rank-1 inhomogeneous random graph satisfying Condition 1.1, since the weights are also sampled from (1). The contribution of six different vertices to the variance is zero, because the presence of distinct edges is independent. Thus, Lemma 4.6 also holds for the rank-1 inhomogeneous random graph. Thus, we can follow the lines of the proof of Theorem 1.4 up to Equation (77). From there, note that

$$\frac{\sum_{1 \leq i < j < k \leq n} q\left(\frac{w_i w_j}{\mu n}\right) q\left(\frac{w_i w_k}{\mu n}\right) q\left(\frac{w_j w_k}{\mu n}\right) \mathbb{1}_{\{(w_i, w_j, w_k) \in B_n(\varepsilon)\}}}{\mathcal{L}(\sqrt{\mu n})^3 n^{\frac{3}{2}(2-\gamma)} \mu^{-\frac{3}{2}\gamma}}$$

$$= \frac{1}{6} \int_F q(t_1 t_2) q(t_1 t_3) q(t_2 t_3) dN_1^{(n)}(t_1) dN_1^{(n)}(t_2) dN_1^{(n)}(t_3).$$

Then we use that the function $q(t_1 t_2) q(t_1 t_3) q(t_2 t_3)$ is a bounded, continuous function by Condition 1.1, so that by [19, Lemma 5],

$$\int_F q(t_1 t_2) q(t_1 t_3) q(t_2 t_3) dN_1^{(n)}(t_1) dN_1^{(n)}(t_2) dN_1^{(n)}(t_3)$$

$$\xrightarrow{\mathbb{P}} \int_F q(t_1 t_2) q(t_1 t_3) q(t_2 t_3) d\lambda(t_1) d\lambda(t_2) d\lambda(t_3).$$

Thus, we can follow the same lines as the proof of the clustering coefficient in the ECM, replacing the term $(1 - e^{-t_1 t_2})(1 - e^{-t_1 t_3})(1 - e^{-t_2 t_3})$ by $q(t_1 t_2) q(t_1 t_3) q(t_2 t_3)$, which then proves Theorem 1.5.

Appendix A

In this section we prove some technical results used in the preceding proofs.

A.1. Erased edges

We start with the proof for the scaling of the number of erased edges.

Proof of Theorem 2.1. Let $K, \delta > 0, \kappa \leq (\gamma - 1)/(1 + \gamma)$. Define the events

$$\mathcal{A}_n = \left\{ |L_n - \mu n| \leq n^{1-\kappa} \right\}, \quad \mathcal{B}_n = \left\{ \max_{1 \leq i \leq n} D_i \leq n^{\frac{1}{\gamma} + \frac{\delta}{2}} \right\}, \quad \mathcal{C}_n = \left\{ \sum_{i=1}^n D_i^2 \leq n^{\frac{2}{\gamma} + \frac{\delta}{2}} \right\},$$

and set $\Lambda_n = \mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n$. Then by Lemma 2.1 and Proposition 2.1, $\mathbb{P}(\Lambda_n) \rightarrow 1$, and hence we only need to prove the result conditioned on the event Λ_n .

First recall that

$$Z_n = \sum_{i=1}^n X_{ii} + \sum_{1 \leq i < j \leq n} \left(X_{ij} - \mathbb{1}_{\{X_{ij} > 0\}} \right).$$

We first consider the conditional expectation of the last term $\mathbb{E}_n \left[\mathbb{1}_{\{X_{ij}>0\}} \right] = 1 - \mathbb{P}_n (X_{ij} = 0)$. It follows from [14, Eq. 4.9] that

$$\begin{aligned} \mathbb{P}_n (X_{ij} = 0) &\leq \prod_{i=0}^{D_i-1} \left(1 - \frac{D_j}{L_n - 2D_i - 1} \right) + \frac{D_i^2 D_j}{(L_n - 2D_i)^2} \\ &\leq \left(1 - \frac{D_j}{L_n - 1} \right)^{D_i} + \frac{D_i^2 D_j}{(L_n - 2D_i)^2} \\ &\leq e^{-\frac{D_i D_j}{L_n - 1}} + \frac{D_i^2 D_j}{(L_n - 2D_i)^2}. \end{aligned}$$

The additional term essentially comes from the fact that we need to consider the cases where a stub of node i connects to another stub of node i .

Next, since $\mathbb{E} [X_{ii}] = D_i(D_i - 1)/(L_n - 1)$ and $\mathbb{E} [X_{ij}] = D_i D_j / (L_n - 1)$ we have

$$\mathbb{E}_n [Z_n] \leq \sum_{1 \leq i < j \leq n} \phi \left(\frac{D_i D_j}{L_n - 1} \right) + \sum_{i=1}^n \frac{D_i^2 - D_i}{L_n - 1} + \sum_{1 \leq i < j \leq n} \frac{D_i^2 D_j}{(L_n - 2D_i)^2},$$

where $\phi(x) = x - 1 + e^{-x}$. Define

$$M_n = \sum_{i=1}^n \frac{D_i^2 - D_i}{L_n - 1} + \sum_{1 \leq i < j \leq n} \frac{D_i^2 D_j}{(L_n - 2D_i)^2}.$$

Then, on the event Λ_n ,

$$\begin{aligned} M_n &\leq \frac{n^{\frac{2}{\gamma} + \frac{\delta}{2}}}{\mu n - n^{1-\kappa} - 1} + \frac{(\mu n - n^{1-\kappa}) n^{\frac{2}{\gamma} + \frac{\delta}{2}}}{\left(\mu n - n^{1-\kappa} - 2n^{\frac{1}{\gamma} + \frac{\delta}{2}} \right)^2} \\ &= O \left(n^{\frac{2}{\gamma} - 1 + \frac{\delta}{2}} \right). \end{aligned}$$

Note that Λ_n is completely determined by the degree sequence. Hence, by Markov’s inequality, we get

$$\begin{aligned} \mathbb{P} \left(Z_n > n^{2-\gamma+\delta}, \Lambda_n \right) &\leq n^{\gamma-2-\delta} \mathbb{E} [Z_n \mathbb{1}_{\Lambda_n}] = n^{\gamma-2-\delta} \mathbb{E} [\mathbb{E}_n [Z_n] \mathbb{1}_{\Lambda_n}] \\ &\leq n^{\gamma-2-\delta} \mathbb{E} \left[\sum_{1 \leq i < j \leq n} \phi \left(\frac{D_i D_j}{L_n - 1} \right) \mathbb{1}_{\{\Lambda_n\}} \right] + n^{\gamma-2-\delta} \mathbb{E} [M_n \mathbb{1}_{\Lambda_n}] \\ &\leq n^{\gamma-2-\delta} \mathbb{E} \left[\sum_{1 \leq i < j \leq n} \phi \left(\frac{D_i D_j}{\mu n - 1 - n^{1-\kappa}} \right) \right] + O \left(n^{\gamma + \frac{2}{\gamma} - 3 - \frac{\delta}{2}} \right) \\ &\leq n^{\gamma-\delta} \mathbb{E} \left[\phi \left(\frac{D_1 D_2}{\mu n - 1 - n^{1-\kappa}} \right) \right] + O \left(n^{\gamma + \frac{2}{\gamma} - 3 - \frac{\delta}{2}} \right) \end{aligned}$$

as $n \rightarrow \infty$. Since

$$\gamma + \frac{2}{\gamma} - 3 < 0$$

for all $1 < \gamma < 2$, the last term goes to zero as $n \rightarrow \infty$. For the other term we note that D_1 and D_2 are independent regularly-varying random variables with exponent $1 < \gamma < 2$, and therefore so is $D_1 D_2$. It then follows from [3] that for any $\delta > 0$

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E} \left[\phi \left(\frac{D_1 D_2}{t} \right) \right]}{t^{-\gamma+\delta}} = 0.$$

Taking $t = \mu n - 1 - n^{1-\kappa}$ we obtain that

$$\lim_{n \rightarrow \infty} n^\gamma \mathbb{E} \left[\phi \left(\frac{D_1 D_2}{\mu n - 1 - n^{1-\kappa}} \right) \right] = 0,$$

from which it follows that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(E_n > n^{2-\gamma+\delta}, \Lambda_n \right) = 0.$$

We proceed with the proof of Corollary 2.1. □

Proof of Corollary 2.1. For the first part we write

$$\sum_{i=1}^n D_i^p Y_i \leq \max_{1 \leq j \leq n} D_j^p \sum_{i=1}^n Y_i \leq 2Z_n \max_{1 \leq j \leq n} D_j^p,$$

and hence, using Theorem 2.1 and Proposition 2.1,

$$\frac{\sum_{i=1}^n D_i^p Y_i}{n^{\frac{p}{\gamma}+2-\gamma+\delta}} \leq \left(\frac{2Z_n}{n^{2-\gamma+\frac{\delta}{2}}} \right) \left(\frac{\max_{1 \leq j \leq n} D_j^p}{n^{\frac{p}{\gamma}+\frac{\delta}{2}}} \right) \xrightarrow{\mathbb{P}} 0.$$

For the second part we bound the main term by

$$\sum_{1 \leq i < j \leq n} X_{ij} D_i D_j \leq \max_{1 \leq j \leq n} D_j \sum_{1 \leq i < j \leq n} X_{ij} D_i \leq \frac{1}{2} \max_{1 \leq j \leq n} D_j \sum_{i=1}^n D_i Y_i.$$

Hence, using the first part of the corollary and Proposition 2.1, it follows that

$$\frac{\sum_{1 \leq i < j \leq n} X_{ij} D_i D_j}{n^{\frac{2}{\gamma}+2-\gamma+\delta}} \leq \left(\frac{\max_{1 \leq j \leq n} D_j}{2n^{\frac{1}{\gamma}+\frac{\delta}{2}}} \right) \left(\frac{\sum_{i=1}^n D_i Y_i}{n^{\frac{1}{\gamma}+2-\gamma+\frac{\delta}{2}}} \right) \xrightarrow{\mathbb{P}} 0. \quad \square$$

A.2. Technical results for clustering

The following result is needed for the proof of Lemma 4.2.

Lemma A.1. *Let X be a nonnegative regularly-varying random variable with distribution (1). Then, for any $0 \leq a < b$,*

$$\mathbb{P} \left(X \in [a, b] \sqrt{n} \right) = \mathcal{L}(\sqrt{\mu n}) n^{-\gamma/2} \int_a^b x^{-\gamma-1} dx (1 + o(1)).$$

Proof. Because \mathcal{L} is a slowly-varying function,

$$\mathcal{L}(c\sqrt{n}) = \mathcal{L}(\sqrt{\mu n})(1 + o(1))$$

for any $c \in (0, \infty)$. Furthermore, using the Taylor expansion of $(a\sqrt{n} - x)^{-\gamma}$ at $x = 0$ yields

$$(b\sqrt{n} + 1)^{-\gamma} = (b\sqrt{n})^{-\gamma} + \gamma(b\sqrt{n})^{-\gamma-1} + O(\gamma(\gamma - 1)(b\sqrt{n})^{-\gamma-2}).$$

Because \mathcal{L} is a slowly-varying function, for every constant t , $\lim_{n \rightarrow \infty} \mathcal{L}(t\sqrt{n})/\mathcal{L}(\sqrt{n}) = 1$. Thus, we obtain

$$\begin{aligned} \mathbb{P}(X \in [a, b]\sqrt{n}) &= \mathcal{L}(a\sqrt{n} - 1)(a\sqrt{n} - 1)^{-\gamma} - \mathcal{L}(b\sqrt{n} + 1)(b\sqrt{n})^{-\gamma} \\ &= \mathcal{L}(\sqrt{\mu n})(1 + o(1))(a\sqrt{n})^{-\gamma} - \mathcal{L}(\sqrt{\mu n})(1 + o(1))(b\sqrt{n})^{-\gamma} \\ &= (1 + o(1))\mathcal{L}(\sqrt{\mu n}) \left((a\sqrt{n})^{-\gamma} - (b\sqrt{n})^{-\gamma} \right) \\ &= (1 + o(1))\mathcal{L}(\sqrt{\mu n})n^{-\gamma/2} \gamma \int_a^b x^{-\gamma-1} dx. \quad \square \end{aligned}$$

Acknowledgements

The work of R. v. d. H. and C. S. was supported by NWO TOP grant 613.001.451. The work of R. v. d. H. is further supported by the NWO Gravitation Networks grant 024.002.003 and the NWO VICI grant 639.033.806. P. v. d. H. and N. L. were supported by the EU FET Open grant NADINE 288956. P. v. d. H. was further supported by ARO grant W911NF1610391.

References

- [1] ANGEL, O., VAN DER HOFSTAD, R. AND HOLMGREN, C. (2019). Limit laws for self-loops and multiple edges in the configuration model. *Ann. Inst. H. Poincaré Prob. Statist.* **55**, 1509–1530.
- [2] BANNINK, T., VAN DER HOFSTAD, R. AND STEGEHUIS, C. (2018). Switch chain mixing times and triangle counts in simple random graphs with given degrees. *J. Complex Networks* **7**, 210–225.
- [3] BINGHAM, N. H. AND DONEY, R. A. (1974). Asymptotic properties of supercritical branching processes I: the Galton–Watson process. *Adv. Appl. Prob.* **6**, 711–731.
- [4] BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1989). *Regular Variation*. Cambridge University Press.
- [5] BOGUÑÁ, M. AND PASTOR-SATORRAS, R. (2003). Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112.
- [6] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Europ. J. Combinatorics* **1**, 311–316.
- [7] BRITTON, T., DEIJFEN, M. AND MARTIN-LÖF, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Statist. Phys.* **124**, 1377–1397.
- [8] CATANZARO, M., BOGUÑÁ, M. AND PASTOR-SATORRAS, R. (2005). Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 27103.
- [9] CHUNG, F. AND LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Nat. Acad. Sci. USA* **99**, 15879–15882.
- [10] COLOMER-DE SIMON, P. AND BOGUÑÁ, M. (2012). Clustering of random scale-free networks. *Phys. Rev. E* **86**, 026120.
- [11] FEDERICO, L. AND VAN DER HOFSTAD, R. (2017). Critical window for connectivity in the configuration model. *Combinatorics Prob. Comput.* **26**, 660–680.
- [12] GAO, J., VAN DER HOFSTAD, R., SOUTHALL, A. AND STEGEHUIS, C. (2018). Counting triangles in power-law uniform random graphs. Preprint. Available at <https://arxiv.org/abs/1812.04289>.
- [13] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks*, Vol. 1. Cambridge University Press.
- [14] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. AND VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27**, 76–123.
- [15] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. AND ZNAMENSKI, D. (2007). Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Prob.* **12**, 703–766.
- [16] VAN DER HOFSTAD, R., JANSSEN, A. J. E. M., VAN LEEUWAARDEN, J. S. H. AND STEGEHUIS, C. (2017). Local clustering in scale-free networks with hidden variables. *Phys. Rev. E* **95**, 022307.
- [17] VAN DER HOFSTAD, R. AND LITVAK, N. (2014). Degree–degree dependencies in random graphs with heavy-tailed degrees. *Internet Math.* **10**, 287–334.

- [18] VAN DER HOFSTAD, R., VAN LEEUWAARDEN, J. S. H. AND STEGEHUIS, C. (2017). Optimal subgraph structures in scale-free networks. To appear in *Ann. Appl. Prob.*
- [19] VAN DER HOFSTAD, R., VAN LEEUWAARDEN, J. S. H. AND STEGEHUIS, C. (2018). Triadic closure in configuration models with unbounded degree fluctuations. *J. Statist. Phys.* **173**, 746–774.
- [20] VAN DER HOORN, P. (2016). Asymptotic analysis of network structures: degree–degree correlations and directed paths. Doctoral Thesis, University of Twente.
- [21] VAN DER HOORN, P., LIPPNER, G. AND KRIOUKOV, D. (2018). Sparse maximum-entropy random graphs with a given power-law degree distribution. *J. Statist. Phys.* **173**, 806–844.
- [22] VAN DER HOORN, P. AND LITVAK, N. (2015). Phase transitions for scaling of structural correlations in directed networks. *Phys. Rev. E* **92**, 022803.
- [23] VAN DER HOORN, P. AND LITVAK, N. (2015). Upper bounds for number of removed edges in the erased configuration model. In *Algorithms and Models for the Web Graph: WAW 2015*, Springer, Cham, pp. 54–65.
- [24] VAN DER HOORN, P. AND OLVERA-CRAVIOTO, M. (2018). Typical distances in the directed configuration model. *Ann. Appl. Prob.* **28**, 1739–1792.
- [25] JANSON, S. (2009). On percolation in random graphs with given vertex degrees. *Electron. J. Prob.* **14**, 86–118.
- [26] JANSON, S. (2009). The probability that a random multigraph is simple. *Combinatorics Prob. Comput.* **18**, 205–225.
- [27] JANSON, S. (2014). The probability that a random multigraph is simple. II. *J. Appl. Prob.* **51**, 123–137.
- [28] KARAMATA, J. (1962). Some theorems concerning slowly varying functions. Tech. Rep., University of Wisconsin–Madison Mathematics Research Center.
- [29] NEWMAN, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701.
- [30] NEWMAN, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E* **67**, 026126.
- [31] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45**, 167–256.
- [32] NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118.
- [33] NORROS, I. AND REITTU, H. (2006). On a conditionally Poissonian graph process. *Adv. Appl. Prob.* **38**, 59–75.
- [34] OSTROUMOVA PROKHORENKOVA, L. (2016). Global clustering coefficient in scale-free weighted and unweighted networks. *Internet Math.* **12**, 54–67.
- [35] OSTROUMOVA PROKHORENKOVA, L. AND SAMOSVAT, E. (2014). Global clustering coefficient in scale-free networks. In *Algorithms and Models for the Web Graph: WAW 2014*, Springer, Cham, pp. 47–58.
- [36] PARK, J. AND NEWMAN, M. E. J. (2004). Statistical mechanics of networks. *Phys. Rev. E* **70**, 066117.
- [37] PIERCE, R. D. (1997). Application of the positive alpha-stable distribution. In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, 1997*, IEEE, Banff, pp. 420–424.
- [38] RIORDAN, O. (2012). The phase transition in the configuration model. *Combinatorics Prob. Comput.* **21**, 265–299.
- [39] SAMORODNITSKY, G. AND TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall/CRC, Boca Raton.
- [40] STEGEHUIS, C. (2019). Degree correlations in scale-free random graph models. *J. Appl. Prob.* **56**, 672–700.
- [41] WATTS, D. J. AND STROGATZ, S. H. (1998). Collective dynamics of small-world networks. *Nature* **393**, 440–442.
- [42] WHITT, W. (2006). *Stochastic-Process Limits*. Springer, New York.
- [43] WORMALD, N. C. (1980). Some problems in the enumeration of labelled graphs. *Bull. Austral. Math. Soc.* **21**, 159–160.
- [44] YAO, D., VAN DER HOORN, P. AND LITVAK, N. (2018). Average nearest neighbor degrees in scale-free networks. *Internet Math. J.* **1**, 38 pp.