

A phase transition for the diameter of the configuration model

Remco van der Hofstad*
Gerard Hooghiemstra[†] and Dmitri Znamenski[‡]

March 4, 2008

Abstract

In this paper, we study the configuration model (CM) with i.i.d. degrees. We establish a phase transition for the diameter when the power-law exponent τ of the degrees satisfies $\tau \in (2, 3)$. Indeed, we show that for $\tau > 2$ and when vertices with degree 1 or 2 are present with positive probability, the diameter of the random graph is, with high probability, bounded from below by a constant times the logarithm of the size of the graph. On the other hand, assuming that all degrees are 3 or more, we show that, for $\tau \in (2, 3)$, the diameter of the graph is, with high probability, bounded from above by a constant times the log log of the size of the graph.

1 Introduction

Random graph models for complex networks have received a tremendous amount of attention in the past decade. See [1, 23, 27] for reviews on complex networks and [2] for a more expository account. Measurements have shown that many real networks share two fundamental properties. The first is the fact that typical distances between vertices are small, which is called the ‘small world’ phenomenon (see [28]). For example, in the Internet, IP-packets cannot use more than a threshold of physical links, and if the distances in terms of the physical links would be large, e-mail service would simply break down. Thus, the graph of the Internet has evolved in such a way that typical distances are relatively small, even though the Internet is rather large. The second and maybe more surprising property of many networks is that the number of vertices with degree k falls off as an inverse power of k . This is called a ‘power law degree sequence’, and resulting graphs often go under the name ‘scale-free graphs’ (see [16] for a discussion where power laws occur in the Internet).

The observation that many real networks have the above two properties has incited a burst of activity in network modelling using random graphs. These models can, roughly speaking, be divided into two distinct classes of models: ‘static’ models and ‘dynamic’ models. In static models, we model with a graph of a given size a snap-shot of a real network. A typical example of this kind of model is the *configuration model* (CM) which we describe below. A related static model, which can be seen as an inhomogeneous version of the Erdős-Rényi random graph, is treated in great generality in [4]. A much studied class of examples of ‘dynamic’ models consist of the so-called *preferential attachment models* (PAM’s), where added vertices and edges are more likely to be attached to vertices that already have large degrees. PAM’s often focus on the growth of the network as a way to explain the power law degree sequences.

Physicists have predicted that distances in PAM’s behave similarly to distances in the CM with similar degrees. Distances in the CM have attracted considerable attention (see e.g., [15, 17, 18, 19]),

*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: rhofstad@win.tue.nl

[†]Delft University of Technology, Electrical Engineering, Mathematics and Computer Science, P.O. Box 5031, 2600 GA Delft, The Netherlands. E-mail: G.Hooghiemstra@ewi.tudelft.nl

[‡]EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: znamenski@eurandom.nl

but distances in PAM's far less (see [5, 20]), which makes it hard to verify this prediction. Together with [20], the current paper takes a first step towards a rigorous verification of this conjecture. At the end of this introduction we will return to this observation, but let us first introduce the CM and present our diameter results.

1.1 The configuration model

The CM is defined as follows. Fix an integer N . Consider an i.i.d. sequence of random variables D_1, D_2, \dots, D_N . We will construct an undirected graph with N vertices where vertex j has degree D_j . We will assume that $L_N = \sum_{j=1}^N D_j$ is even. If L_N is odd, then we will increase D_N by 1. This single change will make hardly any difference in what follows, and we will ignore this effect. We will later specify the distribution of D_1 .

To construct the graph, we have N separate vertices and incident to vertex j , we have D_j stubs or half-edges. The stubs need to be paired to construct the graph. We number the stubs in a given order from 1 to L_N . We start by pairing at random the first stub with one of the $L_N - 1$ remaining stubs. Once paired, two stubs form a single edge of the graph. Hence, a stub can be seen as the left- or the right-half of an edge. We continue the procedure of randomly choosing and pairing the stubs until all stubs are connected. Unfortunately, vertices having self-loops, as well as multiple edges between vertices, may occur, so that the CM is a *multigraph*. However, self-loops are scarce when $N \rightarrow \infty$ (see e.g., [3] or [8]).

The above model is a variant of the *configuration model* [3], which, given a degree sequence, is the random graph with that given degree sequence. The degree sequence of a graph is the vector of which the k^{th} coordinate equals the fraction of vertices with degree k . In our model, by the law of large numbers, the degree sequence is close to the distribution of the nodal degree D of which D_1, \dots, D_N are i.i.d. copies.

The probability mass function and the distribution function of the nodal degree law are denoted by

$$\mathbb{P}(D = k) = f_k, \quad k = 1, 2, \dots, \quad \text{and} \quad F(x) = \sum_{k=1}^{\lfloor x \rfloor} f_k, \quad (1.1)$$

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x . We pay special attention to distributions of the form

$$1 - F(x) = x^{1-\tau} L(x), \quad (1.2)$$

where $\tau > 2$ and L is slowly varying at infinity. This means that the random variables D_j obey a power law, and the factor L is meant to generalize the model. We denote the expectation of D by μ , i.e.,

$$\mu = \sum_{k=1}^{\infty} k f_k. \quad (1.3)$$

1.2 The diameter in the configuration model

In this section we present our results on the bounds on the diameter. The distance considered in this paper is the ordinary graph distance, i.e., the distance $d(A, B)$ between the vertices A and B in graph G is the minimal number of edges of a connecting path. The average or typical distance of a (connected) graph G is the graph distance between two uniformly chosen connected vertices from the vertex set. The diameter of a graph G , denoted by $\text{diam}(G)$, is

$$\text{diam}(G) = \sup_{A, B} d(A, B), \quad (1.4)$$

where the supremum is taken over all pairs of connected vertices in the graph. We use the abbreviation **whp** for a statement that occurs with probability tending to 1 if the number of vertices of the graph N tends to ∞ .

Theorem 1.1 (Lower bound on diameter) *For $\tau > 2$, assuming that $f_1 + f_2 > 0$ and $f_1 < 1$, there exists a positive constant α such that **whp** the diameter of the configuration model is bounded below by $\alpha \log N$.*

A more precise result on the diameter in the CM is presented in [17], where it is proved that under rather general assumptions on the degree sequence of the CM, the diameter of the CM divided by $\log N$ converges to a constant. This result is also valid for related models, such as the Erdős-Rényi random graph, but the proof is quite involved. Since Theorem 1.1, together with Theorem 1.2 below, proves that the diameter of the CM has a phase transition when $\tau \in (2, 3)$, and the proof of Theorem 1.1 is substantially simpler than the more precise result in [17], we decided to include this proof. While Theorem 1.1 is substantially weaker, the fact that a *positive* constant times $\log N$ appears is most interesting, as we will discuss now in more detail.

Indeed, the result in Theorem 1.1 is most interesting in the case when $\tau \in (2, 3)$. By [19, Theorem 1.2], the average distance for $\tau \in (2, 3)$ is proportional to $\log \log N$, whereas we show here that the diameter is bounded below by a positive constant times $\log N$ when $f_1 + f_2 > 0$ and $f_1 < 1$. Therefore, we see that the average distance and the diameter are of a different order of magnitude. The pairs of vertices where the distance is of the order $\log N$ are thus scarce. The proof of Theorem 1.1 reveals that these pairs are along long lines of vertices with degree 2 that are connected to each other. Also in the proof of [17], one of the main difficulties is the identification of the precise length of these long thin lines.

Our second main result states that when $\tau \in (2, 3)$, the above assumption that $f_1 + f_2 > 0$ is *necessary and sufficient* for $\log N$ lower bounds on the diameter. In Theorem 1.2 below, we assume that there exists a $\tau \in (2, 3)$ such that, for some $c > 0$ and all $x \geq 1$,

$$1 - F(x) \geq cx^{1-\tau}, \quad (1.5)$$

which is slightly weaker than the assumption in (1.2). We further define for integer $m \geq 2$ and a real number $\sigma > 1$,

$$C_F = C_F(\sigma, m) = \frac{2}{|\log(\tau - 2)|} + \frac{2\sigma}{\log m}. \quad (1.6)$$

Then our main upper bound on the diameter when (1.5) holds is as follows:

Theorem 1.2 (A $\log \log$ upper bound on the diameter) *Fix $m \geq 2$. We assume that $\mathbb{P}(D \geq m + 1) = 1$, that $\mathbb{E}[D] = \mu < \infty$, and that (1.5) holds for some $\tau \in (2, 3)$. Then, for every $\sigma > (3 - \tau)^{-1}$, the diameter of the configuration model is, **whp**, bounded above by $C_F \log \log N$.*

Summarizing, we show for $\tau > 2$ and when vertices of degree 1 or 2 are present then a constant times $\log N$ is, **whp**, a lower bound for the diameter, whereas for $\tau \in (2, 3)$, and assuming that all degrees are at least 3, the diameter is, **whp**, bounded from above by a constant times $\log \log N$.

1.3 Discussion and related work

Theorem 1.2 has a counterpart for preferential attachment models (PAM) proved in [20]. In these PAM's, at each integer time t , a new vertex with $m \geq 1$ edges attached to it, is added to the graph. The new edges added at time t are then preferentially connected to older edges, i.e., conditionally on the graph at time $t - 1$, which is denoted by $G(t - 1)$, the probability that a given edge is connected to vertex i is proportional to $d_i(t - 1) + \delta$, where $\delta > -m$ is a fixed parameter and $d_i(t - 1)$ is the degree of vertex i at time $t - 1$. A substantial literature exists, see e.g. [7, 11], proving that the degree sequence of PAM's in rather great generality satisfy a power law (see e.g. the references in [12]). In the above setting of *linear preferential attachment*, the exponent τ is equal to [22, 12]

$$\tau = 3 + \frac{\delta}{m}. \quad (1.7)$$

A $\log \log t$ upper bound on the diameter holds for PAM's with $m \geq 2$ and $-m < \delta < 0$, which, by (1.7), corresponds to $\tau \in (2, 3)$ [20]:

Theorem 1.3 (A $\log \log$ upper bound on the diameter of the PAM) *Fix $m \geq 2$ and $\delta \in (-m, 0)$. Then, for every $\sigma > \frac{1}{3-\tau}$, and with*

$$C_G(\sigma) = \frac{4}{|\log(\tau - 2)|} + \frac{4\sigma}{\log m},$$

the diameter of the preferential attachment model is, with high probability, bounded above by $C_G \log \log t$, as $t \rightarrow \infty$.

Observe that the condition $m \geq 2$ in the PAM corresponds to the condition $\mathbb{P}(D \geq m + 1) = 1$ in the CM, where one half-edge is used to attach the vertex, while in PAM's, vertices along a path have degree at least three when $m \geq 2$. Also note from the definition of C_G and C_F that distances in PAM's tend to be twice as big compared to distances in the CM. This is related to the structure of the graphs. Indeed, in both graphs, vertices of high degree play a crucial role in shortest paths. In the CM vertices of high degree are often *directly* linked to each other, while in the PAM, they tend to be connected through a third (later) vertex which links to both vertices of high degree.

Unfortunately, there is no $\log t$ lower bound in the PAM for $\delta > 0$ and $m \geq 2$, or equivalently $\tau > 3$. However, [20] does contain a $(1 - \varepsilon) \log t / \log \log t$ lower bound for the diameter when $m \geq 1$ and $\delta \geq 0$. When $m = 1$, results exist on $\log t$ asymptotics of the diameter, see e.g. [6, 25].

The results in Theorems 1.1–1.3 are consistent with the non-rigorous physics predictions that distances in the PAM and in the CM, for similar degree sequences, behave similarly. It is an interesting problem, for both the CM and PAM, to determine the exact constant $C \geq 0$ such that the diameter of the graph of N vertices divided by $\log \log N$ converges in probability to C . For the CM, the results in [17] imply that $C > 0$, for the PAM, this is not known.

We now turn to related work. Many distance results for the CM are known. For $\tau \in (1, 2)$ distances are bounded [15], for $\tau \in (2, 3)$, they behave as $\log \log N$ [26, 19, 10], whereas for $\tau > 3$ the correct scaling is $\log N$ [18]. Observe that these results induce lower bounds for the diameter of the CM, since the diameter is the supremum of the distance, where the supremum is taken over all pairs of vertices. Similar results for models with conditionally independent edges exist, see e.g. [4, 9, 14, 24]. Thus, for these classes of models, distances are quite well understood. The authors in [17] prove that the diameter of a sparse random graph, with specified degree sequence, has, **whp**, diameter equal to $c \log N(1 + o(1))$, for some constant c . Note that our Theorems 1.1–1.2 imply that $c > 0$ when $f_1 + f_2 > 0$, while $c = 0$ when $f_1 + f_2 = 0$ and (1.5) holds for some $\tau \in (2, 3)$.

There are few results on distances or diameter in PAM's. In [5], it was proved that in the PAM and for $\delta = 0$, for which $\tau = 3$, the diameter of the resulting graph is equal to $\frac{\log t}{\log \log t}(1 + o(1))$. Unfortunately, the matching result for the CM has not been proved, so that this does not allow us to verify whether the models have similar distances.

This paper is organized as follows. In Section 2, we prove the lower bound on the diameter formulated in Theorem 1.1 and in Section 3 we prove the upper bound in Theorem 1.2.

2 A lower bound on the diameter: Proof of Theorem 1.1

We start by proving the claim when $f_2 > 0$. The idea behind the proof is simple. Under the conditions of the theorem, one can, **whp**, find a path $\Gamma(N)$ in the random graph such that this path consists exclusively of vertices with degree 2 and has length at least $2\alpha \log N$. This implies that the diameter is at least $\alpha \log N$, since the above path could be a cycle.

Below we define a procedure which proves the existence of such a path. Consider the process of pairing stubs in the graph. We are free to choose the order in which we pair the free stubs, since

this order is irrelevant for the distribution of the random graph. Hence, we are allowed to start with pairing the stubs of the vertices of degree 2.

Let $N(2)$ be the number of vertices of degree 2 and $S_N(2) = (i_1, \dots, i_{N(2)}) \in \mathbb{N}^{N(2)}$ the collection of these vertices. We will pair the stubs and at the same time define a permutation $\Pi(N) = (i_1^*, \dots, i_{N(2)}^*)$ of $S_N(2)$, and a characteristic $\chi(N) = (\chi_1, \dots, \chi_{N(2)})$ on $\Pi(N)$, where χ_j is either 0 or 1. $\Pi(N)$ and $\chi(N)$ will be defined inductively in such a way that for any vertex $i_k^* \in \Pi(N)$, $\chi_k = 1$, if and only if vertex i_k^* is connected to vertex i_{k+1}^* . Hence, $\chi(N)$ contains a substring of at least $2\alpha \log N$ ones precisely when the random graph contains a path $\Gamma(N)$ of length at least $2\alpha \log N$.

We initialize our inductive definition by $i_1^* = i_1$. The vertex i_1^* has two stubs, we consider the second one and pair it to an arbitrary free stub. If this free stub belongs to another vertex $j \neq i_1^*$ in $S_N(2)$ then we choose $i_2^* = j$ and $\chi_1 = 1$, otherwise we choose $i_2^* = i_2$, and $\chi_1 = 0$. Suppose for some $1 < k \leq N(2)$, the sequences (i_1^*, \dots, i_k^*) and (i_1, \dots, χ_{k-1}) are defined. If $\chi_{k-1} = 1$, then one stub of i_k^* is paired to a stub of i_{k-1}^* , and another stub of i_k^* is free, else, if $\chi_{k-1} = 0$, vertex i_k^* has two free stubs. Thus, for every $k \geq 1$, the vertex i_k^* has at least one free stub. We pair this stub to an arbitrary remaining free stub. If this second stub belongs to vertex $j \in S_N(2) \setminus \{i_1^*, \dots, i_k^*\}$, then we choose $i_{k+1}^* = j$ and $\chi_k = 1$, else we choose i_{k+1}^* as the first stub in $S_N(2) \setminus \{i_1^*, \dots, i_k^*\}$, and $\chi_k = 0$. Hence, we have defined that $\chi_k = 1$ precisely when vertex i_k^* is connected to vertex i_{k+1}^* .

We show that **whp** there exists a substring of ones of length at least $2\alpha \log N$ in the first half of χ_N , i.e., in $\chi_{\frac{1}{2}}(N) = (\chi_{i_1^*}, \dots, \chi_{i_{\lfloor N(2)/2 \rfloor}^*})$. For this purpose, we couple the sequence $\chi_{\frac{1}{2}}(N)$ with a sequence $B_{\frac{1}{2}}(N) = \{\xi_k\}$, where ξ_k are i.i.d. Bernoulli random variables taking value 1 with probability $f_2/(4\mu)$, and such that, **whp**, $\chi_{i_k^*} \geq \xi_k$ for all $k \in \{1, \dots, \lfloor N(2)/2 \rfloor\}$. We write \mathbb{P}_N for the law of the CM *conditionally on the degrees* D_1, \dots, D_N . Then, for any $1 \leq k \leq \lfloor N(2)/2 \rfloor$, the \mathbb{P}_N -probability that $\chi_k = 1$ is at least

$$\frac{2N(2) - C_N(k)}{L_N - C_N(k)}, \quad (2.1)$$

where, as before, $N(2)$ is the total number of vertices with degree 2, and $C_N(k)$ is one plus the total number of paired stubs after $k - 1$ pairings. By definition of $C_N(k)$, for any $k \leq N(2)/2$, we have

$$C_N(k) = 2(k - 1) + 1 \leq N(2). \quad (2.2)$$

Due to the law of large numbers we also have that **whp**

$$N(2) \geq f_2 N/2, \quad L_N \leq 2\mu N. \quad (2.3)$$

Substitution of (2.2) and (2.3) into (2.1) then yields that the right side of (2.1) is at least

$$\frac{N(2)}{L_N} \geq \frac{f_2}{4\mu}.$$

Thus, **whp**, we can stochastically dominate all coordinates of the random sequence $\chi_{\frac{1}{2}}(N)$ with an i.i.d. Bernoulli sequence $B_{\frac{1}{2}}(N)$ of $Nf_2/2$ independent trials with success probability $f_2/(4\mu) > 0$. More precisely, for two vectors x and y , we define $x \succcurlyeq y$ when we have $x_i \geq y_i$ for all coordinates i . Then

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\chi_{\frac{1}{2}}(N) \succcurlyeq B_{\frac{1}{2}}(N)\right) = 1.$$

It is well known (see e.g. [13]) that in i.i.d. Bernoulli sequences the probability of existence of a run of $2\alpha \log N$ ones converges to one whenever

$$2\alpha \log N \leq \varrho \frac{\log(Nf_2/2)}{|\log(f_2/(4\mu))|},$$

for some $0 < \varrho < 1$.

We conclude that **whp** the sequence $B_{\frac{1}{2}}(N)$ contains a substring of $2\alpha \log N$ ones. Since **whp** $\chi_N \succcurlyeq B_{\frac{1}{2}}(N)$, the sequence χ_N also contains, **whp**, the same substring of $2\alpha \log N$ ones, and hence there exists a required path consisting of at least $2\alpha \log N$ vertices with degree 2. Thus, **whp** the diameter is at least $\alpha \log N$, and we have proved Theorem 1.1 in the case that $f_2 > 0$.

We now complete the proof of Theorem 1.1 when $f_2 = 0$ by adapting the above argument. When $f_2 = 0$, and since $f_1 + f_2 > 0$, we must have that $f_1 > 0$. Let $l^* > 2$ be the smallest integer such that $f_{l^*} > 0$. This l^* must exist, since $f_1 < 1$. Denote by $N^*(2)$ the total number of vertices of degree l^* of which its first $l^* - 2$ stubs are connected to a vertex with degree 1. Thus, effectively, after the first $l^* - 2$ stubs have been connected to vertices with degree 1, we are left with a structure which has 2 free stubs. These vertices will replace the $N(2)$ vertices used in the above proof. It is not hard to see that **whp** $N^*(2) \geq f_2^* N/2$ for some $f_2^* > 0$. Then, the argument for $f_2 > 0$ can be repeated, replacing $N(2)$ by $N^*(2)$ and f_2 by f_2^* . In more detail, for any $1 \leq k \leq \lfloor N^*(2)/(2l^*) \rfloor$, the \mathbb{P}_N -probability that $\chi_k = 1$ is at least

$$\frac{2N^*(2) - C_N^*(k)}{L_N - C_N^*(k)}, \quad (2.4)$$

where $C_N^*(k)$ is the total number of paired stubs after $k - 1$ pairings of the free stubs incident to the $N^*(2)$ vertices. By definition of $C_N^*(k)$, for any $k \leq N^*(2)/(2l^*)$, we have

$$C_N(k) = 2l^*(k - 1) + 1 \leq N^*(2). \quad (2.5)$$

Substitution of (2.5), $N^*(2) \geq f_2^* N/2$ and the bound on L_N in (2.3) into (2.4) gives us that the right side of (2.4) is at least

$$\frac{N^*(2)}{L_N} \geq \frac{f_2^*}{4\mu}.$$

Now the proof of Theorem 1.1 in the case where $f_2 = 0$ and $f_1 \in (0, 1)$ can be completed as above. We omit further details. \square

3 A log log upper bound on the diameter for $\tau \in (2, 3)$

In this section, we investigate the diameter of the CM when $\mathbb{P}(D \geq m + 1) = 1$, for some integer $m \geq 2$. We assume (1.5) for some $\tau \in (2, 3)$, and that $\mathbb{E}[D] = \mu < \infty$, so that $L_N \leq 2\mu N$ **whp**. We will show that under these assumptions $C_F \log \log N$ is an upper bound on the diameter of the CM, where C_F is defined in (1.6).

The proof is divided into two key steps. In the first, in Proposition 3.1, we give a bound on the diameter of the *core* of the CM consisting of all vertices with degree at least a certain power of $\log N$. This argument is very close in spirit to the one in [26], the only difference being that we have simplified the argument slightly. After this, in Proposition 3.4, we derive a bound on the distance between vertices with small degree and the core. We note that Proposition 3.1 only relies on the assumption in (1.5), while Proposition 3.4 only relies on the fact that $\mathbb{P}(D \geq m + 1) = 1$, for some $m \geq 2$. The proof of Proposition 3.1 can easily be adapted to a setting where the degrees are *fixed*, by formulating the appropriate assumptions on the number of vertices with degree at least x for a sufficient range of x . This assumption would replace (1.5). Proposition 3.4 can easily be adapted to a setting where there are no vertices of degree smaller than or equal to m . This assumption would replace the assumption $\mathbb{P}(D \geq m + 1) = 1$, for some $m \geq 2$. We refrain from stating these extensions of our results, and start by investigating the core of the CM.

We take $\sigma > \frac{1}{3-\tau}$ and define the *core* Core_N of the CM to be

$$\text{Core}_N = \{i : D_i \geq (\log N)^\sigma\}, \quad (3.1)$$

i.e., the set of vertices with degree at least $(\log N)^\sigma$. Also, for a subset $A \subseteq \{1, \dots, N\}$, we define the diameter of A to be equal to the maximal shortest path distance between any pair of vertices of A . Note, in particular, that if there are pairs of vertices in A that are *not* connected, then the diameter of A is infinite. In the following proposition the diameter of the core is bounded. This proposition appeared earlier as [26, Proposition 3.13]. We included the proof for completeness.

Proposition 3.1 (Diameter of the core) *For every $\sigma > \frac{1}{3-\tau}$, the diameter of Core_N is, **whp**, bounded above by*

$$\frac{2 \log \log N}{|\log(\tau - 2)|} (1 + o(1)). \quad (3.2)$$

Proof. We note that (1.5) implies that **whp** the largest degree $D_{(N)} = \max_{1 \leq i \leq N} D_i$ satisfies

$$D_{(N)} \geq u_1, \quad \text{where} \quad u_1 = N^{\frac{1}{\tau-1}} (\log N)^{-1}, \quad (3.3)$$

because, when $N \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}(D_{(N)} > u_1) &= 1 - \mathbb{P}(D_{(N)} \leq u_1) = 1 - (F(u_1))^N \geq 1 - (1 - cu_1^{1-\tau})^N \\ &= 1 - \left(1 - c \frac{(\log N)^{\tau-1}}{N}\right)^N \sim 1 - \exp(-c(\log N)^{\tau-1}) \rightarrow 1. \end{aligned} \quad (3.4)$$

Define

$$\mathcal{N}^{(1)} = \{i : D_i \geq u_1\}, \quad (3.5)$$

so that, **whp**, $\mathcal{N}^{(1)} \neq \emptyset$. For some constant $C > 0$, which will be specified later, and $k \geq 2$ we define recursively

$$u_k = C \log N (u_{k-1})^{\tau-2}, \quad \text{and} \quad \mathcal{N}^{(k)} = \{i : D_i \geq u_k\}. \quad (3.6)$$

We start by identifying u_k :

Lemma 3.2 (Identification of u_k) *For each $k \in \mathbb{N}$,*

$$u_k = C^{a_k} (\log N)^{b_k} N^{c_k}, \quad (3.7)$$

with

$$c_k = \frac{(\tau - 2)^{k-1}}{\tau - 1}, \quad b_k = \frac{1}{3 - \tau} - \frac{4 - \tau}{3 - \tau} (\tau - 2)^{k-1}, \quad a_k = \frac{1 - (\tau - 2)^{k-1}}{3 - \tau}. \quad (3.8)$$

Proof. The proof is left to the reader. □

The key step in the proof of Proposition 3.1 is the following lemma:

Lemma 3.3 (Connectivity between $\mathcal{N}^{(k-1)}$ and $\mathcal{N}^{(k)}$) *Fix $k \geq 2$, and $C > 4\mu/c$ (see (1.3), and (1.5) respectively). Then, the probability that there exists an $i \in \mathcal{N}^{(k)}$ that is not directly connected to $\mathcal{N}^{(k-1)}$ is $o(N^{-\gamma})$, for some $\gamma > 0$ independent of k .*

Proof. We note that, by definition,

$$\sum_{i \in \mathcal{N}^{(k-1)}} D_i \geq u_{k-1} |\mathcal{N}^{(k-1)}|. \quad (3.9)$$

Also,

$$|\mathcal{N}^{(k-1)}| \sim \text{Bin}(N, 1 - F(u_{k-1})), \quad (3.10)$$

and we have that, by (1.5),

$$N[1 - F(u_{k-1})] \geq cN(u_{k-1})^{1-\tau}, \quad (3.11)$$

which, by Lemma 3.2, grows as a positive power of N , since $c_k \leq c_2 = \frac{\tau-2}{\tau-1} < \frac{1}{\tau-1}$. We use a concentration of probability result

$$\mathbb{P}(|X - \mathbb{E}[X]| > t) \leq 2e^{-\frac{t^2}{2(\mathbb{E}[X]+t/3)}}, \quad (3.12)$$

which holds for binomial random variables [21], and gives that the probability that $|\mathcal{N}^{(k-1)}|$ is bounded below by $N[1 - F(u_{k-1})]/2$ is exponentially small in N . As a result, we obtain that for every k , and **whp**

$$\sum_{i \in \mathcal{N}^{(k)}} D_i \geq \frac{c}{2} N(u_k)^{2-\tau}. \quad (3.13)$$

We note (see e.g., [19, (4.34)]) that for any two sets of vertices A, B , we have that

$$\mathbb{P}_N(A \text{ not directly connected to } B) \leq e^{-\frac{D_A D_B}{L_N}}, \quad (3.14)$$

where, for any $A \subseteq \{1, \dots, N\}$, we write

$$D_A = \sum_{i \in A} D_i. \quad (3.15)$$

On the event where $|\mathcal{N}^{(k-1)}| \geq N[1 - F(u_{k-1})]/2$ and where $L_N \leq 2\mu N$, we then obtain by (3.14), and Boole's inequality that the \mathbb{P}_N -probability that there exists an $i \in \mathcal{N}^{(k)}$ such that i is not directly connected to $\mathcal{N}^{(k-1)}$ is bounded by

$$N \exp \left\{ -\frac{u_k N u_{k-1} [1 - F(u_{k-1})]}{2L_N} \right\} \leq N \exp \left\{ -\frac{c u_k (u_{k-1})^{2-\tau}}{4\mu} \right\} = N^{1 - \frac{cC}{4\mu}}, \quad (3.16)$$

where we have used (3.6). Taking $C > 4\mu/c$ proves the claim. \square

We now complete the proof of Proposition 3.1. Fix

$$k^* = \left\lceil \frac{\log \log N}{|\log(\tau - 2)|} \right\rceil. \quad (3.17)$$

As a result of Lemma 3.3 and since $k^* N^{-\gamma} = o(1)$, we have that every vertex in $\mathcal{N}^{(k-1)}$ is directly connected to a vertex in $\mathcal{N}^{(k)}$ for *all* $k \leq k^*$. Further, all vertices in $\mathcal{N}^{(1)}$ are directly connected, since $u_1^2/N = N^{\frac{2}{\tau-1}-1} \log N \rightarrow \infty$ as a positive power of N for $\tau \in (2, 3)$ (compare the exponent in (3.14)). We conclude that, **whp**, the diameter of $\mathcal{N}^{(k^*)}$ is at most $2k^* + 1$.

Therefore, we are done when we can show that

$$\text{Core}_N \subseteq \mathcal{N}^{(k^*)}. \quad (3.18)$$

For this, we note that

$$\mathcal{N}^{(k^*)} = \{i : D_i \geq u_{k^*}\}, \quad (3.19)$$

so that it suffices to prove that $u_{k^*} \geq (\log N)^\sigma$, for any $\sigma > \frac{1}{3-\tau}$. According to Lemma 3.2,

$$u_{k^*} = C^{a_{k^*}} (\log N)^{b_{k^*}} N^{c_{k^*}}. \quad (3.20)$$

It is not hard to see that $N^{c_{k^*}} = O(1)$, $(\log N)^{b_{k^*}} = (\log N)^{\frac{1}{3-\tau} + o(1)}$, and $C^{a_{k^*}} = O(1)$, so that

$$u_{k^*} = (\log N)^{\frac{1}{3-\tau} + o(1)}, \quad (3.21)$$

so that, by picking N sufficiently large, we can make $\frac{1}{3-\tau} + o(1) \leq \sigma$. This completes the proof of Proposition 3.1. \square

For an integer $m \geq 2$, we define

$$C(m) = \sigma / \log m. \quad (3.22)$$

Proposition 3.4 (Maximal distance between periphery and core) *Assume that $\mathbb{P}(D \geq m+1) = 1$, for some $m \geq 2$. Then, for every $\sigma > (3 - \tau)^{-1}$ the maximal distance between any vertex and the core is, **whp**, bounded from above by $C(m) \log \log N$.*

Proof. We start from a vertex i and will show that the probability that the distance between i and Core_N is at least $C(m) \log \log N$ is $o(N^{-1})$. This proves the claim. For this, we explore the neighborhood of i as follows. From i , we connect the first $m + 1$ stubs (ignoring the other ones). Then, successively, we connect the first m stubs from the closest vertex to i that we have connected to and have not yet been explored. We call the arising process when we have explored up to distance k from the initial vertex i the k -exploration tree.

When we never connect two stubs between vertices we have connected to, then the number of vertices we can reach in k steps is *precisely* equal to $(m + 1)m^{k-1}$. We call an event where a stub on the k -exploration tree connects to a stub incident to a vertex in the k -exploration tree a *collision*. The number of collisions in the k -exploration tree is the number of cycles or self-loops in it. When k increases, the probability of a collision increases. However, for k of order $\log \log N$, the probability that more than *two* collisions occur in the k -exploration tree is small, as we will prove now:

Lemma 3.5 (Not more than one collision) *Take $k = \lceil C(m) \log \log N \rceil$. Then, the \mathbb{P}_N -probability that there exists a vertex of which the k -exploration tree has at least two collisions, before hitting the core Core_N , is bounded by $(\log N)^d L_N^{-2}$, for $d = 4C(m) \log(m + 1) + 2\sigma$.*

Proof. For any stub in the k -exploration tree, the probability that it will create a collision before hitting the core is bounded above by $(m + 1)m^{k-1}(\log N)^\sigma L_N^{-1}$. The probability that two stubs will both create a collision is, by similar arguments, bounded above by $[(m + 1)m^{k-1}(\log N)^\sigma L_N^{-1}]^2$. The total number of possible pairs of stubs in the k -exploration tree is bounded by

$$\left((m + 1)(1 + m + \dots + m^{k-1}) \right)^2 \leq \left((m + 1)m^k \right)^2,$$

so that, by Boole's inequality, the probability that the k -exploration tree has at least two collisions is bounded by

$$\left((m + 1)m^k \right)^4 (\log N)^{2\sigma} L_N^{-2}. \quad (3.23)$$

When $k = \lceil C(m) \log \log N \rceil$, we have that $\left((m + 1)m^k \right)^4 (\log N)^{2\sigma} \leq (\log N)^d$, where d is defined in the statement of the lemma. \square

Finally, we show that, for $k = \lceil C(m) \log \log N \rceil$, the k -exploration tree will, **whp** connect to the Core_N :

Lemma 3.6 (Connecting exploration tree to core) *Take $k = \lceil C(m) \log \log N \rceil$. Then, the probability that there exists an i such that the distance of i to the core is at least k is $o(N^{-1})$.*

Proof. Since $\mu < \infty$ we have that $L_N/N \sim \mu$. Then, by Lemma 3.5, the probability that there exists a vertex for which the k -exploration tree has at least 2 collisions before hitting the core is $o(N^{-1})$. When the k -exploration tree from a vertex i does *not* have two collisions, then there are at least $(m - 1)m^{k-1}$ stubs in the k^{th} layer that have not yet been connected. When $k = \lceil C(m) \log \log N \rceil$ this number is at least equal to $(\log N)^{C(m) \log m + o(1)}$. Furthermore, the expected number of stubs incident to the Core_N is at least $N(\log N)^\sigma \mathbb{P}(D_1 \geq (\log N)^\sigma)$ so that **whp** the number of stubs incident to Core_N is at least (compare (1.5))

$$\frac{1}{2} N (\log N)^\sigma \mathbb{P}(D_1 \geq (\log N)^\sigma) \geq \frac{c}{2} N (\log N)^{\frac{2-\tau}{3-\tau}}. \quad (3.24)$$

By (3.14), the probability that we connect none of the stubs in the k^{th} layer of the k -exploration tree to one of the stubs incident to Core_N is bounded by

$$\exp \left\{ -\frac{cN(\log N)^{\frac{2-\tau}{3-\tau}+C(m)\log m}}{2L_N} \right\} \leq \exp \left\{ -\frac{c}{4\mu}(\log N)^{\frac{2-\tau}{3-\tau}+\sigma} \right\} = o(N^{-1}), \quad (3.25)$$

because **whp** $L_N/N \leq 2\mu$, and since $\frac{2-\tau}{3-\tau} + \sigma > 1$. □

Proof of Theorem 1.2. Take any two vertices, say A and B . Then Proposition 3.4 shows that, **whp**, the distance from node A to the core is bounded by $C(m)\log \log N$, and similarly for node B . According to Proposition 3.1, any two nodes within the core are within mutual distance of size $2k^*(1+o(1))$, where k^* is given in (3.2). This shows that the diameter of the graph is, **whp**, bounded above by

$$2C(m)\log \log N + \frac{2\log \log N}{|\log(\tau-2)|} = C_F \log \log N,$$

where the additional contribution of $o(k^*)$ is compensated by the fact that $\sigma > (3-\tau)^{-1}$. □

Acknowledgements. The work of RvdH and DZ was supported in part by Netherlands Organisation for Scientific Research (NWO).

References

- [1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47-97, (2002).
- [2] A.-L. Barabási. *Linked, The New Science of Networks*. Perseus Publishing, Cambridge, Massachusetts, (2002).
- [3] B. Bollobás. *Random Graphs*, 2nd edition, Academic Press, (2001).
- [4] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms* **31**, 3-122, (2007).
- [5] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, **24**(1):5-34, (2004).
- [6] B. Bollobás and O. Riordan. Shortest paths and load scaling in scale-free trees. *Phys. Rev. E.*, **69**:036114, (2004).
- [7] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**(3):279-290, (2001).
- [8] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, **124**(6):1377-1397, (2006).
- [9] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**(25):15879-15882 (electronic), (2002).
- [10] R. Cohen and S. Havlin. Scale free networks are ultrasmall, *Physical Review Letters* **90**, 058701, (2003).
- [11] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures Algorithms*, **22**(3):311-335, (2003).

- [12] M. Deijfen, H. van den Esker, R. van der Hofstad and G. Hooghiemstra. A preferential attachment model with random initial degrees. Preprint (2007). To appear in *Arkiv för Matematik*.
- [13] P. Erdős and A. Rényi. On a new law of large numbers, *J. Analyse Math.* **23**, 103–111, (1970).
- [14] H. van den Esker, R. van der Hofstad and G. Hooghiemstra. Universality for the distance in finite variance random graphs. Preprint (2006). Available from arXiv:math/0605414.
- [15] H. van den Esker, R. van der Hofstad, G. Hooghiemstra and D. Znamenski. Distances in random graphs with infinite mean degrees, *Extremes* **8**, 111-141, 2006.
- [16] C. Faloutsos, P. Faloutsos and M. Faloutsos. On power-law relationships of the internet topology, *Computer Communications Rev.*, **29**, 251-262, (1999).
- [17] D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. Preprint (2004). Available from <http://www.cs.utexas.edu/~fernholz/>
- [18] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures and Algorithms* **26**, 76-123, 2005.
- [19] R. van der Hofstad, G. Hooghiemstra and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees, *Electronic Journal of Probability* **12**, 703-766, 2007.
- [20] R. van der Hofstad and G. Hooghiemstra. Diameters in preferential attachment models. Preprint (2007), available from arXiv:0705.4153v1[math.PR].
- [21] S. Janson. On concentration of probability, *Contemporary Combinatorics*, ed. B. Bollobás, Bolyai Soc. Math. Stud. **10**, János Bolyai Mathematical Society, Budapest, 289-301, (2002).
- [22] J. Jordan. The degree sequences and spectra of scale-free random graphs, *Random Structures Algorithms* **29**(2), 226-242, (2006).
- [23] M.E.J. Newman. The structure and function of complex networks, *SIAM Rev.* **45**(2), 167–256, (2003).
- [24] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, **38**(1):59–75, (2006).
- [25] B. Pittel. Note on the heights of random recursive trees and random m -ary search trees. *Random Structures Algorithms*, **5**(2):337–347, (1994).
- [26] H. Reittu and I. Norros. On the power law random graph model of massive data networks, *Performance Evaluation*, **55** (1-2), 3-23, (2004).
- [27] S. H. Strogatz. Exploring complex networks, *Nature*, **410**(8), 268–276, March (2001).
- [28] D. J. Watts. *Small Worlds, The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, New Jersey, (1999).