The Diameter of Sparse Random Graphs*

Daniel Fernholz
Department of Computer Sciences
University of Texas at Austin
fernholz@cs.utexas.edu

Vijaya Ramachandran
Department of Computer Sciences
University of Texas at Austin
vlr@cs.utexas.edu

December 13, 2004

Abstract

We derive an expression of the form $c \ln n \pm o(\ln n)$ for the diameter of a sparse random graph with a specified degree sequence. The result holds a.a.s., assuming certain convergence and supercriticality conditions are met. The proof is constructive and yields a method for computing the constant c. For the random graph $\mathcal{G}_{n,p}$ with $np = \Theta(1) + 1$, we solve for c in closed form.

Keywords: random graph, diameter, degree sequence, configuration model.

1 Introduction

The *diameter* of a random graph is the maximum distance between two connected vertices. For the diameter of a sparse random graph with a specified degree sequence we derive an expression of the form

$$c \ln n \pm o(\ln n). \tag{1}$$

Our result holds a.a.s. for degree sequences satisfying certain natural convergence conditions. We determine the value of the constant c for all degree distributions which are super-critical, meaning that a giant connected component is a.a.s. present, and for which the average degree is constant. This includes $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$ when the expected degree is a constant and power-law graphs. For sparse random graphs diameter results of such precision were known earlier only for regular graphs of constant degree [4]. Weaker results (to within a constant factor) were known for diameter of sparse $\mathcal{G}_{n,p}$ [5] and random 'expected-degree' power law graphs [10].

Our results have several applications. We make it possible derive the explicit value of the constant c in equation (1) for many natural random graphs. For the random graph $\mathcal{G}_{n,p}$ with $d = n \cdot p > 1$, we derive a closed-form expression for c as a function of d. This function can be characterized asymptotically by $c = \frac{3-o(1)}{\ln d}$ as $d \to 1$ and $c = \frac{1}{\ln d} + \frac{2}{d} + O\left(\frac{\ln d}{d^2}\right)$ as $d \to \infty$; a bound of $\frac{1}{\ln d} \cdot (1 + o(1)) \cdot \ln n$ for the diameter of $\mathcal{G}_{n,p}$ in this range was conjectured in [5]. We also show that, for $d > 1$, the diameter of $\mathcal{G}_{n,p}$ is a.a.s. equal to the diameter of the giant component, as conjectured in [5]. Our results show that almost all pairs of vertices in the giant component have the same shortest path length to within an arbitrarily small constant factor, and also that simple path finding algorithms will find a shortest path between a given pair of vertices a.a.s. in $O(\sqrt{n} \log n)$ time.

*This work was supported in part by NSF grant CCR-9988160.

Among the tools and techniques we use and develop are the following. We give a formulation of asymptotic degree sequences via metrics and convergence of probability distributions. We use couplings to apply well-known results for sums of i.i.d. random variables in situations where a clean i.i.d. condition does not hold; in particular, we achieve bounds on large deviations for neighborhood sizes of a breadth-first search in a random graph. Our actual result on diameter uses our earlier results [7, 8] on κ-core (especially 2-core) to develop a method of identifying and bounding the worst-case distance between two connected vertices. For instance, in a random graph that has a constant fraction of vertices of degree 1, such a path will consist of essentially a longest path (from each of two vertices of degree 1) to just past the 2-core, followed by a typical path across the 2-core. We also make extensive use of probability generating functions (p.g.f.); using the p.g.f., we relate key properties of the degree distribution of the 2-core of a random graph to the degree distribution with the giant component removed.

Organization of this paper. In section 2 we present background material on random variables and random graphs. In section 3 we describe the CM (or *configuration model*) algorithm for generating a random graph with a fixed degree sequence as well the CM κ-core algorithm from [7, 8] (other results are known for the κ-core problem [14, 6], but our results depend mainly on the treatment in [8]). In this section, we also introduce an important tool in our analysis, the *CM breadth-first search algorithm*, and analyze its key properties in terms of neighborhood expansion. In section 4 we state and prove our diameter result. In section 5 we briefly describe some applications of our main result.

2 Preliminaries.

2.1 Random Variables and Probability.

In this section we present some basic probabilistic definitions and notation. We adopt the convention that random variables and other random structures are typeset in boldface, and we write probabilities and expectations using the script symbols $\mathcal{P}$ and $\mathcal{E}$, respectively.

We shall be dealing exclusively with discrete probabilities, and as such nothing in this paper requires more than basic familiarity with elementary probabilistic concepts such as random variables, probability distributions, etc. Specifically, for any random element $\mathbf{a}$ drawn from a set A, we assume that there exists a subset $A' \subseteq A$ such that $\mathcal{P}[\mathbf{a} = a'] > 0$ for all $a' \in A'$, and $\mathcal{P}[\mathbf{a} \in A'] = 1$. Hence, we can safely discard all sets of measure zero, and use the terms "always" and "with probability 1" interchangeably.

2.1.1 Random variables, moments, and generating functions

Most of the random variables we discuss will be defined on a particular discrete set, namely, the set $\mathbb{Z}^* = \{0, 1, 2, \ldots\}$ of non-negative integers. In particular, we have the following assumption.

Assumption 2.1 *Any probability distribution discussed in this paper is assumed to be a distribution on the set $\mathbb{Z}^*$ unless otherwise specified.*

We let Φ denote the set of all discrete probability distributions on $\mathbb{Z}^*$, and we refer to an element of Φ as simply a *distribution.*

Given a $\mathbb{Z}^*$-valued random variable $\mathbf{X}$, the *distribution* of $\mathbf{X}$ is defined by

$$\mathfrak{D}_{\mathbf{X}}(i) = \mathcal{P}[\mathbf{X} = i] \tag{2}$$

for all $i \in \mathbb{Z}^*$. Conversely, for a distribution μ, we let $\mathbf{X}_\mu$ denote a random variable with distribution $\mathfrak{D}_{\mathbf{X}_\mu} = \mu$. For $\mathbb{Z}^*$-valued random variables $\mathbf{X}, \mathbf{Y}$, we write

$$\mathbf{X} \stackrel{d}{=} \mathbf{Y} \tag{3}$$

if $\mathfrak{D}_{\mathbf{X}} = \mathfrak{D}_{\mathbf{Y}}$.

Next, we introduce two standard tools used to describe a random variable $\mathbf{X}_\mu$ [9].

Definition 2.1 *The* k'th factorial moment *of* $\mathbf{X}_\mu$ *is*

$$M_k(\mu) = \mathcal{E}[(\mathbf{X}_\mu)_k] = \sum_{i=0}^{\infty} (i)_k \cdot \mu(i), \tag{4}$$

where $(i)_k = i(i-1)\cdots(i-k+1)$.

Definition 2.2 *The* probability generating function (p.g.f.) *of* $\mathbf{X}_\mu$ *is the function*

$$\psi_\mu(z) = \mathcal{E}[z^{\mathbf{X}_\mu}] = \sum_{i=0}^{\infty} z^i \mu(i) \tag{5}$$

for $z \in [0,1]$.

In general, all generic random variables are assumed to be independent unless certain dependencies are made explicit. For any integer $n \geq 0$ and any distribution μ, we let

$$\sum\nolimits^n \mathbf{X}_\mu \tag{6}$$

denote the sum of n independent random variables distributed according to μ. For any $0 \leq p \leq 1$, we let $\mathbf{1}_p$ denote a Bernoulli random variable with $\mathcal{P}[\mathbf{1}_p = 1] = p$ and $\mathcal{P}[\mathbf{1}_p = 0] = 1 - p$.

2.1.2 Conditional and random distributions

Given $\mathbb{Z}^*$-valued random variables $\mathbf{X}, \mathbf{Y}$, for any i with $\mathcal{P}[\mathbf{Y} = i] > 0$, we define the distribution of $\mathbf{X}$ *conditional on* $\mathbf{Y} = i$ by

$$\mathfrak{D}_{[\mathbf{X}|\mathbf{Y}=i]}(j) = \mathcal{P}[\mathbf{X} = j \mid \mathbf{Y} = i] \tag{7}$$

for all $j \in \mathbb{Z}^+$. We note that the conditional distribution $\mathfrak{D}_{[\mathbf{X}|\mathbf{Y}=i]}$ is not well defined if $\mathcal{P}[\mathbf{Y} = i] = 0$. Therefore, any statement which specifies a property of the conditional distribution $\mathfrak{D}_{[\mathbf{X}|\mathbf{Y}=i]}$ "for all i" should be understood as meaning "for all i such that $\mathcal{P}[\mathbf{Y} = (i)] > 0$."

A *random probability distribution on* $\mathbb{Z}^*$ (or simply a *random distribution*) $\boldsymbol{\mu}$ is a random element drawn from the set Φ of discrete probability distributions on $\mathbb{Z}^*$. Random distributions arise when we consider the behavior of a random variable $\mathbf{X}$ conditional on $\mathbf{Y}$ without specificying a particular event of the form $\mathbf{Y} = i$. Specifically, the random distribution

$$\mathfrak{D}_{\mathbf{X}|\mathbf{Y}} = \begin{cases} \mathfrak{D}_{[\mathbf{X}|\mathbf{Y}=0]} & \text{if } \mathbf{Y} = 0 \\ \mathfrak{D}_{[\mathbf{X}|\mathbf{Y}=1]} & \text{if } \mathbf{Y} = 1 \\ \vdots \end{cases} \tag{8}$$

is the the *distribution of* $\mathbf{X}$ *conditional on* $\mathbf{Y}$.

Assumption 2.1 states that all probability distributions explicitly referenced will be defined on the set $\mathbb{Z}^*$ of non-negative integers; the same is true of random distributions. Hence, can offer an alternate characterization of a random distributions. Specifically, we can consider a random distribution $\boldsymbol{\mu}$ to be a set of $\mathbb{R}$-valued random variables $\boldsymbol{\mu}(i)$ for $i = 0, 1, 2, \ldots$, with the restriction that $0 \leq \boldsymbol{\mu}(i) \leq 1$ for all i and $\sum_i \boldsymbol{\mu}(i) = 1$.

In addition to the $\boldsymbol{\mu}(i)$, a random distribution $\boldsymbol{\mu}$ also induces random variables $M_k(\boldsymbol{\mu})$ and $\psi_{\boldsymbol{\mu}}(z)$ for each $z \in [0, 1]$ (definitions 2.1 and 2.2). In general, these random variables will be $\mathbb{R}$-valued rather than $\mathbb{Z}^*$-valued, so the definitions given in section 2.1.1 for $\mathbb{Z}^*$-valued random variables will not be applicable.

2.1.3 Discrete random processes

For any set A, an *A-valued random process* is a sequence $(\mathbf{a}_0, \ldots, \mathbf{a}_n)$ of random elements drawn from A. The set A is the *state space* of the random process. Intuitively, a random process can be considered as a single random element whose value changes over time, so $\mathbf{a}_t$ denotes the state of the process at time t. In situations where the time parameter t is clear from the context, we simply write $\mathbf{a}$ as an abbreviation for $\mathbf{a}_t$.

The *history* of a random process at time t is the subsequence $(\mathbf{a}_0, \ldots, \mathbf{a}_t)$. A *Markov chain* is a random process satisfying

$$\mathcal{P}[\mathbf{a}_{t+1} = a \mid (\mathbf{a}_0, \ldots \mathbf{a}_t) = (a_0, \ldots a_t)] = \mathcal{P}[\mathbf{a}_{t+1} = a \mid \mathbf{a}_t = a_t] \tag{9}$$

for any sequence $(a_0 \ldots a_t)$ and any element a. For any set $A' \subseteq A$, the *hitting time* of the event $\mathbf{a} \in A'$ is the random time defined by

$$\boldsymbol{\tau}_{[\mathbf{a} \in A']} = \inf\{t : \mathbf{a}_t \in A'\}.$$

2.1.4 Dominated distributions and couplings

Definition 2.3 *Let μ, ν be distributions. We say μ* dominates *ν and write $\mu \rhd \nu$ if, for all i,*

$$\mathcal{P}[\mathbf{X}_\mu \geq i] \geq \mathcal{P}[\mathbf{X}_\nu \geq i]. \tag{10}$$

For random variables $\mathbf{X}, \mathbf{Y}$, analogous to our usage of $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ if $\mathfrak{D}_{\mathbf{X}} = \mathfrak{D}_{\mathbf{Y}}$ (as in equation 3), we write $X \stackrel{d}{\rhd} Y$ if $\mathfrak{D}_{\mathbf{X}} \rhd \mathfrak{D}_{\mathbf{Y}}$ and say $\mathbf{X}$ *dominates* $\mathbf{Y}$ *in distribution.*

Definition 2.4 *Let $\mathbf{X}$ and $\mathbf{Y}$ be random variables. A* coupling *of $\mathbf{X}$ and $\mathbf{Y}$ is a pair of random variables $(\mathbf{X}', \mathbf{Y}')$ defined on a common probability space such that $\mathbf{X}' \stackrel{d}{=} \mathbf{X}$ and $\mathbf{Y}' \stackrel{d}{=} \mathbf{Y}$ (as defined in equation (3)).*

Dominated distribution can also be understood in terms of couplings as follows.

Proposition 2.1 *Let $\mathbf{X}, \mathbf{Y}$ be random variables. Then $\mathbf{X} \stackrel{d}{\rhd} \mathbf{Y}$ if and only if there exists a coupling $(\mathbf{X}', \mathbf{Y}')$ of $\mathbf{X}$ and $\mathbf{Y}$ such that $\mathcal{P}[\mathbf{X}' \geq \mathbf{Y}'] = 1$.*

2.2 Random Graphs

In this section, we define the random graph model which will be used throughout the paper. A *graph* is a pair $G = (V, E)$, where V is a set of vertices and E is a set of edges. We shall treat graphs as generic combinatorial structures, and accordingly, we make the following assumption.

Assumption 2.2 *All n-vertex graphs discussed in this paper are drawn on the same canonical vertex set $V_n = \{v_1, \ldots, v_n\}$.*

This assumption will also hold for any other structures introduced in this section which involve a set of vertices.

2.2.1 Configurations and Graphs.

Definition 2.5 *An* endpoint arrangement *(or simply an* arrangement*) is a pair (V, S) where*

- *V is a set of* vertices*.*
- *S is a set of* endpoints*; each endpoint belongs to exactly one vertex.*

As discussed above, we shall work with a canonical n-vertex set V_n. Accordingly, we may simplify our notation by dropping the reference to the vertex set, and simply refer to the set S itself as an endpoint arrangement. When using this abbreviation, we let $n(S)$ denote the number of vertices corresponding to the arrangement S.[1]

Given an endpoint arrangement S:

- $S(v)$ denotes the set of endpoints which belong to vertex $v \in V_n$.
- $v(s)$ denotes the vertex to which the endpoint $s \in S$ belongs.

Definition 2.6 *A* configuration *is a triple (V, S, E) where (V, S) forms an endpoint arrangement and E is a set of* edges *such that each edge $e \in E$ is a pair of endpoints $\{s_1, s_2\}$, and E forms a perfect matching of S.*

Given a configuration and an endpoint $s \in S$, we denote by $E(s)$ the endpoint matched to s.

A graph is naturally associated with a configuration by associating an edge $(v(s_1), v(s_2))$ with each pair of matched endpoints (s_1, s_2). In this context, a graph can be considered an equivalence class of configurations modulo permutation of endpoints assigned to the same vertex.

In general, we shall analyze configurations directly rather than graphs. For expository purposes, though, we prefer the simple and traditional term *graph*. Accordingly, if the context is clear, we may abuse our terminology slightly and refer to a configuration as a "graph."

We adopt the convention that the variable n will denote the number of vertices $|V|$ in an arrangement, and m will denote the number of endpoints $|S|$. Note that this differs from the traditional convention by which m denotes the number of edges, which would be half the number of endpoints.

The *degree* of a vertex v in an endpoint arrangement S is $|S(v)|$, the number of endpoints assigned to v. The degree of a vertex is denoted by $d_S(v)$, or simply $d(v)$, if the set of endpoints S is clear from the context. For an endpoint s, we also abbreviate

$$d_S(s) = d_S(v(s))$$

[1] Note that it is not (necessarily) the case that $n(S)$ can be determined by counting the number of vertices which contain endpoints in S; there may be vertices which contain no endpoints but are still included as part of the endpoint arrangement.

and refer to this value as the *degree* of the endpoint s. The *degree sequence* of an arrangement S is the sequence of degrees $D_S = (d_S(v_1), \ldots, d_S(v_n))$.

2.2.2 Random configurations

In this subsection, we describe the *configuration model* [3] for generating a random graph with a specified degree sequence. For an endpoint arrangement $S = (V, S)$, we let $\mathbf{E}_S$ denote a uniformly random matching of the endpoints in S, and we define an associated random configuration by

$$\mathbf{G}_S = (V, S, \mathbf{E}_S).$$

For a sequence $D = (d_1, \ldots, d_n)$ of non-negative integers, we let S_D denote an endpoint arrangement with degree sequence D, and, with slight abuse of notation, we define

$$\mathbf{G}_D = \mathbf{G}_{S_D} = (V, S_D, \mathbf{E}_{S_D}).$$

We say $\mathbf{G}_D$ is a *random configuration with degree sequence* D. We note that $\mathbf{G}_D$ is only well-defined if the sum of the degrees in D is odd. Accordingly, we define a *degree sequence* to be a sequence D where $\sum_i d_i$ is even.

Since each simple graph with degree sequence D occurs with the same probability under the configuration model, then conditioning on simplicity produces a uniformly random simple graph with degree sequence D. The following fact follows from a result of McKay and Wormald [11]

Fact 2.1 *If the maximum degree of a degree sequence is $o(n^{1/3})$ and the average degree is $\Theta(1)$, then a random configuration produces a simple graph with constant probability.*

If a degree sequence D satisfies the conditions in fact 2.1, then for any graph property A,

$$\mathcal{P}\left[\mathbf{G}_D \text{ satisfies } A \;\middle|\; \mathbf{G}_D \text{ is simple}\right] = O\left(\mathcal{P}\left[\mathbf{G}_D \text{ satisfies } A\right]\right).$$

In general, we shall ignore the simplicity requirement and just study the random configuration $\mathbf{G}_D$, noting that asymptotic results derived using the configuration model are also applicable to random simple graphs if the maximum and average degree requirements is met.

2.3 Asymptotics.

In the previous section, we defined the random graph $\mathbf{G}_D$ for any degree sequence D. We seek to study such random graphs asymptotically. Several authors, including Molloy and Reed [12, 13], and Aiello, Chung, and Lu [1], have accomplished this by creating infinite sequences $D_1, D_2, \ldots$ of degree sequences, and examining limits as in

$$\lim_{n\to\infty} \mathcal{P}[\mathbf{G}_{D_n} \text{ satisfies } A].$$

(Here D_n is a degree sequence on n vertices.) This involves a considerable amount of overhead, since an entire sequence D_n must be specified for each value of n. We specify asymptotic degree sequences similar to the *smooth sequences* defined by Molloy and Reed [12, 13], but we make no explicit references to sequences of degree sequences. Instead, we embed the set of degree sequences in a topological space, and deal with convergence in the topological sense.

We review some notation and terminology involving limits and convergence in topological spaces. Let X^* be a topological space and consider a subset $X \subset X^*$ and an element $x^* \in X^*$. For any

property P defined on the set X, we say P *holds asymptotically* as $x \to x^*$ if there exists a neighborhood N_{x^*} of x^* in X^* such that P holds for all $x \in N_{x^*} \cap X$.

Consider a mapping $x \mapsto y_x$ from X to a topological space Y. For any element $y \in Y$, we say

$$y_x \to y \text{ as } x \to x^*$$

if, for any neighborhood N_y of y in Y, the property that $y_x \in N_y$ holds asymptotically as $x \to x^*$.

Consider mappings $x \mapsto z_x$ and $x \mapsto z'_x$ from X to $\mathbb{R}$, and assume $z_x \geq 0$ and $z'_x > 0$ for $x \to x^*$. We recall the standard "big O" notation:

- $z_x = O(z'_x)$ as $x \to x^*$ if there exists a constant $C > 0$ such that $\frac{z_x}{z'_x} < C$ holds asymptotically as $x \to x^*$;
- $z_x = o(z'_x)$ as $x \to x^*$ if $\frac{z_x}{z'_x} \to 0$ as $x \to x^*$;
- Ω, Θ, and ω are defined accordingly.

Finally, as a convention, we will often write asymptotic statements as follows:

$$\text{Assume } x \to x^*. \text{ Then } y_x \to y.$$

This statement is equivalent to $y_x \to y$ as $x \to x^*$.

2.3.1 Asymptotics and probability

In this subsection we discuss asymptotic statements involving probabilities and random variables.

Definition 2.7 *For each $x \in X$, let H_x denote an event in some probability space.*

1. *We say H_x occurs* asymptotically almost surely (a.a.s.) *as $x \to x^*$ if $\mathcal{P}[H_x] \to 1$ as $x \to x^*$.*
2. *For a mapping $x \to n_x$ from X to $\mathbb{Z}^+$, we say H_x occurs* with exponentially high probability (w.e.h.p.) in n_x *as $x \to x^*$ if $\mathcal{P}[H_x] = 1 - e^{-n_x^{\Omega(1)}}$ as $x \to x^*$.*

Next, we deal with probabilistic convergence.

Definition 2.8 *For each $x \in X$, let $\mathbf{y}_x$ denote a random element in a topological space Y.*

1. *We say $\mathbf{y}_x \to y$ a.a.s. as $x \to x^*$ if, for every neighborhood N_y of y in Y, the event $\mathbf{y}_x \in N_y$ occurs a.a.s. as $x \to x^*$.*
2. *We say $\mathbf{y}_x \to y$ w.e.h.p. in n_x as $x \to x^*$ if, for every neighborhood N_y of y in Y, the event $\mathbf{y}_x \in N_y$ occurs w.e.h.p. in n_x as $x \to x^*$.*

Using definitions 2.7 and 2.8, we can link together a.a.s. and w.e.h.p. statements, as demonstrated in the following lemma for a.a.s.

Lemma 2.2 *Let X^*, Y^* be topological spaces, let $X \subseteq X^*$ and $Y \subseteq Y^*$, and let $x^* \in X^*$ and $y^* \in Y^*$. For each $x \in X$, let $\mathbf{y}_x$ be a random element in Y, and for each $y \in Y$, let H_y be an event in some probability space. Assume that $\mathbf{y}_x \to y^*$ a.a.s. as $x \to x^*$ and that H_y occurs a.a.s. as $y \to y^*$.*

For each $x \in X$, define an event $H_{\mathbf{y}_x}$ such that

$$\mathcal{P}[H_{\mathbf{y}_x} \mid \mathbf{y}_x = y] = \mathcal{P}[H_y].$$

Then $H_{\mathbf{y}_x}$ occurs a.a.s. as $x \to x^*$.

Proof. For any $\epsilon > 0$, there exists a neighborhood N_{y^*} of y^* in Y^* such that if $y \in N_{y^*} \cap Y$ then $\mathcal{P}[H_y] > 1 - \epsilon$, and there exists a neighborhood N_{x^*} of x^* in X^* such that if $x \in N_{x^*} \cap X$ then $\mathcal{P}[\mathbf{y}_x \in N_{y^*}] > 1 - \epsilon$. It follows that

$$\mathcal{P}[H_{\mathbf{y}_x}] > 1 - 2\epsilon$$

for $x \in N_{x^*}$, and since ϵ is arbitrary, the proof is complete. ∎

We note that this lemma can be generalized in several ways to deal with w.e.h.p. convergence and various other conditions.

2.3.2 Degree distributions

The topologies we use to asymptotically specify degree sequences will be defined implicitly via mappings from the set of degree sequences to various topological parameter spaces. The main parameter will be the degree distribution, defined as follows.

Definition 2.9 *For any degree sequence $D = (d_1, \ldots, d_n)$, the* degree distribution λ_D *is defined by*

$$\lambda_D(i) = \frac{|\{j : d_j = i\}|}{n}. \tag{11}$$

The degree distribution can be understood intuitively as follows. Given an arrangement S with degree sequence D, if we choose a vertex $\mathbf{v}$ uniformly at random from V_n, we have

$$\mathfrak{D}_{d_S(\mathbf{v})} = \lambda_D.$$

Note that, from definition 2.1, $M_1(\lambda_D)$ gives the expected degree of a vertex chosen uniformly at random (i.e. the average degree).

For any distribution λ with $0 < M_1(\lambda) < \infty$, we define an associated *residual distribution* by

$$\mu_\lambda(i) = \frac{(i+1)\lambda(i+1)}{M_1(\lambda)} \tag{12}$$

for all i. If $M_1(\lambda) = 0$ we define $\mu_\lambda(0) = 1$ and $\mu_\lambda(i) = 0$ for $i > 1$. For any degree sequence D, we abbreviate $\mu_D = \mu_{\lambda_D}$ and refer to μ_D as the *residual degree distribution* of D. Since $M_1(\lambda_D) < \infty$ for any degree sequence D (or any sequence of integers), μ_D is well defined. Intuitively, if we choose an endpoint $\mathbf{s}$ uniformly at random from S, we have

$$\mathfrak{D}_{d_S(\mathbf{s})-1} = \mu_D.$$

The term *residual* reflects the fact that $d_S(\mathbf{s}) - 1$ counts the number of endpoints which belong to $v(\mathbf{s})$ other than $\mathbf{s}$ itself.

Our asymptotic specifications will rely mainly on the degree distribution λ_D of a degree sequence D. However, the residual distribution μ_D is often more useful than λ_D in analyzing the CM algorithm. Therefore, we shall be mindful of how our definitions affect the behavior of the residual distribution as well as the degree distribution.

We now define a family of metrics on the set Φ of probability distributions on $\mathbb{Z}^*$.

Definition 2.10 *The* variation distance *between distributions $\mu, \nu \in \Phi$ is given by*

$$g_\Phi(\mu, \nu) = \frac{1}{2} \sum_{i=0}^{\infty} |\mu(i) - \nu(i)|. \tag{13}$$

It is clear that the variation distance defines a metric on Φ. The variation distance is often defined equivalently by

$$g_\Phi(\mu,\nu) = \sum_{i=0}^{\infty} \max(\mu(i) - \nu(i), 0) = \sum_{i=0}^{\infty} \max(\nu(i) - \mu(i), 0). \tag{14}$$

This alternate definition accounts for the topologically irrelevant factor of $\frac{1}{2}$ in equation (13).

For any $k \geq 1$, we define a metric which takes into account the k'th factorial moment of a probability distribution. For distributions μ, ν satisfying $M_k(\mu), M_k(\nu) < \infty$, we define

$$g_{\Phi,k}(\mu,\nu) = g_\Phi(\mu,\nu) + |M_k(\mu) - M_k(\nu)|\,. \tag{15}$$

We refer to convergence of a probability distribution with respect to the metric $g_{\Phi,k}$ as *k-convergence*, and we use the notation $\lambda \xrightarrow{k} \lambda^*$ to indicate k-convergence of λ to λ^*. Accordingly, the notation $\lambda \to \lambda^*$ indicates convergence with respect to the variation distance metric g_Φ.

Note that if $M_k(\lambda) < \infty$, the residual distribution μ_λ given in equation (12) is well-defined. The residual distribution is typically a more useful tool than the degree distribution in studying random graphs, and we deal with the residual distribution much more than the degree distribution itself. We note that the moments of the residual distribution are given by

$$M_k(\mu_\lambda) = \frac{M_{k+1}(\lambda)}{M_1(\lambda)},$$

and therefore $\lambda \xrightarrow{k} \lambda^*$ implies $\mu_\lambda \xrightarrow{k-1} \mu_{\lambda^*}$.

For the rest of this paper, we shall only consider limiting degree distributions with the property that $0 < M_1(\lambda) < \infty$. Hence, a degree sequence D satisfying $\lambda_D \xrightarrow{1} \lambda$, will have average degree $M_1(\lambda_D) = \Theta(1)$. In particular, this implies that the number of endpoints m satisfies $m = \Theta(n)$, where n is the number of vertices. We note that it may be the case that the residual distribution satisfies $M_1(\mu_\lambda) = \infty$, even if $M_1(\lambda) < \infty$. This occurs, for example, for a power law or Pareto distribution $\lambda(i) = \frac{i^{-\beta}}{\zeta(\beta)}$ for values of $2 < \beta \leq 3$ (here $\zeta(\beta)$ denotes the Riemann zeta function). In our study of the diameter of the random graph $\mathbf{G}_D$, we shall often need to consider the two cases $M_1(\mu_\lambda) < \infty$ and $M_1(\mu_\lambda) = \infty$ separately.

2.4 Properties of the Metric Spaces $(\Phi, g_{\Phi,k})$.

In this section, we give some useful results about the metrics we have defined on the set Φ of distributions on $\mathbb{Z}^*$.

For any distribution μ in $\mathbb{Z}^*$, we define the *complementary distribution function* of μ by

$$F_\mu(i) = \mathcal{P}[\mathbf{X}_\mu \geq i] = \sum_{j \geq i} \mu(j).$$

Proposition 2.3 *Let μ^* be a distribution. Then the following statements are equivalent:*

1. *$\mu \to \mu^*$ with respect to the variation distance metric g_Φ.*
2. *$\max_i |\mu(i) - \mu^*(i)| \to 0$.*
3. *$\max_i |F_\mu(i) - F_{\mu^*}(i)| \to 0$.*

The complementary distribution function is related to domination (definition 2.3), since $\mu \rhd \nu$ if and only if $F_\mu(i) \geq F_\nu(i)$ for all i.

Lemma 2.4 *Consider a mapping $\mu \mapsto \nu$ from Φ to Φ such that $\nu \lhd \mu$ for all μ. If there exist distributions μ^* and ν^* such that $\nu \to \nu^*$ as $\mu \xrightarrow{k} \mu^*$, then $\nu \xrightarrow{k} \nu^*$ as $\mu \xrightarrow{k} \mu^*$ as well.*

Proof. It suffices to show that $M_k(\nu) \to M_k(\nu^*) < \infty$. Note that, for any distribution λ,

$$M_k(\lambda) = \sum_{i=0}^{\infty} (i)_k \lambda(i) = \sum_{j=1}^{\infty} ((j)_k - (j-1)_k) F_\lambda(j).$$

Since $(j)_k - (j-1)_k \geq 0$ for all $j \geq 1$, the lemma follows from the statement of Lesbegue dominated convergence in [9]. ∎

This yields a useful corollary regarding endpoint arrangements. For any arrangement S, we write λ_S and μ_S to denote the degree distribution λ_{D_S} and the residual distribution μ_{D_S}, respectively.

Corollary 2.5 *Let S be an arrangement such that $\lambda_S \xrightarrow{k} \lambda$. Let $T \subseteq S$ and assume $\lambda_T \to \lambda'$. Then $\lambda_T \xrightarrow{k} \lambda'$.*

Finally, we introduce truncated distributions.

Definition 2.11 *For any distribution μ and any $0 < \epsilon \leq 1$, the* ϵ-truncated distribution $\mu_{[\epsilon]}$ *is specified by the complementary distribution function*

$$F_{\mu_{[\epsilon]}}(i) = \max\{F_\mu(i) - \epsilon, 0\}$$

for $i > 0$, and $F_{\mu_{[\epsilon]}}(i) = 1$ for $i \leq 0$.

Informally, the distribution $\mu_{[\epsilon]}$ is constructed by removing a total amount ϵ from the weights $\mu(i)$ for the highest values of i, and increasing the weight $\mu(0)$ by ϵ.

Proposition 2.6 *The truncated distribution $\mu_{[\epsilon]}$ satisfies the following properties:*

1. *$g_\Phi(\mu_{[\epsilon]}, \mu) \leq \epsilon$.*
2. *For any ν, if $g_\Phi(\mu, \nu) \leq \epsilon$ then $\mu_{[\epsilon]} \lhd \nu$.*

3 The Configuration Model Algorithm.

As defined in section 2.2, a random configuration is generated from an endpoint arrangement (V, S) by choosing a random matching of the set S of endpoints. The *configuration model (CM) algorithm* is a procedure which generates this matching one edge at a time. The CM algorithm is customizable, in the sense that we have some flexibility regarding the order in which the edges of the matching are revealed.

The intuition is as follows. For any endpoint $s \in S$, a uniformly random matching $\mathbf{E}$ of s will choose $\mathbf{E}(s)$ uniformly at random from the set $S - \{s\}$. Also, the remaining endpoints in $S - \{s, \mathbf{E}(s)\}$ will be matched uniformly at random. Hence, a random matching can be constructed by successively choosing an unmatched endpoint and choosing its match uniformly at random.

This section proceeds as follows. In section 3.1, we give a formal description of the CM algorithm as a discrete random process. Since as described above, the CM algorithm can be customized, we then describe two particular variants of the CM algorithm. The first is the κ-core CM algorithm, which we studied in [7, 8]. The second is the CM BFS algorithm, which performs a standard breadth-first-search while generating a random configuration, and which will be used extensively in this paper.

3.1 The CM Process

In this section we introduce the notation which we shall use to describe the discrete random process associated with the execution of the CM algorithm. All endpoint arrangements, configurations, etc. are defined with respect to the canonical vertex set V_n. The CM algorithm generates a random matching of an initial arrangement U containing $m = |U|$ endpoints.

Definition 3.1 *A* CM process *of an endpoint arrangement U is a random permutation $(\mathbf{s}_1, \ldots, \mathbf{s}_m)$ of U such that for any endpoint s and any even time t,*

$$\mathcal{P}[\mathbf{s}_t = s \mid \mathbf{s}_1, \ldots, \mathbf{s}_{t-1}] = \begin{cases} \frac{1}{m-t+1} & \text{if } s \notin \{\mathbf{s}_1, \ldots, \mathbf{s}_{t-1}\} \\ 0 & \text{otherwise.} \end{cases}$$

We associate a random matching $\mathbf{E}$ of the endpoints in U with a CM process by setting $\mathbf{E}(\mathbf{s}_t) = \mathbf{s}_{t+1}$ for all odd values $t = 1, 3, \ldots, m-1$. It is easy to verify that the CM algorithm does in fact produce a random matching of the endpoints.

At time t, we say the endpoints $\mathbf{s}_1, \ldots, \mathbf{s}_t$ are *exposed*, so the set $\mathbf{U}_t$ contains all unexposed endpoints. The set $\mathbf{U}_t$ plays a central role in the CM algorithm. Accordingly, we define:

- The *unexposed degree* a vertex v is $\mathbf{d}_t(v) = d_{\mathbf{U}_t}(v)$.
- The *unexposed degree distribution* is $\boldsymbol{\lambda}_t = \lambda_{\mathbf{U}_t}$.
- The *unexposed residual distribution* is $\boldsymbol{\mu}_t = \mu_{\mathbf{U}_t}$.

We refer to the value $\mathbf{d}_t(s) - 1$ as the *residual unexposed degree* of the endpoint s at time t. Note that the endpoint $\mathbf{s}_t$ becomes exposed at time t, hence

$$\mathbf{d}_t(\mathbf{s}_t) = \mathbf{d}_{t-1}(\mathbf{s}_t) - 1. \tag{16}$$

In other words, the residual unexposed degree of $\mathbf{s}_t$ at time $t-1$ becomes the total unexposed degree of $\mathbf{s}_t$ at time t. Recall that, if t is even, then the endpoint $\mathbf{s}_t$ is chosen uniformly at random from $\mathbf{U}_{t-1}$. This yields the following fundamental property of the residual unexposed degree of an endpoint chosen at an even time step.

Proposition 3.1 *For any even time step t of the CM algorithm,*

$$\mathfrak{D}_{\mathbf{d}_t(\mathbf{s}_t)|\mathbf{s}_1,\ldots,\mathbf{s}_{t-1}} = \boldsymbol{\mu}_{t-1}. \tag{17}$$

Definition 3.1 specifies that the endpoint $\mathbf{s}_t$ must be chosen uniformly at random from $\mathbf{U}_{t-1}$ for even values of t. However, the behavior for odd values is unspecified; hence, different strategies for choosing endpoints at odd times can be formulated to study various aspects of the resulting graph. Informally, we define a *specialization* of the CM algorithm to be a method for choosing $\mathbf{s}_t$ for odd values of t.

In this paper, we will define several specializations of the CM algorithm, which we will piece together to study the diameter of the random graph $\mathbf{G}_D$. A *halted CM process* is a CM process along with a random halting time $\boldsymbol{\tau}$. Using halted CM processes, we can define a single specialization of the CM algorithm which uses different strategies for choosing endpoints at different times.

Recall that, as defined in section 2.3.2, the notation $\lambda_U \xrightarrow{k} \lambda$ indicates k-convergence of the degree distribution λ_D to the distribution λ. Typically, we will study CM processes for which the initial arrangement U satisfies $\lambda_U \xrightarrow{k} \lambda$ for either $k = 1$ or $k = 2$. The following proposition states that if we only look at $o(m)$ steps, then the residual distribution will not change significantly during this time.

Proposition 3.2 *Assume $\lambda_U \xrightarrow{k} \lambda$ for some distribution λ and $|U| \to \infty$. Then, for any CM process with initial arrangement U, and any time $t = o(|U|)$,*

$$\boldsymbol{\mu}_t \xrightarrow{k-1} \mu_\lambda \tag{18}$$

holds asymptotically always. In particular, $M_1(\boldsymbol{\mu}_t) \to M_1(\mu_\lambda)$ holds asymptotically always for $k \geq 2$.

3.2 The CM κ-core Algorithm.

The κ-core of a graph is the maximal induced subgraph of minimum degree k. The *CM κ-core algorithm* [7, 8] is a specialization of the CM algorithm which chooses an endpoint $\mathbf{s}_t$ of minimum unexposed degree at each odd time step t. Formally, a CM κ-core process is a CM process satisfying

$$\mathbf{d}_{t-1}(\mathbf{s}_t) = \min\{\mathbf{d}_{t-1}(s) : s \in \mathbf{U}_{t-1}\} \tag{19}$$

always for all odd values $t < m$. We note that since there may be several endpoints of minimum degree, the CM κ-core algorithm can be further specialized if necessary. The κ-core algorithm finds the κ-cores of a random graph by exposing all endpoints which are not part of the κ-core and leaving the κ-core unexposed.

For our purposes, we are only concerned with the 2-core phase of the CM κ-core algorithm. We shall further specialize the 2-core algorithm, by introducing a set of vertices W, called the *protected set*.

Definition 3.2 *Given an initial arrangement U and a protected set W, a* protected 2-core CM process *is a halted CM process satisfying $\mathbf{d}_{t-1}(\mathbf{s}_t) = 1$ and $v(\mathbf{s}_t) \notin W$ for all odd time steps $t < \boldsymbol{\tau}$, and halting at the hitting time $\boldsymbol{\tau}$ of the event*

$$\mathbf{d}(v) \neq 1 \text{ for } v \notin W.$$

Hence, the protected 2-core algorithm chooses endpoints of unexposed degree 1 outside of W and halts when no such endpoints remain.

The state of the CM process at the halting time $\boldsymbol{\tau}$ of the CM 2-core algorithm can be determined w.e.h.p. from the results in [8] as described briefly below. For details, see appendix A. Recall that the probability generating function (p.g.f.) of a distribution μ is defined by $\psi_\mu(z) = \mathcal{E}[z^{\mathbf{X}_\mu}]$ for $z \in [0, 1]$ (see equation 5). Accordingly, for any distribution μ, we let z_μ denote the lowest fixed point in ψ_μ, so

$$z_\mu = \min\{z \in [0, 1] : \psi_\mu(z) = z\}, \tag{20}$$

and we define the distribution ρ_μ by specifying the p.g.f.

$$\psi_{\rho_\mu}(z) = \frac{\psi_\mu((1 - z_\mu)z + z_\mu) - z_\mu}{1 - z_\mu}. \tag{21}$$

The following lemma is an implication of [8] (see Theorem A.2 in appendix A).

Lemma 3.3 *Let U be an initial arrangement satisfying $\lambda_U \xrightarrow{k} \lambda$, where λ has residual distribution $\mu = \mu_\lambda$. Let W be a protected set and assume $|W| = o(n)$. Then*

1. *The stopping time of the protected 2-core algorithm satisfies $\frac{\boldsymbol{\tau}}{m} \to 1 - (1 - z_\mu)^2$ w.e.h.p., where z_μ is defined in (20).*
2. *The residual distribution at time $\boldsymbol{\tau}$ satisfies $\boldsymbol{\mu}_{\boldsymbol{\tau}} \xrightarrow{k-1} \rho_\mu$ w.e.h.p., where ρ_μ is defined in (21)*

We note that the second condition is as strong as possible, since k-convergence of the distribution λ corresponds to $(k-1)$-convergence of the residual distribution μ_λ.

3.3 The CM BFS Algorithm.

Here we consider a breadth-first-search (BFS) specialization of the κ-core algorithm. Beginning with a set of vertices W, the CM BFS algorithm performs a standard breadth-first search and exposes endpoints accordingly. Typically, BFS involves a queue of vertices; our implementation differs slightly in that the queue will contain endpoints rather than vertices. We first define a single BFS iteration.

Definition 3.3 *Let U be an initial arrangement and let W be a vertex set with endpoint set $R = U(W)$. A* CM BFS iteration process *is a halted CM process where $\mathbf{s}_t \in R$ for all odd time steps $t \leq \boldsymbol{\tau}$, and halting at the hitting time $\boldsymbol{\tau}$ of the event $R \cap \mathbf{U} = \emptyset$.*

For the CM BFS iteration process, we expand the state of the BFS algorithm as follows.

- $\mathbf{R} = R \cap \mathbf{U}$ denotes the remaining unexposed endpoints in R.
- $\mathbf{Q} = \mathbf{R} \cup \{s \in \mathbf{U} : \mathbf{d}(s) < d_U(s)\}$. We call $\mathbf{Q}$ the *endpoint queue*; accordingly, the *initial endpoint queue* is the set $\mathbf{Q}_0 = R$. This definition specifies that any endpoint s whose current unexposed degree $\mathbf{d}(s)$ differs from its initial degree $d_U(s)$ belongs to $\mathbf{Q}$, as well as any remaining unexposed endpoints in R.
- $\mathbf{T} = \mathbf{U} - \mathbf{Q}$ contains all unexposed endpoints not in the queue. Endpoints in $\mathbf{T}$ are called *unexplored*. The set $U' = U - R$ is the initial set of unexplored endpoints.

We note that, in the context of BFS, the set $\mathbf{T}$ of unexplored endpoints is typically more relevant than $\mathbf{U}$, since the set $\mathbf{Q}$ contains endpoints which are already connected to a vertex in the set W. Hence we denote the residual distribution with respect to the set of unexplored endpoints $\mathbf{T}$ by

$$\boldsymbol{\nu}_t = \mu_{\mathbf{T}_t}, \tag{22}$$

and refer to $\boldsymbol{\nu}_t$ as the *unexplored distribution* at time t.

Note that, at each odd time step, the endpoint $\mathbf{s}_t$ must be chosen from R and hence from $\mathbf{Q}_{t-1}$, so $|\mathbf{Q}_t| = |\mathbf{Q}_{t-1}| - 1$ always for odd t. However, at each even time step, two qualitative outcomes are possible regarding the matched endpoint $\mathbf{s}_t$. If $\mathbf{s}_t \notin \mathbf{Q}_{t-1}$, then $|\mathbf{Q}_t| \geq |\mathbf{Q}_{t-1}|$, with equality if $d_{\mathbf{U}_t}(\mathbf{s}_t) = 0$. Otherwise, $\mathbf{s}_t \in \mathbf{Q}_{t-1}$ so $|\mathbf{Q}_t| < |\mathbf{Q}_{t-1}|$, and a *cross-edge* has occurred. We also distinguish two different kinds of cross edges.

- A *horizontal edge* occurs if $\mathbf{s}_t \in \mathbf{R}_{t-1}$.
- A *diagonal edge* occurs if $\mathbf{s}_t \in \mathbf{Q}_{t-1} - \mathbf{R}_{t-1}$.

We define the *BFS neighborhood* of a CM BFS iteration to be the set of endpoints $\mathbf{N} = \mathbf{Q}_{\boldsymbol{\tau}}$ in the endpoint queue when the BFS iteration ends.

Definition 3.4 *An* extended CM BFS process *(or just* CM BFS process*) is a sequence of CM BFS iterations, where the BFS neighborhood $\mathbf{N}$ of each iteration is used as the initial endpoint queue R for next iteration.*

For an extended CM BFS process we define:

- The *i'th BFS neighborhood* $\mathbf{N}_i$ is the BFS neighborhood of the i'th CM BFS iteration. For consistency, we let $\mathbf{N}_0 = R$, the initial endpoint queue.

- The *i'th halting time* $\boldsymbol{\tau}_i$ is halting time of the i'th CM BFS iteration.

In the following subsections, we will give probabilistic bounds on the sizes of the BFS neighborhoods. Our strategy will be to show that the random variable $|\mathbf{N}_i|$ is distributed "similarly" to the sum of $|\mathbf{N}_{i-1}|$ independent random variables distributed according to the residual unexplored distribution $\boldsymbol{\nu}$. Specifically, we will show that $|\mathbf{N}_i|$ is dominated in distribution from below by the truncated distribution $\boldsymbol{\nu}_{[\epsilon]}$ for arbitrary $\epsilon > 0$.

Our proof will involve a coupling of the CM BFS process with a random process which does generate a random variable which is distributed identically the sum of independent random variables. First, we derive some results for coupled random processes.

3.3.1 Coupled Random Processes.

We begin by recalling some notation from section 2.

From section 2.4, recall that for a distribution μ, $\mu_{[\epsilon]}$ denotes the ϵ-truncated distribution, which satisfies $g_\Phi(\mu_{[\epsilon]}, \mu) \leq \epsilon$ and $\mu_{[\epsilon]} \lhd \nu$ for all ν with $g_\Phi(\nu, \mu) < \epsilon$. From section 2.1.1, recall that $\sum^n \mathbf{X}_\mu$ denotes the sum of n independent random variables distributed according to μ and $\mathbf{1}_p$ denotes a Bernoulli random variable with parameter p. For a definition of coupled random variables, see section 2.1.4.

Lemma 3.4 *For all $\epsilon > 0$, all $\mu \in \Phi$, and all $r > 0$,*

$$\sum^r \mathbf{X}_{\mu_{[\epsilon]}} \stackrel{d}{\lhd} \sum^{\sum^r \mathbf{1}_{(1-\epsilon)}} \mathbf{X}_\mu.$$

Proof. Write

$$\sum^{\sum^r \mathbf{1}_{(1-\epsilon)}} \mathbf{X}_\mu = \sum_{i=1}^{r} \mathbf{Y}_i,$$

where the $\mathbf{Y}_i$ are independent and distributed identically to $\mathbf{Y} \stackrel{d}{=} \mathbf{1}_{(1-\epsilon)} \cdot \mathbf{X}_\mu$, and note that

$$g_\Phi(\mathfrak{D}_\mathbf{Y}, \mu) \leq \epsilon. \quad \blacksquare$$

Lemma 3.5 *For any values $n > 0$ and $0 \leq p \leq 1$,*

$$2 \cdot \sum^n \mathbf{1}_p \stackrel{d}{\lhd} \sum^{2n} \mathbf{1}_{\sqrt{p}}. \tag{23}$$

Proof. The result follows immediately from the following coupling:

$$\sum^{2n} \mathbf{1}_{\sqrt{p}} \stackrel{d}{=} \sum_{i=1}^{2n} \mathbf{X}_i, \quad \text{where the } \mathbf{X}_i \text{ are independent and } \mathbf{X}_i \stackrel{d}{=} \mathbf{1}_{\sqrt{p}},$$

$$\sum^{n} \mathbf{1}_{p} \stackrel{d}{=} \sum_{j=1}^{n} \mathbf{Y}_j, \quad \text{where } \mathbf{Y}_j = \mathbf{X}_{2j-1} \cdot \mathbf{X}_{2j} \stackrel{d}{=} \mathbf{1}_p.$$

Clearly, for any k, if $\sum \mathbf{Y}_j = k$, then $\sum \mathbf{X}_i \geq 2k$. $\blacksquare$

Next, we give some simple results regarding couplings and random processes. We begin with a lemma about couplings and dependent random variables.

Lemma 3.6 *Let $\mathbf{X}_1, \mathbf{X}_2$ and $\mathbf{Y}_1, \mathbf{Y}_2$ be (not necessarily independent) random variables, and suppose:*

1. $\mathbf{X}_1 \overset{d}{\rhd} \mathbf{Y}_1$, *and*

2. *for every* $i \geq j$,

$$\mathfrak{D}_{[\mathbf{X}_2|\mathbf{X}_1=i]} \rhd \mathfrak{D}_{[\mathbf{Y}_2|\mathbf{Y}_1=j]}. \tag{24}$$

Then $\mathbf{X}_2 \overset{d}{\rhd} \mathbf{Y}_2$.

Proof. Since $\mathbf{X}_1 \overset{d}{\rhd} \mathbf{Y}_1$, we can create a coupling $(\mathbf{X}'_1, \mathbf{Y}'_1)$ such that $\mathcal{P}[\mathbf{X}'_1 \geq \mathbf{Y}'_1] = 1$. Similarly, by the second hypothesis, for any $i \geq j$, we can define a coupling $(\mathbf{X}'_{(i,j)}, \mathbf{Y}'_{(i,j)})$ of the conditional random variables $[\mathbf{X}_2|\mathbf{X}_1 = i]$ and $[\mathbf{Y}_2|\mathbf{Y}_1 = j]$ such that

$$\mathcal{P}[\mathbf{X}'_{(i,j)} \geq \mathbf{Y}'_{(i,j)}] = 1.$$

Finally, we define a pair of random variables $(\mathbf{X}'_2, \mathbf{Y}'_2)$ by

$$\mathbf{X}'_2 = \mathbf{X}'_{(\mathbf{X}'_1, \mathbf{Y}'_1)}$$
$$\mathbf{Y}'_2 = \mathbf{Y}'_{(\mathbf{X}'_1, \mathbf{Y}'_1)}.$$

We easily verify that for any value k,

$$\begin{aligned}
\mathcal{P}[\mathbf{X}'_2 = k] &= \sum_{i,j} \mathcal{P}[\mathbf{X}'_{(i,j)} = k \mid \mathbf{X}'_1 = i, \mathbf{Y}'_1 = j] \cdot \mathcal{P}[\mathbf{X}'_1 = i, \mathbf{Y}'_1 = j] \\
&= \sum_{i} \mathcal{P}[\mathbf{X}_2 = k \mid \mathbf{X}_1 = i] \cdot \mathcal{P}[\mathbf{X}_1 = i] \\
&= \mathcal{P}[\mathbf{X}_2 = k]
\end{aligned}$$

and therefore $\mathbf{X}'_2 \overset{d}{=} \mathbf{X}_2$. Similarly, $\mathbf{Y}'_2 \overset{d}{=} \mathbf{Y}_2$. Also, since $\mathcal{P}[\mathbf{X}'_1 \geq \mathbf{Y}'_1] = 1$, and $\mathcal{P}[\mathbf{X}'_2 \geq \mathbf{Y}'_2 \mid \mathbf{X}'_1 \geq \mathbf{Y}'_1] = 1$, it follows that $\mathcal{P}[\mathbf{X}'_2 \geq \mathbf{Y}'_2] = 1$. ∎

Recall that for any distribution μ, $\sum^n \mathbf{X}_\mu$ denotes the sum of n independent random variables distributed according to μ. The following proposition follows easily by applying lemma 3.6 inductively.

Proposition 3.7 *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be random variables, and let* μ *be a distribution such that for all* i,

$$\mathcal{P}[\mathfrak{D}_{\mathbf{X}_i|\mathbf{X}_1,\ldots,\mathbf{X}_{i-1}} \lhd \mu] = 1. \tag{25}$$

Then $\sum_{i=1}^n \mathbf{X}_i \overset{d}{\lhd} \sum^n \mathbf{X}_\mu$.

The next proposition deals with couplings for random processes. Again, the result follows from a simple inductive application of lemma 3.6.

Proposition 3.8 *Let* $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_r$ *be a* $\mathbb{Z}^*$*-valued random process and let* $\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_r$ *be a* $\mathbb{Z}^*$*-valued Markov chain such that* $\mathbf{X}_0 \overset{d}{\rhd} \mathbf{Y}_0$, *and such that for any* $i \geq j$ *and any* t,

$$\mathcal{P}\left[\mathfrak{D}_{\mathbf{X}_t|\mathbf{X}_1,\ldots,\mathbf{X}_{t-1}} \rhd \mathfrak{D}_{[\mathbf{Y}_t|\mathbf{Y}_{t-1}=j]} \mid \mathbf{X}_{t-1} = i\right] = 1. \tag{26}$$

Then $\mathbf{X}_t \overset{d}{\rhd} \mathbf{Y}_t$ *for all* t.

3.3.2 A single iteration of CM BFS

In this subsection we give a probabilistic lower bound for the size of the neighborhood of a single BFS iteration. We first relate the size of a BFS neighborhood to the sum of i.i.d. random variables. This allows us to use standard large deviation techniques (i.e., Cramer's theorem) to derive concentration results for the BFS neighborhood size.

Theorem 3.9 *Consider a CM BFS iteration for which the following hold.*

1. *The initial arrangement U satisfies $\lambda_U \xrightarrow{1} \lambda$, and λ has residual distribution $\mu = \mu_\lambda$.*
2. *The initial queue $R = U(W)$ has size $r = |R| = o(m)$.*

Then, for any $\epsilon > 0$, the BFS neighborhood $\mathbf{N}$ satisfies

$$|\mathbf{N}| \overset{d}{\triangleright} \sum^{r} \mathbf{X}_{\mu_{[\epsilon]}}. \tag{27}$$

To prove Theorem 3.9, we begin by pointing out the obvious fact that a BFS iteration can last at most $2 \cdot r$ steps, and therefore, for $r = o(m)$, the residual distribution $\boldsymbol{\mu}_t \to \mu$ always for any $0 \le t \le \boldsymbol{\tau} \le 2r$. It is also true, though less obvious, that $\boldsymbol{\nu}_t \to \mu$, and therefore $|\mathbf{Q}_t| = o(m)$, for any $0 \le t \le \boldsymbol{\tau} \le 2r$. This is because, for any $v \notin W$, if $\mathbf{d}_t(v) = d_U(v)$, then $d_{\mathbf{T}_t}(v) = \mathbf{d}_t(v) = d_U(v)$, and the unexposed degree can only change for at most $r = o(m)$ vertices outside of W during the iteration.

Now, we bound the number of endpoints in horizontal edges with the following lemma. Recall that $\mathbf{1}_p$ is a Bernoulli random variable with parameter p.

Lemma 3.10 *Let $\mathbf{A} \subset R$ denote the set of endpoints $s \in R$ such the endpoint $\mathbf{E}(s)$ matched to s satisfies $\mathbf{E}(s) \in R$. Then, for any $\epsilon > 0$,*

$$|\mathbf{A}| \overset{d}{\triangleleft} \sum^{r} \mathbf{1}_\epsilon.$$

Proof. Note that the distribution of $|\mathbf{A}|$ depends only on r and m, since the actual degrees of vertices do not affect the number of horizontal edges. Let $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_r$ denote the endpoints of R in the order that they are exposed by CM BFS. Any horizontal edge must consist of two consecutive endpoints according to this ordering. So, for $1 \le i \le r-1$, define random variables

$$\mathbf{K}_i = \begin{cases} 1 & \text{if } \mathbf{E}(\mathbf{q}_i) = \mathbf{q}_{i+1}, \\ 0 & \text{otherwise,} \end{cases}$$

so $|\mathbf{A}| = 2 \cdot \sum_{i=1}^{r-1} \mathbf{K}_i$.

Note that since $r = o(m)$, the probability that an endpoint chosen uniformly at random from $\mathbf{U}_t$ lies in R is at most $\frac{r}{m-2r} = O(\frac{r}{m}) = o(1)$ at any time during the BFS iteration. Also, if $\mathbf{K}_i = 1$ then $\mathbf{K}_{i+1} = 0$, since the endpoint $\mathbf{q}_{i+1}$ cannot be part of more than one horizontal edge. Hence, for $i > 1$,

$$\mathcal{P}[\mathbf{K}_i = 1 \mid \mathbf{K}_1, \ldots, \mathbf{K}_{i-1}] = \begin{cases} 0 & \text{if } \mathbf{K}_{i-1} = 1, \\ O(\frac{r}{m}) & \text{if } \mathbf{K}_{i-1} = 0. \end{cases} \tag{28}$$

This observation allows us to consider the $\mathbf{K}_i$ in pairs. So, for $1 \le i \le \lfloor r/2 \rfloor$, we define

$$\mathbf{L}_i = \mathbf{K}_{2i-1} + \mathbf{K}_{2i},$$

and note that $|\mathbf{A}| = 2\sum_{i=1}^{\lfloor r/2 \rfloor} \mathbf{L}_i$. Also, $\mathbf{L}_i \leq 1$ always for all i, and

$$\mathcal{P}[\mathbf{L}_i = 1 \mid \mathbf{L}_1, \ldots, \mathbf{L}_{i-1}] = \mathcal{P}[\mathbf{L}_i = 1 \mid \mathbf{L}_{i-1}] \leq 2O(\tfrac{r}{m}) = O(\tfrac{r}{m})$$

always, and since the $\mathbf{L}_i$ are Bernoulli random variables, then

$$\mathcal{P}[\mathfrak{D}_{\mathbf{L}_i|\mathbf{L}_1,\ldots,\mathbf{L}_{i-1}} \lhd \mathbf{1}_{O(\frac{r}{m})}] = 1.$$

Hence, we conclude by proposition 3.7 that

$$|\mathbf{A}| \lhd 2 \cdot \sum^{\lfloor r/2 \rfloor} \mathbf{1}_{O(\frac{r}{m})}.$$

Finally, by lemma 3.5, it follows that that

$$|\mathbf{A}| \overset{d}{\lhd} \sum^{r} \mathbf{1}_{O(\sqrt{\frac{r}{m}})},$$

and therefore

$$|\mathbf{A}| \overset{d}{\lhd} \sum^{r} \mathbf{1}_{\epsilon}$$

for any $\epsilon > 0$. ∎

Now, note that the set of endpoints $R - \mathbf{A}$ will all match to the set $U' = U - R$. In particular, conditional on $|\mathbf{A}|$, exactly $r - |\mathbf{A}|$ endpoints chosen uniformly at random from U' will become exposed during the BFS iteration.

Lemma 3.11 *For any $r' \leq r$, and for any $\epsilon > 0$, the size of the BFS neighborhood $|\mathbf{N}|$ satisfies*

$$|\mathbf{N}| \overset{d}{\rhd} \sum^{r'} \mathbf{X}_{\mu_{[\epsilon]}}$$

conditional on $r - |\mathbf{A}| = r'$.

Proof. As noted above, conditional on $r - |\mathbf{A}| = r'$, exactly r' endpoints chosen uniformly at random from U' become exposed during the BFS iteration. Without loss of generality, we may assume that the endpoints in $\mathbf{A}$ become exposed first, so after $|\mathbf{A}|$ steps, we have $\mathbf{Q}_{|\mathbf{A}|} = \mathbf{R}_{|\mathbf{A}|} = R - \mathbf{A}$, and that the next r' endpoints exposed at even time steps are drawn from U'. Then, for any $0 \leq t' \leq r'$, we define $\mathbf{K}_{t'} = \left|\mathbf{Q}_{|\mathbf{A}|+2t'}\right| - \left|\mathbf{R}_{|\mathbf{A}|+2t'}\right|$, so

$$\mathbf{K}_{r'} = |\mathbf{Q}_{\boldsymbol{\tau}}| = |\mathbf{N}|\,.$$

The random process $\mathbf{K}_0, \ldots, \mathbf{K}_{r'}$ thus keeps track of the number of "new" endpoints in the BFS queue which are not part of the original queue R. Now, the value of $\mathbf{K}$ drops by exactly 1 if a cross-edge occurs. Hence,

$$\mathcal{P}[\mathbf{K}_{t'} = \mathbf{K}_{t'-1} - 1 \mid \mathbf{K}_1, \ldots, \mathbf{K}_{t'-1}] = O\left(\tfrac{\mathbf{K}_{t'-1}}{m}\right). \tag{29}$$

Otherwise, when a cross-edge does not occur, $\mathbf{K}$ will increase by the value of the unexplored degree of the endpoint $\mathbf{s}_{|\mathbf{A}|+2t'}$ chosen at time $|\mathbf{A}| + 2t'$. So, for any $i \geq 0$,

$$\mathcal{P}[\mathbf{K}_{t'} = \mathbf{K}_{t'-1} + i \mid \mathbf{K}_1, \ldots, \mathbf{K}_{t'-1}] = \boldsymbol{\nu}_{|\mathbf{A}|+2t'}(i) - O\left(\tfrac{\mathbf{K}_{t'-1}}{m}\right) = \mu(i) \pm o(1), \tag{30}$$

since $\frac{\mathbf{K}_{t'-1}}{m} = o(1)$ and $t' = o(m)$; here $\boldsymbol{\nu}_{|\mathbf{A}|+2t'}$ denotes the unexplored residual distribution (equation 22) at time $|\mathbf{A}| + 2t'$.

We wish to show that

$$\mathbf{K}_{t'} \overset{d}{\triangleright} \sum^{t'} \mathbf{X}_{\mu_{[\epsilon]}} \tag{31}$$

for arbitrary $\epsilon > 0$. However, since, as described above, the value of $\mathbf{K}_{t'}$ can change in two different ways, it is difficult to relate $\mathbf{K}_{t'}$ to the sum if i.i.d. random variables directly. Instead, we construct an auxilliary $\mathbb{Z}^*$-valued Markov chain $\mathbf{Z}_1, \ldots, \mathbf{Z}_{r'}$ and show that

$$\mathbf{K}_{t'} \overset{d}{\triangleright} \mathbf{Z}_{t'} \overset{d}{\triangleright} \sum^{t'} \mathbf{X}_{\mu_{[\epsilon]}}.$$

The Markov chain $\mathbf{Z}_{t'}$ must be "easy" to compare to both the random process $\mathbf{K}_{t'}$ and the sum of i.i.d. random variables $\sum^{t'} \mathbf{X}_{\mu_{[\epsilon]}}$. Intuitively, it is useful to think of the Markov chain $\mathbf{Z}_{t'}$ as counting the number of elements in some set $\mathbf{H}_{t'}$, so

$$\mathbf{Z}_{t'} = |\mathbf{H}_{t'}|.$$

The contents of the set $\mathbf{H}_{t'}$ change over time in a way that resembles a CM BFS iteration, but exhibits enough independence that the process $\mathbf{Z}_{t'}$ can be easily compared to a sequence of i.i.d. random variables.

Let us define such a process. At time 0, we set $\mathbf{H}_0 = \emptyset$ and therefore $\mathbf{Z}_0 = |\mathbf{H}_0| = 0$. Then, at each time step t', the contents of the set $\mathbf{H}$ change as follows.

1. A random number of new elements is added to $\mathbf{H}_{t'}$; the number of new elements is distributed according to $\mu_{[\epsilon_1]}$ where $\epsilon_1 > 0$ is arbitrary.
2. Each element in $\mathbf{H}_{t'-1}$ is removed independently with probability $\frac{\epsilon_2}{r'}$, where $\epsilon_2 > 0$ is arbitrary.

Formally, then, $\mathbf{Z}_{t'}$ is a Markov chain defined by $\mathbf{Z}_0 = 0$, and, if we let $\mathbf{L}_{t'}$ denote the random variable $\mathbf{Z}_{t'}$ conditional on $\mathbf{Z}_{t'-1}$, then

$$\mathbf{L}_{t'} \overset{d}{=} \mathbf{X}_{\mu_{[\epsilon_1]}} + \sum^{\mathbf{Z}_{t'-1}} \mathbf{1}_{(1-\frac{\epsilon_2}{r'})}.$$

We also give an alternate description of the process $\mathbf{Z}_{t'}$. For $1 \le i \le r'$ and $0 \le j \le r' - i$, let $\mathbf{Y}_{(i,j)}$ denote the number of elements that remain in the set $\mathbf{H}$ at time $i + j$ among those that were added at time i. As described above, we have

$$\mathbf{Y}_{(i,0)} \overset{d}{=} \mathbf{X}_{\mu_{[\epsilon_1]}}.$$

and

$$\mathbf{Y}_{(i,j)} \overset{d}{=} \sum^{\mathbf{Y}_{(i,j-1)}} \mathbf{1}_{(1-\frac{\epsilon_2}{r'})}$$

for $j > 0$. It follows by definition that

$$\mathbf{Z}_{t'} = \sum_{i=1}^{t'} \mathbf{Y}_{(i,t'-i)}.$$

We now show that $\mathbf{Z}_{t'} \overset{d}{\triangleleft} \mathbf{K}_{t'}$. Note that $\mathbf{Z}_{t'} \le \mathbf{Z}_{t'-1} - 1$ occurs if (but not only if) $\mathbf{Y}_{(t',0)} = 0$ and at least one element in $\mathbf{H}_{t'-1}$ is removed at time t'. Hence

$$\begin{aligned}\mathcal{P}[\mathbf{Z}_{t'} \le \mathbf{Z}_{t'-1} - 1 \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{t'-1}] &\ge \epsilon_1 \cdot \left(1 - (1 - \tfrac{\epsilon_2}{r'})^{\mathbf{Z}_{t'-1}}\right) \\ &\ge \epsilon_1 \cdot \left(1 - e^{-\frac{\epsilon_2 \mathbf{Z}_{t'-1}}{r'}}\right).\end{aligned}$$

We claim that

$$\mathcal{P}[\mathbf{Z}_{t'} \leq \mathbf{Z}_{t'-1} - 1 \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{t'-1}] = \omega\left(\frac{\mathbf{Z}_{t'-1}}{m}\right) \tag{32}$$

for $\mathbf{Z}_{t'-1} = o(m)$. For $\mathbf{Z}_{t'-1} = \Omega(r')$, we have

$$\epsilon_1 \cdot \left(1 - e^{-\frac{\epsilon_2 \mathbf{Z}_{t'-1}}{r'}}\right) = \Theta(1),$$

so (32) follows immediately, and for $\frac{\epsilon_2 \mathbf{Z}_{t'-1}}{r'}$ sufficiently small, (32) follows from the power series

$$\epsilon_1 \cdot \left(1 - e^{-\frac{\epsilon_2 \mathbf{Z}_{t'-1}}{r'}}\right) = \frac{\epsilon_1 \epsilon_2 \mathbf{Z}_{t'-1}}{r'} - \frac{\epsilon_1}{2!}\left(\frac{\epsilon_2 \mathbf{Z}_{t'-1}}{r'}\right)^2 + \cdots.$$

Now, for any $i \geq 0$, we have

$$\mathcal{P}[\mathbf{Z}_{t'} \geq \mathbf{Z}_{t'-1} + i \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{t'-1}] \leq \mathcal{P}[\mathbf{X}_{\mu_{[\epsilon_1]}} \geq i]. \tag{33}$$

Hence, we conclude from (29), (30), (32), and (33) and from proposition 3.8 that $\mathbf{Z}_{t'} \overset{d}{\lhd} \mathbf{K}_{t'}$ for all $1 \leq t' \leq r'$.

Next, we claim that

$$\mathbf{Z}_{r'} \overset{d}{\rhd} \sum^{r'} \mathbf{X}_{\mu_{[\epsilon]}}, \tag{34}$$

where $\epsilon > 0$ depends on ϵ_1 and ϵ_2 and can be made arbitrarily small. Note that, for $j > 0$,

$$\begin{aligned}\mathbf{Y}_{(i,j)} \overset{d}{=} \sum^{\mathbf{Y}_{(i,j-1)}} \mathbf{1}_{(1-\frac{\epsilon_2}{r'})} &\overset{d}{=} \sum^{\mathbf{Y}_{(i,0)}} \mathbf{1}_{(1-\frac{\epsilon_2}{r'})^j} \\ &\overset{d}{\rhd} \sum^{\mathbf{X}_{\mu_{[\epsilon_1]}}} \mathbf{1}_{(1-j\frac{\epsilon_2}{r'})}\end{aligned}$$

so

$$\mathbf{Y}_{(i,r')} \overset{d}{\rhd} \sum^{\mathbf{X}_{\mu_{[\epsilon_1]}}} \mathbf{1}_{(1-\epsilon_2)}.$$

Note that, for any $h \in \mathbb{Z}^*$,

$$\lim_{\epsilon_1, \epsilon_2 \to 0} \mathcal{P}\left[\sum^{\mathbf{X}_{\mu_{[\epsilon_1]}}} \mathbf{1}_{(1-\epsilon_2)} = h\right] = \mathcal{P}[\mathbf{X}_\mu = h] = \mu(h).$$

It follows by proposition 2.3 that, for arbitrary $\epsilon > 0$, we can choose $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$g_\Phi\left(\mu, \mathfrak{D}_{\left(\sum^{\mathbf{X}_{\mu_{[\epsilon_1]}}} \mathbf{1}_{(1-\epsilon_2)}\right)}\right) < \epsilon,$$

and therefore

$$\mathbf{Y}_{(i,r')} \overset{d}{\rhd} \sum^{\mathbf{X}_{\mu_{[\epsilon_1]}}} \mathbf{1}_{(1-\epsilon_2)} \overset{d}{\rhd} \mathbf{X}_{\mu_{[\epsilon]}}.$$

Hence

$$|\mathbf{N}| = \mathbf{K}_{r'} \overset{d}{\rhd} \mathbf{Z}_{r'} \overset{d}{\rhd} \sum^{r'} \mathbf{X}_{\mu_{[\epsilon]}}, \tag{35}$$

where $\epsilon > 0$ is arbitrarily small. ∎

Proof of Theorem 3.9. Choose arbitrary $\epsilon_1, \epsilon_2 > 0$ such that $\epsilon_1 + \epsilon_2 < \epsilon$, and, combining the two previous lemmas with lemma 3.4, we deduce

$$|\mathbf{N}| \overset{d}{\rhd} \sum^{r - |\mathbf{A}|} \mathbf{X}_{\mu_{[\epsilon_1]}} \overset{d}{\rhd} \sum^{\sum^{r} \mathbf{1}_{1-\epsilon_2}} \mathbf{X}_{\mu_{[\epsilon_1]}} \overset{d}{\rhd} \sum^{r} \mathbf{X}_{\mu_{[\epsilon_1+\epsilon_2]}}. \quad ∎$$

This theorem allow us to derive a large deviation inequality regarding the growth rate of BFS neighborhoods. The proof is adapted from the upper bound proof of Cramer's Theorem in [9].

Lemma 3.12 *Let U be an endpoint arrangement satisfying $\lambda_U \xrightarrow{1} \lambda$ and $m = |U| \to \infty$, where λ has residual distribution $\mu = \mu_\lambda$. Let $R \subseteq U$ be the initial endpoint queue for an iteration of CM BFS and assume $r = |R| = o(m)$. Then, for any $\epsilon > 0$, there exists a value C_ϵ such that the following statements hold asymptotically:*

1. *If $M_1(\mu) < \infty$ then the size of the BFS neighborhood $\mathbf{N}$ satisfies*

$$\mathcal{P}\left[\frac{|\mathbf{N}|}{r} \leq M_1(\mu) - \epsilon\right] \leq e^{-C_\epsilon r}. \tag{36}$$

2. *If $M_1(\mu) = \infty$ then the size of the BFS neighborhood $\mathbf{N}$ satisfies*

$$\mathcal{P}\left[\frac{|\mathbf{N}|}{r} \leq 1/\epsilon\right] \leq e^{-C_\epsilon r}. \tag{37}$$

Proof. For any distribution μ, define

$$\Lambda_\mu(z) = \ln \mathcal{E}[e^{-z\mathbf{X}_\mu}]. \tag{38}$$

Note that if $\nu \to \mu$ then $\Lambda_\nu(z) \to \Lambda_\nu(z)$ for all $z \geq 0$, and since

$$\mathcal{E}[e^{-z\sum^r \mathbf{X}_\nu}] = \mathcal{E}[e^{-z\mathbf{X}_\nu}]^r = e^{r\Lambda_\nu(z)},$$

then by Chebyshev's Inequality,

$$\ln \mathcal{P}\left[\sum^r \mathbf{X}_\nu \leq rx\right] \leq r(\Lambda_\nu(s) + xs) \to r(\Lambda_\mu(s) + xs) \tag{39}$$

for all z.

Now, by Theorem 3.9, we have

$$|\mathbf{N}| \overset{d}{\triangleright} \sum^r \mathbf{X}_{\mu_{[\delta]}},$$

for arbitrary $\delta > 0$. If $M_1(\mu) < \infty$, then for any $\epsilon > 0$, we can choose z and δ appropriately such that $-\Lambda_{\mu_{[\delta]}}(z)/z \geq (M_1(\mu) - \epsilon)$ and set C_ϵ accordingly. If $M_1(\mu) = \infty$, then for any $1/\epsilon$, there exists values z, δ such that $-\Lambda_{\mu_{[\delta]}}(z)/z \geq 1/\epsilon$, and again we set C_ϵ accordingly. ∎

3.4 Halted CM BFS

We define a *halted CM BFS process* to be a CM BFS process which halts upon termination of the iteration in progress after a given number of CM steps. For halted BFS:

- σ denotes the number of BFS iterations that were completed. Hence the halting time of halted CM BFS is $\boldsymbol{\tau_\sigma}$.
- We define the i'th *halted BFS neighborhood* by

$$\mathbf{N}'_i = \begin{cases} \mathbf{N}_i & \text{if } i \leq \boldsymbol{\sigma}, \\ \emptyset & \text{if } i > \boldsymbol{\sigma}. \end{cases}$$

The following lemma will be used to obtain the lower bound for the diameter.

Lemma 3.13 *Consider an endpoint arrangement U such that $\lambda_U \xrightarrow{2} \lambda$, where λ has residual distribution $\mu = \mu_\lambda$, let the initial endpoint queue $R = U(W)$ have size $r = |R| = o(m)$, and consider CM BFS halted after $l = o(m)$ steps. Then, for any i and any $\epsilon > 0$,*

$$\mathcal{E}\left[\left|\mathbf{N}_i'\right|\right] \leq r \cdot M_1(\mu+\epsilon)^i. \tag{40}$$

Proof. Note that $|\mathbf{N}_1|$ is at most equal to the sum of the residual unexplored degrees of the endpoints matched to each $s \in R$. Hence, by linearity of expectation,

$$\mathcal{E}[|\mathbf{N}_1|] \leq r \cdot M_1(\boldsymbol{\nu}_0),$$

and since $|R| = o(m)$, then $\boldsymbol{\nu}_0 \xrightarrow{1} \mu$ holds asymptotically always.

Let us write $\mathbf{n}_i' = |\mathbf{N}_i'|$. Then, for arbitrary i, if $\boldsymbol{\tau}_{i-1} < l$ then

$$\mathcal{E}\left[\mathbf{n}_i' \;\middle|\; \mathbf{n}_{i-1}', \mathbf{U}'_{\boldsymbol{\tau}_{i-1}}\right] \leq \mathbf{n}_{i-1}' \cdot M_1(\boldsymbol{\mu}_{\boldsymbol{\tau}_{i-1}}),$$

and if $\boldsymbol{\tau}_{i-1} \geq l$ then

$$\mathbf{n}_i' = 0.$$

Since $l = o(m)$, then for any $t \leq l$,

$$M_1(\boldsymbol{\nu}_t) < M_1(\mu) + \epsilon$$

for arbitrary $\epsilon > 0$. Hence, we conclude that

$$\mathcal{E}\left[\mathbf{n}_i'\right] \leq r \cdot M_1(\mu+\epsilon)^i$$

for all i and for arbitrary $\epsilon > 0$. ∎

4 The Diameter of a Random Graph.

Given vertices u, v in a graph G, let $\delta(u,v)$ denote the distance from u to v, that is, the length of a shortest path from u to v. We set $\delta(u,v) = \infty$ if u and v are not connected. The diameter $\Delta(G)$ of a graph G is the maximum distance between any connected pair of vertices in G. In this section we compute $\Delta(\mathbf{G}_D)$ with asymptotic precision.

We begin by stating our main theorem regarding the diameter of a random graph. For a distribution μ, recall that z_μ denotes the least fixed point of the p.g.f. ψ_μ in the interval $[0,1]$ (see equation 20). We are only interested in distributions for which $z_\mu < 1$, or equivalently, $M_1(\mu) > 1$. For such a distribution μ, a random graph $\mathbf{G}_D$ with residual distribution $\mu_D \to \mu$ a.a.s. contains a giant connected component and a giant 2-core (see appendix).

Assuming $z_\mu < 1$, we define

$$M^*(\mu) = \psi_\mu'(z_\mu), \tag{41}$$

the derivative of the p.g.f. at z_μ. The significance of the value $M^*(\mu)$ is discussed in the appendix. In particular, by Theorem A.2 and statement 1 of proposition A.3, the fraction of endpoints of residual degree 1 in the 2-core of $\mathbf{G}_D$ converges to $M^*(\mu)$ w.e.h.p.

Also, let

$$a(\mu) = \begin{cases} 2 & \text{if } \mu(0) > 0 \\ 1 & \text{if } \mu(0) = 0 \text{ and } \mu(1) > 0 \\ 0 & \text{if } \mu(0) = \mu(1) = 0, \end{cases} \tag{42}$$

Theorem 4.1 *Let μ be a distribution satisfying $M_1(\mu) > 1$, consider a degree sequence D satisfying $\lambda_D \xrightarrow{2} \lambda$ or $\lambda_D \xrightarrow{1} \lambda$ with $M_2(\lambda) = \infty$, where λ has residual distribution $\mu_\lambda = \mu$. In addition, assume that if $\mu(0) = 0$ then $\mathbf{G}_D$ has no vertices of degree 1 (i.e. $\mu_D(0) = 0$) and if $\mu(1) = 0$ then then $\mathbf{G}_D$ has no vertices of degree 2 (i.e. $\mu_D(1) = 0$). Then,*

$$\frac{\Delta(\mathbf{G}_D)}{\ln n} \to \Delta_\mu \quad a.a.s. \tag{43}$$

where

$$\Delta_\mu = \frac{a(\mu)}{-\ln M^*(\mu)} + \frac{1}{\ln M_1(\mu)} \tag{44}$$

with $a(\mu)$ as defined in equation (42).

Informally, the two terms in equation (44) correspond to two different characteristics of a random graph $\mathbf{G}_D$ that determine its diameter. The second term measures the "average" distance across $\mathbf{G}_D$, while the first term gives the length of the longest isolated paths or cycles which can cause the distance between a particular pair of vertices to be significantly longer than the average.

In the simplest (and possibly the most "typical") situation, we have $\mu(0) > 0$ and $M_1(\mu) < \infty$; for example, sparse $\mathcal{G}_{n,p}$ falls into this category. In this situation, the diameter is determined by a longest shortest path between two vertices of degree 1 and will consist of a path from each of the vertices to the 2-core of the graph and a path connecting across the 2-core.

The proof of Theorem 4.1 proceeds as follows. First, we prove the upper bound, which is the more substantial portion of the proof. We show the upper bound first in the "typical" case described above, and then generalize to other situations. Finally, we compute a matching lower bound.

4.1 Upper Bound Proof of Theorem 4.1

Our principal proof strategy will be to examine the rate of growth of the neighborhoods around specific vertices. For any vertex v in a configuration (V, S, E), we define the i'th endpoint neighborhood $N_i(v)$ of v by

$$N_i(v) = \{s \in S : \delta(v, v(s)) = i \text{ and } \delta(v, v(E(s))) \geq i\},$$

so if we perform CM BFS in a random graph beginning with vertex set $W = \{v\}$, then the set $N_i(v)$ will correspond to the i'th BFS neighborhood $\mathbf{N}_i$ as defined in section 3.3.

For any vertex v, let

$$\gamma(v) = \begin{cases} \min\{i : |N_i(v)| \geq 3m^{1/2} \ln m\} \text{ if this set is nonempty,} \\ \frac{1}{2}\min\{i : |N_i(v)| = 0\} \text{ otherwise.} \end{cases}$$

Also, if v is a vertex in a random graph, we write we write $\mathbf{N}_i(v)$ and $\boldsymbol{\gamma}(v)$ to denote the respective random set and random quantity.

Lemma 4.2 *For a degree sequence D with $\lambda_D \xrightarrow{1} \lambda$, the graph $\mathbf{G}_D$ a.a.s. exhibits the property that*

$$\delta(u, v) \leq \gamma(u) + \gamma(v) + 1$$

for all connected pairs of vertices $u, v \in V$.

Proof. Given any pair of vertices u, v, we perform CM BFS, first starting with u and then starting with v, until endpoint neighborhoods $\mathbf{N}_u = \mathbf{N}_{\boldsymbol{\gamma}(u)}(u)$ and $\mathbf{N}_v = \mathbf{N}_{\boldsymbol{\gamma}(v)}(v)$ are exposed. Now, either

1. u and v are not connected,

2. A path from u to v has been exposed, or

3. 1 and 2 do not hold, and both $\mathbf{N}_u$ and $\mathbf{N}_v$ contain at least $3m^{1/2} \ln m$ unexposed endpoints.

In the first case, $\delta(u,v) = \infty$. In the second case, clearly $\delta(u,v) \leq \boldsymbol{\gamma}(u) + \boldsymbol{\gamma}(v) + 1$ Hence, we need only consider the third case.

Suppose we now expose all of the endpoints in $\mathbf{N}_u$; if any endpoint in $\mathbf{N}_u$ is matched to an endpoint in $\mathbf{N}_v$, then

$$\delta(u,v) \leq \boldsymbol{\gamma}(u) + \boldsymbol{\gamma}(v) + 1.$$

We claim that this occurs with probability $1 - o(n^2)$. To see this, we observe that a given endpoint in $\mathbf{N}_u$ matches to $\mathbf{N}_v$ with probability at least

$$|\mathbf{N}_v| / m \geq m^{-1/2}.$$

Now, if a particular endpoint in $\mathbf{N}_u$ does not match to $\mathbf{N}_v$, it may match into $\mathbf{N}_u$, reducing the number of unexposed endpoints in $\mathbf{N}_u$ by 2. Nevertheless, if we sequentially match all of the endpoints in $\mathbf{N}_u$, there are at least $|\mathbf{N}_u| / 2$ chances to find a connection to $\mathbf{N}_v$.

The probability that no connection is found is therefore at most

$$\begin{aligned}
(1 - m^{-1/2})^{3m^{1/2} \ln m} &= \left(\left(1 - \frac{1}{m^{1/2}} \right)^{m^{1/2}} \right)^{3 \ln m} \\
&= \left(\frac{1}{e - o(1)} \right)^{3 \ln m} \\
&= O(m^{-3}).
\end{aligned}$$

By considering all $\binom{n}{2} = O(m^2)$ pairs of vertices, we conclude that this event a.a.s. occurs for no such pair, and the lemma is proved. ∎

Lemma 4.2 proves that in order to find an upper bound on the diameter of $\mathbf{G}_D$, it suffices to bound the maximum value of $\boldsymbol{\gamma}(v)$ for $v \in V$. However, it is not necessary to consider all vertices in V; the next lemma we prove allows us to narrow down the set of vertices which can contribute to the diameter of $\mathbf{G}_D$.

Recall that the 2-core of a graph is the maximal induced subgraph of minimum degree 2. For a graph $G = (V, E)$, let $d_2(v)$ denote the degree of a vertex v in the 2-core of G, that is, the number of edges in the 2-core of G which are incident on V.

Lemma 4.3 *For any graph $G = (V, E)$, and any vertex $v \in V$, if there exists a vertex v' such that $\delta(v, v') = \Delta(G)$, then either $d_2(v) = d(v)$ or $d_2(v) = 0$ and $d(v) = 1$.*

Proof. If $0 < d_2(v) < d(v)$, then for any vertex v' connected to v, we can find a vertex v'' such that any path from v' to v'' must pass through v and therefore $\delta(v', v'') > \delta(v, v')$. Hence either $d_2(v) = 0$ or $d_2(v) = d(v)$. Now, if $d_2(v) = 0$ it follows that d_2 is not in the 2-core, and therefore there are no cycles in G which contain v. Therefore, if $d(v) \geq 2$ and $d_2(v) = 0$, then once again, for every v' connected to v, we can find a vertex v'' such that any path from v' to v'' must path through v and so $\delta(v', v'') > \delta(v, v')$. ∎

By lemma 4.3, we only need to consider vertices which are either entirely in the 2-core or have degree 1. Note, however, that a vertex of degree 1 may or may not be connected to the 2-core by a path. Hence, our proof of the upper bound of Theorem 4.1 will consider the following three cases separately:

1. $d(v) = 1$ and v is connected to the 2-core;
2. $d(v) = 1$ and v is in a tree component;
3. the degree of v in the 2-core satisfies $d_2(v) = d(v) \geq 2$. In this situation, we also distinguish between the case when the minimum degree of the entire graph is 3 or greater, and the case where the minimum degree is at most 2.

In most cases, the diameter of $\mathbf{G}_D$ will occur between two vertices of type 1, that is, two vertices of degree 1 which are connected to the 2-core.

4.1.1 Vertices of degree 1, connected to the 2-core

In this subsection, we consider the neighborhoods in $\mathbf{G}_D$ of a vertex of degree 1 which is connected to the 2-core of $\mathbf{G}_D$. We assume throughout $\lambda_D \xrightarrow{2} \lambda$ or $\lambda_D \xrightarrow{1} \lambda$ with $M_2(\lambda) = \infty$, where λ has residual distribution $\mu = \mu_\lambda$. Recall that $M^*(\mu)$ gives the derivative of the p.g.f. ψ_μ at the fixed point z_μ (equation 41). $M^*(\mu)$ has an alternate interpretation which we shall make use of in this section. Recall that by Lemma 3.3, if $\boldsymbol{\tau}$ is the halting time of the CM 2-core algorithm, then $\boldsymbol{\mu}_{\boldsymbol{\tau}} \to \rho_\mu$ w.e.h.p., where ρ_μ is the distribution defined in (21). By manipulating generating functions, it can be shown (see appendix A) that

$$M^*(\mu) = \rho_\mu(1).$$

Lemma 4.4 *Choose any $v \in \mathbf{G}_D$ such that $d(v) = 1$. Then, for any $\epsilon > 0$,*

$$\mathcal{P}\left[\frac{\boldsymbol{\gamma}(v)}{\ln n} \geq \frac{1}{-\ln M^*(\mu)} + \frac{1}{2\ln M_1(\mu)} + \epsilon\right] = o(n^{-1}). \tag{45}$$

Proof. In order to bound $\boldsymbol{\gamma}(v)$, we shall execute a specialization of the CM algorithm which combines the CM 2-core algorithm and CM BFS breadth-first-search. First, we execute the protected 2-core algorithm with protected set $W = \{v\}$. Then, if the single endpoint belonging to v remains unexposed at the halting time $\boldsymbol{\tau}$, we execute CM BFS starting with the vertex v.

Based on lemma 3.3, the unexposed residual distribution at the halting time satisfies $\boldsymbol{\mu}_{\boldsymbol{\tau}} \to \rho_\mu$ w.e.h.p. Hence, we discard the exponentially small probability that this convergence fails, and assume that $\boldsymbol{\mu}_{\boldsymbol{\tau}}$ is arbitrarily close to ρ_μ.

Also, at time $\boldsymbol{\tau}$, there are no endpoints of residual degree 1 other than v. We note that if v's unique endpoint has been exposed at time $\boldsymbol{\tau}$ then v is not connected to the 2-core of $\mathbf{G}_D$ and belongs to a tree component; we shall deal with tree components separately.

We analyze the BFS in three phases. For this proof we let $\mathbf{n}_i = |\mathbf{N}_i|$ denote the number of endpoints in the BFS queue after i iterations of the BFS. The phases are:

1. $\mathbf{n}_i = 1$ to $\mathbf{n}_i \geq \ln\ln n$;
2. $\mathbf{n}_i = \ln\ln n$ to $\mathbf{n}_i \geq \ln^2 n$;
3. $\mathbf{n}_i = \ln^2 n$ to $\mathbf{n}_i \geq 3m^{1/2}\ln n$.

In the original graph, phase 1 corresponds to first performing BFS from v until the 2-core reached, and then continuing the BFS in the 2-core until a neighborhood of size $\ln\ln n$ is found. Intuitively, phase 1 will include a large number of iterations if the path from v to the 2-core is very long, if closest vertex to v in the 2-core is part of a long isolated cycle, or, more generally, if the small BFS neighborhoods around v grow at an abnormally slow rate.

Phase 2 transitions from a BFS neighborhood of size $\ln\ln n$ to a BFS neighborhood of size $\ln^2 n$ in the 2-core. Typically, phase 2 will include only a small number of iterations, and phase 2 serves mainly to transition from the "small" neighborhoods in phase 1 to the "large" neighborhoods of phase 3. Then, in phase 3, the neighborhoods are large enough so that their growth rate is highly predictable using the tools developed in section 3.3.

We compute the total number of BFS iterations by considering "good" and "bad" iterations in each phase (their properties are defined below); we let $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$, $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ and denote the number of good and bad iterations in each phase, respectively. Informally, a "good" iteration occurs if the size of the BFS neighborhood grows sufficiently quickly, and a "bad" iteration occurs otherwise.

Phase 1. In phase 1, a good iteration occurs if $\mathbf{n}_{i+1} > \mathbf{n}_i$ or $\mathbf{n}_{i+1} = 0$, and a bad iteration occurs otherwise. Now, if no cross-edges occur, then $\mathbf{n}_{i+1}$ is equal to the sum of the residual degrees of the endpoints matched to the endpoints in $\mathbf{N}_i$. Recall that the unexposed residual distribution satisfies $\boldsymbol{\mu}_t \to \rho_\mu$ w.e.h.p., and therefore the probability that an unexposed endpoint chosen uniformly at random has residual degree 1 is

$$\boldsymbol{\mu}_t(1) \to \rho_\mu(1) = M^*(\mu)$$

w.e.h.p. Also, since all vertices outside of the BFS queue have unexposed degree at least 2 (or 0), then the only way a bad iteration can occur without a cross-edge is if every endpoint in $\mathbf{N}$ matches to an endpoint of residual unexposed degree 1. This probability is maximized if $\mathbf{n}_i = 1$, in which case the single endpoint in $\mathbf{N}_i$ matches to an endpoint of unexposed residual degree 1 with probability at most $\rho_\mu(1) + o(1) = M^*(\mu) + o(1)$.

Recall that, in the context of CM BFS (section 3.3), $\mathbf{Q}$ denotes the set of unexposed endpoints in the BFS queue; in particular, a cross-edge occurs if and only if an endpoint in $\mathbf{Q}$ is chosen at random during an even time step. Hence, the probability of encountering a cross-edge at any given time is $|\mathbf{Q}| \,/\, |\mathbf{U}|$.

Also, at any time during any BFS iteration in phase 1, if $|\mathbf{Q}| > 3\ln\ln n$, it is clear that, even if all of the (at most $\ln\ln n$) remaining unexposed endpoints in the initial queue R form cross-edges, the size of the queue at the the end of the iteration will be greater than $\ln\ln n$ and therefore phase 1 will end. Hence, during phase 1, we can assume that the probability of encountering a cross-edge at any particular step is $O((\ln\ln n)/m)$. And, therefore, the probability of encountering more than one cross edge during any of the first $O(\ln^{O(1)} n)$ steps of phase 1 is $O(n^{-2}\ln^{O(1)} n) = \tilde{O}(n^{-2})$. Here $O(\ln^{O(1)} n)$ serves simply as an upper bound to the number of CM steps which occur in phase 1, as we shall see below.

For any given BFS iteration in phase 1, barring a cross-edge, we must have $\mathbf{n}_{i+1} \geq \mathbf{n}_i$, and with at most one cross-edge, we have $\mathbf{n}_{i+1} \geq \mathbf{n}_i - 2$. Hence, if at most one cross-edge occurs in phase one, then $\mathbf{G}_1 \leq \ln\ln n + 2$, since each good iteration increases the number of endpoints in $\mathbf{N}$. Recall that the total number of iterations in phase 1 is $\mathbf{G}_1 + \mathbf{B}_1$. It follows that for any constant c, we have

$$\begin{aligned}\mathcal{P}[\mathbf{G}_1 + \mathbf{B}_1 \geq c\ln n] &= \mathcal{P}[\mathbf{B}_1 + \ln\ln n + 2 \geq c\ln n] + \tilde{O}(n^{-2}) \\ &= \mathcal{P}[\mathbf{B}_1 \geq (c - o(1))\ln n] + \tilde{O}(n^{-2}).\end{aligned}$$

It follows that, for any constant c, with probability $1 - \tilde{O}(n^{-2})$, the event $\mathbf{B}_1 \geq c\ln n$ occurs if and only at least $c\ln n$ of the first $\ln\ln n + 2 + c\ln n$ iteratations in phase 1 are bad. As shown above, the probability that any given iteration in phase 1 is a bad iteration is w.e.h.p. bounded

above by $\rho_\mu(1) + o(1) = M^*(\mu) + o(1)$. Thus, we compute

$$\begin{aligned}\mathcal{P}[\mathbf{B}_1 \geq c \ln n] &\leq \binom{\ln\ln n + 2 + c\ln n}{c\ln n}(M^*(\mu)+o(1))^{c\ln n} + \tilde{O}(n^{-2}) \\ &\leq \left(1 + O\left(\frac{\ln\ln n}{\ln n}\right)\right)^{c\ln n}(M^*(\mu)+o(1))^{c\ln n} + \tilde{O}(n^{-2}) \\ &\leq (M^*(\mu)+o(1))^{c\ln n} + \tilde{O}(n^{-2}). \end{aligned} \tag{46}$$

In particular, for any $\epsilon_1 > 0$, we let

$$c = \frac{1}{-\ln M^*(\mu)} + \epsilon_1$$

and compute $\mathcal{P}[\mathbf{B}_1 \geq c\ln n] \leq o(n^{-1})$. Hence, for arbitrary $\epsilon_1 > 0$, with probability $1 - o(n^{-1})$,

$$\mathbf{B}_1 \leq \left(\frac{1}{-\ln M^*(\mu)} + \epsilon_1\right)\ln n, \text{ and } \mathbf{G}_1 = O(\ln\ln n). \tag{47}$$

We point out that it is not necessarily the case that $\mathbf{n}_i$ ever reaches $\ln\ln n$; it is possible that $\mathbf{n}_i$ becomes 0 at some point, for example if $r^{i-1} = 2$ and the two endpoints in $\mathbf{N}_{i-1}$ are matched to each other. This occurs with probability $\Theta(m^{-1}) = \Theta(n^{-1})$; however, if this event does occur, the analysis above shows that, with probability $1-o(n^{-1})$, it occurs after at most $\left(\frac{1}{-\ln M^*(\mu)} + o(1)\right)\ln n$ iterations.

Phase 2. In phase 2 we transition from endpoint sets of size $\ln\ln n$ to size $\ln^2 n$. Unlike phase 1, we do not consider a phase 2 iteration to be good simply because $\mathbf{n}_{i+1} > \mathbf{n}_i$. Instead, we will consider the actual rate of neighborhood growth.

Proposition A.3 in the appendix shows that $M_1(\rho_\mu) = M_1(\mu)$. Also, by assumption of Theorem 4.1, either $\mu_D \xrightarrow{1} \mu$ or $\mu_D \to \mu$ with $M_1(\mu) = \infty$, so it follows from lemma 3.3 that $\boldsymbol{\mu}_t \xrightarrow{1} \rho_\mu$ w.e.h.p. or $\boldsymbol{\mu}_t \to \rho_\mu$ w.e.h.p. with $M_1(\rho_\mu) = M(\mu) = \infty$. In particular, the average residual unexposed degree at the halting time of the protected 2-core algorithm converges to $M_1(\mu)$ w.e.h.p. Hence, in this phase, we define a bad iteration to be an iteration in which

- $\mathbf{n}_{i+1} \leq (M_1(\mu) - \delta)\mathbf{n}_i$ if $M_1(\mu) < \infty$.
- $\mathbf{n}_{i+1} \leq (1/\delta)\mathbf{n}_i$ if $M_1(\mu) = \infty$.

for arbitrarily small $\delta > 0$. The number of good iterations is thus bounded above by

$$\frac{2\ln\ln n}{\ln(M_1(\mu) - \delta)} = O(\ln\ln n)$$

in the first case and

$$\frac{2\ln\ln n}{\ln(1/\delta)} = O(\ln\ln n)$$

in the second.

By lemma 3.12, the probability of a bad iteration is at most

$$e^{-C_\delta \mathbf{n}_i} \leq (\ln n)^{-\Theta(1)}$$

for a constant C_δ. For any $\epsilon_2 > 0$, a routine manipulation of binomial distributions yields

$$\mathcal{P}[\mathbf{B}_2 \geq \epsilon_2 \ln n] \leq (\ln n)^{-\Theta(\ln n)} = n^{-\Theta(\ln\ln n)}.$$

Hence, for arbitrary $\epsilon_2 > 0$, with probability $1 - o(n^{-1})$,

$$\mathbf{B}_2 \leq \epsilon_2 \ln n, \text{ and } \mathbf{G}_2 = O(\ln \ln n). \tag{48}$$

Phase 3. In phase 3, a bad iteration defined as in phase 2; now, however, we have $\mathbf{n}_i \geq \ln^2 n$, so the probability of a bad iteration is at most

$$e^{-C_\delta \ln^2 n} = n^{-\Omega(\ln n)}.$$

It follows that, with probability $1 - o(n^{-1})$, $\mathbf{B}_3 = 0$, and in phase 3 we need only count good iterations. By setting δ appropriately, for arbitrary $\epsilon_3 > 0$, we attain

$$\mathbf{G}_3 \leq \log_{M_1(\mu)-\delta}(3m^{1/2} \ln n) \leq \left(\frac{1}{2 \ln M_1(\mu)} - \epsilon_3\right) \ln n, \quad \text{if } M_1(\mu) < \infty, \tag{49}$$

and

$$\mathbf{G}_3 \leq \log_{1/\delta} m^{1/2} \leq \epsilon_3 \ln n, \quad \text{if } M_1(\mu) = \infty. \tag{50}$$

Finally, we set $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$ and we add up the good and bad iterations in the three phases as given by (47), (48), and (49) or (50) to yield equation (45). ∎

4.1.2 Vertices of higher degree

In this subsection, we consider the neighborhoods vertices of degree 2 or greater in $\mathbf{G}_D$. For any vertex v, recall that $\mathbf{d}_2(v)$ denotes the degree of v in the 2-core of $\mathbf{G}_D$; hence, if $\boldsymbol{\tau}$ is the halting time of the CM 2-core process, then $\mathbf{d}_2(v) = \mathbf{d}_{\boldsymbol{\tau}}(v)$. Also, recall that by lemma 4.3, for $d(v) \geq 2$, we are only interested in vertices for which $\mathbf{d}_2(v) = d(v)$, since otherwise a longest shortest path in $\mathbf{G}_D$ cannot begin or end at v.

Lemma 4.5 *Choose any $v \in \mathbf{G}_D$ such that $d(v) \geq 2$. Then, for any $\epsilon > 0$*

$$\mathcal{P}\left[\frac{\boldsymbol{\gamma}(v)}{\ln n} \geq \frac{1}{-2 \ln M^*(\mu)} + \frac{1}{2 \ln M_1(\mu)} + \epsilon \;\middle|\; \mathbf{d}_2(v) = d(v)\right] = o(n^{-1}). \tag{51}$$

Proof. The proof proceeds as in lemma 4.4. We execute first the CM 2-core algorithm but now we use the protected set $W = \emptyset$. Once the 2-core has been found, we execute CM BFS beginning with the vertex v. By assumption, all of v's endpoints remain unexposed at the time that the 2-core is found. Our analysis will employ the same three stages as in lemma 4.4. Clearly, the arguments regarding phases 2 and 3 are identical to the case where $d(v) = 1$. Hence, we must deal with phase 1.

We note that, comparing equations (45) and (51), in order to attain the required bound, we must reduce the duration of phase 1 from $\frac{\ln n}{-\ln M^*(\mu)}$ to $\frac{\ln n}{-2 \ln M^*(\mu)}$. In order to do so, we consider phase 1 in slightly more detail.

Since our vertex v has degree at least 2, the CM BFS begins with a neighborhood of size $\mathbf{n}_0 \geq 2$. Recall that a cross-edge occurs during phase 1 with probability $\tilde{O}(n^{-1})$. Now, the probability of experiencing a bad iteration without a cross-edge is at most $(\rho_\mu(1) + o(1))^{\mathbf{n}_i} = (M^*(\mu) + o(1))^{\mathbf{n}_i}$. Without any cross-edges, we must have $\mathbf{n}_{i+1} \geq \mathbf{n}_i$, which implies $\mathbf{n}_i \geq d(v) \geq 2$ throughout the phase.

On the other hand, if a cross-edge does occur, then $\mathbf{n}_i$ can decrease. If $\mathbf{n}_{i+1} = 0$ (for example, if $\mathbf{n}_i = 2$ and a horizontal cross-edge matches both endpoints in $\mathbf{n}_i$), then phase 1 ends immediately, as does the entire BFS. However, if $\mathbf{n}_{i+1} = 1$, then a bad iteration becomes more probable. In

order to handle this situation, we note that a cross-edge is sufficiently unlikely that, for any $\epsilon > 0$, the probability that a cross-edge occurs either preceded of followed by at least $\epsilon \ln n$ iterations of phase 1 is at most

$$\tilde{O}(n^{-1})(\rho_\mu(1) + o(1))^{\epsilon \ln n} = o(n^{-1}).$$

Therefore, with probability $1 - o(n^{-1})$, this does not occur, and we may assume that the neighborhood size is at least 2 throughout phase 1.

Hence, a bad iteration requires that at least 2 endpoints of residual degree 1 are chosen consecutively, and this occurs with probability at most $(M^*(\mu) + o(1))^2$. Similarly to equation (46), we now deduce that

$$\begin{aligned}\mathcal{P}[\mathbf{B}_1 \geq c \ln n] &\leq \left(M^*(\mu) + o(1))^2\right)^{c \ln n} + o(n^{-1}) \\ &\leq (M^*(\mu) + o(1))^{2c \ln n} + o(n^{-1}),\end{aligned}$$

and the factor of 2 in the exponent carries through the computations in 4.4 to yield equation (51). ∎

4.1.3 Graphs with minimum degree at least 3

Here we consider the case where $\mu(0) = \mu(1) = 0$, hence, as assumed in Theorem 4.1, $\mathbf{G}_D$ has no vertices of degree 1 or 2.

Lemma 4.6 *Assume $\mu(0) = \mu(1) = 0$, and assume $\mathbf{G}_D$ has minimum degree 3. Then, for any vertex v and any $\epsilon > 0$,*

$$\mathcal{P}\left[\frac{\boldsymbol{\gamma}(v)}{\ln n} \geq \frac{1}{2 \ln M_1(\mu)} + \epsilon\right] = o(n^{-1}). \tag{52}$$

Proof. Again we use the same three phases as in the proof of lemma 4.4. However, we change the definition of good and bad iterations during the first phase. Now, we consider a bad iteration to be any iteration in which $\mathbf{n}_{i+1} \leq 2\mathbf{n}_i$. We note that since all vertices have degree at least 3 (and residual degree at least 2), a bad iteration can only occur with a cross-edge. As shown above, with probability $1 - o(n^{-1})$ at most one cross-edge occurs during the first phase, and with a single cross-edge, we have $\mathbf{n}_{i+1} \geq \mathbf{n}_i - 2$.

Since a good iteration doubles the size of the BFS endpoint queue, then (barring multiple cross-edges), at most $O(\ln \ln \ln n) = o(\ln n)$ good iterations can occur during phase 1. Hence, with probability $1 - o(n^{-1})$, both phases 1 and 2 have $o(\ln n)$ iterations and the previous result regarding the duration of phase 3 yields equation (52). ∎

4.1.4 Tree Components

The structure of a random graph with a fixed degree sequence D satisfying $\lambda_D \xrightarrow{2} \lambda$ or $\lambda_D \xrightarrow{1} \lambda$ with $M_2(\lambda) = \infty$ was described by Molloy and Reed [12, 13]. The results of Molloy and Reed which are pertinent to this paper can be summarized as follows (see appendix for more details). If the residual distribution $\mu = \mu_\lambda$ satisfies $M_1(\mu) > 1$, then the graph $\mathbf{G}_D$ a.a.s. contains a giant connected component; and, if the giant component is removed, the residual graph has the structure of the random graph $\mathbf{G}_{\mathbf{D}'}$ where the random degree sequence $\mathbf{D}'$ satisfies $\lambda_{\mathbf{D}'} \xrightarrow{2} \lambda'$ a.a.s. In particular, this limiting distribution λ' has residual distribution $\mu_{\lambda'} = \xi_\mu$, and ξ_μ satisfies $M_1(\xi_\mu) = M^*(\mu)$ (see proposition A.3). Also, all tree components a.a.s. lie outside of the giant component (or, equivalently, the giant component is a.a.s. not a tree component).

A straightforward argument shows that for

$$k > \left(\frac{1}{-\ln M_1(\xi_\mu)} + \epsilon\right) \ln n = \left(\frac{1}{-\ln M^*(\mu)} + \epsilon\right) \ln n,$$

the k'th BFS neighborhood for a vertex of degree 1 in the graph $\mathbf{G_{D'}}$ (which corresponds to $\mathbf{G}_D$ with the giant component removed) has expected size $o(n^{-1})$ and therefore is empty with probability $1 - o(n^{-1})$. Hence, the diameter of the largest tree component is a.a.s. less than $(\frac{1}{-\ln M^*(\mu)} + \epsilon) \ln n$ for arbitrary ϵ; in particular, the diameter of the giant component is greater than the diameter of any tree component.

4.1.5 Upper Bound

The proof of the upper bound of Theorem 4.1 follows immediately from the previous lemmas using the first moment method. Specifically, the above lemmas show that, for any $\epsilon > 0$, the expected number of vertices in $\mathbf{G}_D$ with

$$\frac{\boldsymbol{\gamma}(v)}{\ln n} \geq \frac{1}{2}\left(\frac{a(\mu)}{-\ln M^*(\mu)} + \frac{1}{\ln M_1(\mu)} + \epsilon\right)$$

is $n \cdot o(n^{-1}) = o(1)$. Hence, with probability $1 - o(1)$, no such vertex exists, and the upper bound follows. ∎

4.2 Lower Proof of Theorem 4.1

In order to prove the lower bound of Theorem 4.1, we shall demonstrate that for any $\epsilon > 0$, there a.a.s. exist a pair of vertices u, v in $\mathbf{G}_D$ with distance $\delta(u, v) \geq (\Delta_\mu - \epsilon) \ln n$, where

$$\Delta_\mu = \frac{a(\mu)}{-\ln M^*(\mu)} + \frac{1}{\ln M_1(\mu)}$$

is the value computed in equation (43) in Theorem 4.1. Recall that, intuitively, the second term $\frac{1}{\ln M_1(\mu)}$ in the equation above gives the distance separating a typical pair of vertices, while the first term $\frac{a(\mu)}{-\ln M^*(\mu)}$ describes the length of the long isolated paths or cycles which cause the diameter to differ from the "average" distance. Accordingly, we prove the lower bound by first showing that almost all vertices in a random graph with $\mu(0) = 0$ are separated by a shortest path of distance at least $\frac{\ln n}{\ln M_1(\mu)}(1 - o(1))$, and then finding a pair of vertices separated by an additional distance of $\frac{a(\mu) \ln n}{-\ln M^*(\mu)}$.

For the following lemma, we assume $\mu(0) = 0$, but we drop the assumption that $\mathbf{G}_D$ must have minimum degree 2. That is, for this lemma only, it suffices that $\mathbf{G}_D$ have $o(n)$ vertices of degree 1.

Lemma 4.7 *Let $\lambda_D \xrightarrow{2} \lambda$ or $\lambda_D \xrightarrow{1} \lambda$ with $M_2(\lambda) = \infty$, where λ has residual distribution $\mu_\lambda = \mu$, and assume $\mu(0) = 0$ and $\mu(1) < 1$. Let u, v be vertices in $\mathbf{G}_D$ such that $d(u) = \Theta(1)$ and $d(v) = \Theta(1)$. Then,*

1. *u and v are connected with probability $1 - o(1)$.*
2. *For any $\epsilon > 0$, with probability $1 - o(1)$,*

$$\delta(u, v) > (1 - \epsilon)\frac{\ln n}{\ln M_1(\mu)}. \tag{53}$$

Proof. We trace the BFS neighborhoods of u and v as in the upper bound proof. First, we execute the CM 2-core algorithm with protected set $\{u, v\}$. Since $\mu(0) = 0$, the probability generating function of μ satisfies $\psi_\mu(0) = 0$. Hence, by lemma 3.3 the CM 2-core algorithm terminates after $o(m)$ steps w.e.h.p. Since u and v have constant degree, then with probability $1-o(1)$, all endpoints belonging to u and v remain unexposed at this time.

Then, beginning with u, we perform a CM BFS search, halted after $m^{1-\epsilon_0}$ steps, for a value $\epsilon_0 > 0$ to be specified afterwards. Now, as with the upper bound proof (phases 1-3), the probability that $\mathbf{n}_{i+1} < \mathbf{n}_i$ for any given iteration is $O(m^{-1})$, hence with probability $1-o(1)$, this never occurs. Hence, by analysis similar to the upper bound proof, when the halted BFS terminates, with probability $1 - o(1)$, we will have found a large set of at least $3m^{1/2} \ln n$ unexposed endpoints connected to u. We then perform a similar BFS beginning with v, and conclude that with probability $1-o(1)$, a path connecting u to v will be found.

For the second part of the lemma, we must show that, with probability $1 - o(1)$, any path from u to v is longer than $(1-\epsilon)\frac{\ln n}{\ln M_1(\mu)}$. Note that, since the vertex v has $\Theta(1)$ unexposed endpoints, then with probability $1-o(1)$, none of these endpoints become exposed during $m^{1-\epsilon_0} = o(m)$ steps of the CM algorithm. Hence, with probability $1 - o(1)$, the distance $\delta(u, v)$ is greater than the number of BFS iterations completed during the halted BFS. We shall prove that, for any $\epsilon > 0$, with probability $1 - o(1)$, at least $(1-\epsilon)\frac{\ln n}{\ln M_1(\mu)}$ BFS iterations will have been completed when the CM BFS halts after $m^{1-\epsilon_0}$ steps.

Let $\mathbf{n}'_i = |\mathbf{N}'_i|$, so $\mathbf{n}'_i$ denotes size of the i'th BFS endpoint neighborhood if the BFS has not yet halted and 0 otherwise. Hence, by lemma 3.13, for any i, and for any $\epsilon_1 > 0$

$$\mathcal{E}[\mathbf{n}'_i] \leq (M_1(\mu) + \epsilon_1)^i .$$

Let $h = (1-\epsilon)\frac{\ln n}{\ln M_1(\mu)}$ and note that by linearity of expectation,

$$\mathcal{E}\left[\sum_{i=1}^{h} \mathbf{n}'_i\right] \leq \sum_{i=1}^{h} (M_1(\mu) + \epsilon_1)^i \leq h(M_1(\mu) + \epsilon_1)^h.$$

Choosing ϵ_1 sufficiently small, we have

$$\mathcal{E}\left[\sum_{i=1}^{h} \mathbf{n}'_i\right] \leq O(n^{1-C}) \tag{54}$$

for a constant $C > 0$.

Now, recall that the CM BFS is halted after $m^{1-\epsilon_0}$ steps, for arbitrary $\epsilon_0 > 0$. By definition, h iterations of CM BFS are completed at the halting time if and only if $\sum_{i=1}^{h} \mathbf{n}'_i \leq m^{1-\epsilon_0}$. We now set $\epsilon_0 < C$, where C is the constant in equation (54), and conclude by Chevyshev's inequality that

$$\mathcal{P}\left[\sum_{i=1}^{h} \mathbf{n}'_h > m^{1-\epsilon_0}\right] = o(1). \tag{55}$$

Hence, with probability $1 - o(1)$, h iterations complete during the halted BFS, and no endpoints which belong to v are encountered within distance $h = (1-\epsilon)\frac{\ln n}{\ln M_1(\mu)}$ of u. ∎

For graphs with minimum degree 3, the constant Δ_μ in Theorem 4.1 is given by

$$\Delta_\mu = \frac{1}{\ln M_1(\mu)}.$$

The lower bound in this case, as stated in the following corollary, is obtained by choosing any pair of vertices u, v and invoking lemma 4.7.

Corollary 4.8 *Assume $\mu(0) = \mu(1) = 0$, and assume $\mathbf{G}_D$ has no vertices of degree less than 3. Then, for every $\epsilon > 0$, with probability $1 - o(1)$, there exists a pair of vertices u, v such that $\delta(u, v) \geq (1 - \epsilon)\Delta_\mu \ln n$.*

Next, we consider graphs for which $\mu(0) > 0$, so the minimum nonzero degree is 1.

Lemma 4.9 *Assume $\mu(0) > 0$. Then, for every $\epsilon > 0$, with probability $1 - o(1)$, there exists a pair of vertices u, v such that $\delta(u, v) \geq (1 - \epsilon)\Delta_\mu \ln n$.*

Proof. Again we execute the CM 2-core algorithm, but now we choose a protected set W which consists of $\Theta(n^{C_\epsilon})$ vertices of degree 1; the constant $0 < C_\epsilon < 1$ will be specified further on. We let W' denote the set of vertices $w \in W$ for which the (unique) endpoint belonging to w remains unexposed when the protected 2-core algorithm halts. For any vertex $w \in W$, the event $w \in W'$ clearly occurs with probability $\Theta(1)$, and it is straightforward to show that W' a.a.s. (and w.e.h.p.) contains a constant fraction of the vertices in W.

At this point we attempt to find two vertices $u, v \in W'$ with k consecutive endpoint neighborhoods of size 1, where

$$k = \left(\frac{1}{-\ln M^*(\mu)} - \epsilon_0 \right) \ln n \tag{56}$$

for an arbitrary $\epsilon_0 > 0$. We shall refer to this as the *BFS probe* algorithm. First, choose any vertex $w \in W'$, and execute CM BFS until either

- k iterations are completed, and the sizes of the first k BFS neighborhoods satisfy

  $$\mathbf{n}_0 = \mathbf{n}_1 = \cdots = \mathbf{n}_k = 1,$$

  or

- for some $j \leq k$, $\mathbf{n}_j \neq 1$.

We repeat this procedure, choosing a new vertex from W' each time until all endpoints belonging to vertices in W' have been exposed.

We note that the BFS probe will expose at most $O(\ln n)$ endpoints for each vertex in W, and since $|W| = o(m)$, then $o(m)$ endpoints overall will become exposed. Hence, for any $\epsilon_1 < 0$, the fraction of endpoints of residual degree 1 will never drop below $M^*(\mu) - \epsilon_1$ at any time during these BFS searches. Therefore, the probability that any particular $w \in W'$ produces a chain of length k is at least

$$(M^*(\mu) - \epsilon_1)^k = (M^*(\mu) - \epsilon_1)^{\left(\frac{1}{\ln M^*(\mu)} - \epsilon_0\right) \ln n} = \Omega(n^{-(1 - C_{\epsilon_0})}). \tag{57}$$

The probability of failing to find such a chain in j attempts is therefore at most

$$(1 - \Omega(n^{-(1-C_{\epsilon_0})}))^j = e^{j \ln(1 - \Omega(n^{-(1-C_{\epsilon_0})}))} = e^{-\Omega(jn^{-(1-C_{\epsilon_0})})}, \tag{58}$$

and since the number of attempts is $j = |W'| = \Theta(n^{C_\epsilon})$, we choose $C_\epsilon > 1 - C_{\epsilon_0}$. This guarantees that at least 2 such chains are found with probability $1 - o(1)$. Also, if a particular vertex w produces such a chain, with probability $1 - o(1)$, the unexposed endpoint at the end of the chain will remain unexposed through the rest of the BFS probe with probability $1 - o(1)$. Hence, at the end of the BFS probe, with probability $1 - o(1)$, at least two such chains will be remain.

Now, let u, v be the two vertices in W' which are found to have k BFS neighborhoods of size 1, and let u' and v' respectively denote the vertices in the k'th BFS neighborhoods of u and v. At this point, we note that any path connecting u to v must pass through u' and v'. Thus,

$$\delta(u, v) = \delta(u', v') + \delta(u, u') + \delta(v, v'). \tag{59}$$

Both $\delta(u,u')$ and $\delta(v,v')$ have length at least k as given in equation (56). A lower bound for $\delta(u',v')$ follows from lemma 4.7. This completes the proof. ∎

The generalization to the case where $\mathbf{G}_D$ has minimum degree exactly 2 (so $\mu(0) = 0$ and $\mu(1) > 0$) is straightforward, as shown below.

Lemma 4.10 *Assume $\mu(0) = 0$ and $\mu(1) > 0$, and assume $\mathbf{G}_D$ has minimum degree 2. Then, for every $\epsilon > 0$, with probability $1 - o(1)$, there exist a pair of vertices u, v such that $\delta(u,v) \geq (1-\epsilon)\Delta_\mu \ln n$.*

Proof. The proof is very similar to the proof for graphs with minimum degree 1, except that the BFS probe now searches for a chain of BFS neighborhoods containing 2 endpoints, rather than only 1. We set

$$k = \left(\frac{1}{2 \ln M^*(\mu)} - \epsilon_0 \right) \ln n \tag{60}$$

so the probability that, beginning with a vertex v of degree 2, we find k consecutive neighborhoods of size 2 is

$$\left((M^*(\mu) - \epsilon_1)^2 \right)^k = (M^*(\mu) - \epsilon_1)^{\left(\frac{1}{\ln M^*(\mu)} - \epsilon_0 \right) \ln n} = \Omega(n^{-(1-C_{\epsilon_0})}). \tag{61}$$

Hence, with probability $1 - o(1)$, we find at least 2 vertices u, v with chains of length k.

At this point, the neighborhoods at distance k from u and v will contain 2 unexposed endpoints each. We now consider these neighborhoods as "vertices" u', v' of unexposed degree 2, and invoke lemma 4.7 to derive an appropriate lower bound on $\delta(u',v')$. This completes the proof. ∎

The lower bound proof of Theorem 4.1 is now complete.

5 Applications

5.1 The Diameter of $\mathcal{G}_{n,p}$

Consider computing the diameter of the classical random graph $\mathcal{G}_{n,p}$, for $p = \frac{d}{m}$, where $d > 1$. It can be shown that degree distribution of $\mathcal{G}_{n,p}$ is w.e.h.p. k-convergent to the Poisson distribution $\pi_d(i) = \frac{e^{-d} d^i}{i!}$. Also, the Poisson distribution has the property that the residual distribution is the same as the original distribution, so $\mu_{\pi_d} = \pi_d$. Hence, let $\Delta(d) = \frac{2}{-\ln(M^*(\pi_d))} + \frac{1}{\ln(M_1(\pi_d))}$, so that, by Theorem 4.1, we have $\frac{\Delta(\mathcal{G}_{n,p})}{\ln n} \to \Delta(d)$ a.a.s. for $p = \frac{d}{n}$.

The p.g.f. for the Poisson distribution π_d has the simple expression $\psi_{\pi_d}(z) = e^{d(z-1)}$. The fixed point of this function is given by $z_{\pi_d} = \frac{-W(-de^{-d})}{d}$, where the *Lambert W-function* $W(z)$ is the principal inverse of $f(z) = ze^z$. This gives a closed-form expression for $\Delta(d)$:

$$\Delta(d) = \frac{2}{\ln -W(de^{-d})} + \frac{1}{\ln d}. \tag{62}$$

From equation (62), it can be shown that $\Delta(d) \ln d \to 3$ as $d \to 1$ and $\Delta(d) \ln d \to 1$ as $d \to \infty$, and it is a simple exercise to derive increasingly accurate asymptotic characterizations of $\Delta(d)$, as in

$$\Delta(d) = \frac{1}{\ln d} + \frac{2}{d} + O\left(\frac{\ln d}{d^2} \right) \text{ as } d \to \infty$$

and so on.

5.2 Finding a Shortest Path Quickly

The proof of Theorem 4.1 shows that one can quickly find a shortest path between a pair of vertices u, v in a random graph $\mathbf{G}_D$ using a simple algorithm. Specifically, we perform BFS starting from u until either

- a path connecting u and v is found;
- the BFS neighborhood is empty (so u and v are not connected);
- the BFS neighborhood reaches size $3m^{1/2} \ln m$.

Then we perform BFS from v until either a connecting path is found or the search terminates without finding a connecting path.

An alternate algorithm (which is a better heuristic in practice for general graphs) is one that performs the BFS search simultaneously from u and v, starting the search for each new level from the BFS neighborhood of smaller size. Assuming an adjacency-lists representation for the graph, the following corollary of Theorem 4.1 holds for either algorithm.

Corollary 5.1 *Let D be a degree sequence satisfying $\lambda_D \stackrel{1}{\to} \lambda$ where $M_1(\mu_\lambda) > 1$. Then a.a.s. for any connected pair of vertices u, v in the random graph $\mathbf{G}_D$, a shortest path connecting u and v can be found by a simple algorithm in time $O(m^{1/2} \ln m) = O(n^{1/2} \ln n)$.*

5.3 The Distance Distribution

For a graph $G = (V, E)$ let

$$\delta(G) = \{\delta(u, v) : u, v \in V\}$$

denote the multiset of distances between pairs of vertices in G, and note that $|\delta(G)| = n^2$, since there are n^2 pairs of vertices. The proof of Theorem 4.1 shows that almost all finite distances in $\delta(G)$ are very close to $M_1(\mu_\lambda) \cdot \ln n$. We state this more precisely in the following corollary.

Corollary 5.2 *Let D be a degree sequence satisfying $\lambda_D \stackrel{2}{\to} \lambda$ or $\lambda_D \stackrel{1}{\to} \lambda$ with $M_2(\lambda) = \infty$, and assume $M_1(\mu_\lambda) > 1$. Then, for any $\epsilon > 0$,*

$$\left| \left\{ \delta \in \delta(\mathbf{G}_D) : \delta < \infty \text{ and } \left| \tfrac{\delta}{\ln n} - M_1(\mu_\lambda) \right| > \epsilon \right\} \right| = o(n^2) \text{ a.a.s.}$$

APPENDIX

A Random Graphs and Probability Generating Functions

In this section we discuss various distributions related to random graphs which can be described using manipulations of probability generating functions. Recall that, by definition 2.2, for a distribution μ, the p.g.f. ψ_μ is given by

$$\psi_\mu(z) = \mathcal{E}[z^{\mathbf{X}_\mu}]$$

for $z \in [0, 1]$. Various characteristics of a random graph with a fixed degree sequence D satisfying $\lambda_D \to \lambda$ can be understood in terms of the p.g.f. of the limiting residual distribution $\mu = \mu_\lambda$. The p.g.f. also plays a key role in the theory of branching processes [2].

We begin by noting that the k'th derivative of the p.g.f. is given by

$$\psi_\mu^{(k)}(z) = \mathcal{E}\left[\mathbf{X}_\mu \cdot (\mathbf{X}_\mu - 1)\cdots(\mathbf{X}_\mu - k + 1)\cdot z^{\mathbf{X}_\mu - k}\right] = \mathcal{E}\left[(\mathbf{X}_\mu)_k \cdot z^{\mathbf{X}_\mu - k}\right],$$

and therefore

$$\frac{\psi_\mu^{(k)}(0)}{k!} = \mu(k) \tag{63}$$

and

$$\psi_\mu^{(k)}(1) = M_k(\mu). \tag{64}$$

Now, for any distribution μ, let

$$z_\mu = \min\{z \in [0,1] : \psi_\mu(z) = z\}. \tag{65}$$

denote the lowest fixed point of ψ_μ in the interval $[0,1]$, and note that $z_\mu < 1$ if and only if $M_1(\mu) > 1$ (see, for example [2]). For a distribution μ with $M_1(\mu) > 1$, we define distributions ξ_μ and ρ_μ, by giving the generating functions

$$\psi_{\xi_\mu}(z) = \frac{\psi_\mu(z \cdot z_\mu)}{z_\mu} \tag{66}$$

and

$$\psi_{\rho_\mu}(z) = \frac{\psi_\mu(z + (1-z)z_\mu) - z_\mu}{1 - z_\mu} \tag{67}$$

The result of Molloy and Reed [13] regarding the degree sequence of a random graph with the giant component removed can now be expressed as follows.

Theorem A.1 ([13]) *Let D be a degree sequence and assume $\lambda_D \xrightarrow{1} \lambda$ where $\mu = \mu_\lambda$ satisfies $M_1(\mu) > 1$. Consider the random graph $\mathbf{G}_D$, and let $\boldsymbol{\xi}$ denote the residual degree distribution of the graph which results from removing the largest connected component from $\mathbf{G}_D$.*

Then $\boldsymbol{\xi} \to \xi_\mu$ a.a.s.

Proof. From the formula of Molloy and Reed for the limiting degree distribution of the graph with the largest component removed (i.e. Theorem 2 of [13]), we derive

$$\boldsymbol{\xi}(i) \to \mu(i) z_\mu^{i-1} \tag{68}$$

a.a.s. for all i, and by equation (66), $\psi_{\xi_\mu}(z) = \sum_{i=0}^\infty z^i z_\mu^{i-1}\mu(i)$ therefore $\xi_\mu(i) = \mu(i)z_\mu^{i-1}$, so ξ_μ is in fact the limiting residual distribution described in equation (68).[2] ∎

The residual degree distribution of the 2-core of the random graph $\mathbf{G}_D$, as described in [8], can be expressed similarly.

Theorem A.2 ([8]) *Let D be a degree sequence and assume $\lambda_D \xrightarrow{k} \lambda$ where $\mu = \mu_\lambda$ satisfies $M_1(\mu) > 1$. Consider the random graph $\mathbf{G}_D$, and let $\boldsymbol{\rho}$ denote the residual degree distribution of the 2-core of $\mathbf{G}_D$.*

Then $\boldsymbol{\rho} \xrightarrow{k-1} \rho_\mu$ w.e.h.p.

[2] We note that this result can be strengthened to achieve k-convergence and/or a w.e.h.p. guarantee, but this is unnecessary for our purposes.

Proof. In Lemma 1 in the appendix of [8], the expected number of vertices of degree $i \geq 2$ at time t of the 2-core algorithm, assuming that the 2-core has not yet been found, is given as

$$\sum_{j=i}^{\infty} \lambda(j) \binom{j}{i} p(t)^i (1-p(t))^{j-i} \pm o(1),$$

where $p(t) = (\frac{m-t}{m})^{1/2}$, and [8] gives w.e.h.p. concentration. Also, the halting time $\boldsymbol{\tau}$ of the 2-core algorithm is shown to satisfy $\frac{\boldsymbol{\tau}}{m} \to 1 - (1-z_\mu)^2$ or, equivalently, $p(\boldsymbol{\tau}) \to 1 - z_\mu$, with w.e.h.p. concentration.

From these two results, we compute the limiting residual distribution at the halting time of the CM 2-core algorithm. First, let λ_2 denote the limiting degree distribution of the 2-core, so

$$\lambda_2(i) = \sum_{j=i}^{\infty} \lambda(j) \binom{j}{i} (1-z_\mu)^i z_\mu^{j-i} \tag{69}$$

for $i \geq 2$ (and by definition $\lambda_2(1) = 0$).

Let us denote corresponding residual distribution by $\mu_2 = \mu_{\lambda_2}$, and compute, for $i \geq 1$,

$$\begin{aligned}
\mu_2(i) = \frac{(i+1)\lambda_2(i+1)}{M_1(\lambda_2)} &= \frac{(i+1)}{M_1(\lambda_2)} \sum_{j=i+1}^{\infty} \lambda(j) \binom{j}{i+1} (1-z_\mu)^{i+1} z_\mu^{j-i-1} \\
&= \frac{(i+1)}{M_1(\lambda_2)} \sum_{h=i}^{\infty} \lambda(h+1) \binom{h+1}{i+1} (1-z_\mu)^{i+1} z_\mu^{h-i} \\
&= \frac{(i+1)}{M_1(\lambda_2)} \sum_{h=i}^{\infty} \lambda(h+1) \frac{(h+1)!}{(h-i)!(i+1)!} (1-z_\mu)^{i+1} z_\mu^{h-i} \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \sum_{h=i}^{\infty} \mu(h) \binom{h}{i} (1-z_\mu)^i z_\mu^{h-i}.
\end{aligned}$$

In particular, note the similarity between the expression above for $\mu_2(i)$ and equation (69) for $\lambda_2(i)$. Note also that $\mu_2(0) = 0$.

We now show that $\psi_{\mu_2} = \psi_{\rho_\mu}$, where ρ_μ is defined in equation (67). First, we compute

$$\begin{aligned}
\psi_{\mu_2}(z) = \mathcal{E}[z^{\mathbf{X}_{\mu_2}}] &= \sum_{i=0}^{\infty} \mu_2(i) z^i \\
&= \sum_{i=1}^{\infty} \mu_2(i) z^i \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \sum_{i=1}^{\infty} \sum_{h=i}^{\infty} \mu(h) \binom{h}{i} (1-z_\mu)^i z_\mu^{h-i} z^i \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \sum_{i=1}^{\infty} \sum_{h=i}^{\infty} \mu(h) \binom{h}{i} (z(1-z_\mu))^i z_\mu^{h-i} \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \sum_{h=1}^{\infty} \mu(h) \sum_{i=1}^{h} \binom{h}{i} (z(1-z_\mu))^i z_\mu^{h-i} \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \sum_{h=0}^{\infty} \mu(h) \Big((z(1-z_\mu)+z_\mu)^h - z_\mu^h \Big) \\
&= \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \Big(\psi_\mu \left(z(1-z_\mu)+z_\mu \right) - \psi_\mu(z_\mu) \Big).
\end{aligned}$$

Since z_μ is a fixed point of ψ_μ, this yields

$$\psi_{\mu_2}(z) = \frac{M_1(\lambda)(1-z_\mu)}{M_1(\lambda_2)} \Big(\psi_\mu \left(z(1-z_\mu)+z_\mu \right) - z_\mu \Big). \tag{70}$$

In order to demonstrate that equations (67) and (70) are equal (and therefore $\psi_{\rho_\mu} = \psi_{\mu_2}$), it suffices to show that

$$\frac{M_1(\lambda)}{M_1(\lambda_2)} = (1-z_\mu)^{-2}.$$

While this can be achieved algebraically by a computation similar to what is shown above, we give a more intuitive, if less formal, argument. Since λ gives the initial degree distribution and λ_2 gives the degree distribution of the 2-core, then $m = M_1(\lambda) \cdot n$ gives the number of endpoints in the original graph, and $M_1(\lambda_2) \cdot n$ gives the number of endpoints in the 2-core. It follows that the fraction of endpoints in the 2-core is given by

$$\frac{M_1(\lambda_2)}{M_1(\lambda)}.$$

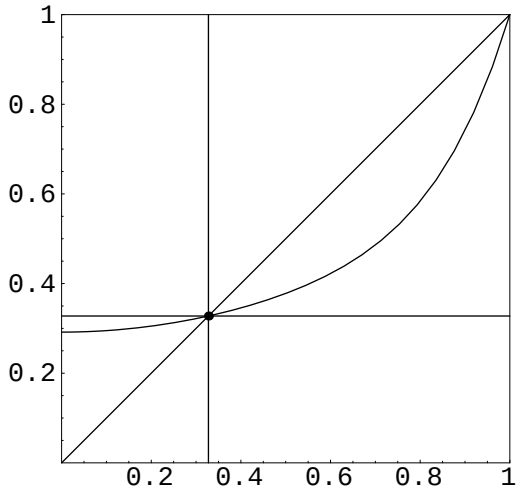
As noted above, the halting time of the 2-core algorithm concentrates about $m(1-(1-z_\mu)^2)$, and since each time step exposes a single endpoint, we have

$$\frac{M_1(\lambda_2)}{M_1(\lambda)} = (1-z_\mu)^2.$$

Therefore $\mu_2 = \rho_\mu$, and we conclude that $\boldsymbol{\rho} \xrightarrow{k-1} \rho_\mu$ w.e.h.p. ∎

Next, we give a visual interpretation of the functions ψ_{ξ_μ} and ψ_{ρ_μ}, and their relation to the p.g.f. ψ_μ. For a typical distribution μ, let us examine a plot of the p.g.f. ψ_μ as shown below; here

the x-axis is z and the y-axis is $\psi_\mu(z)$.



The value z_μ corresponds to the lowest fixed point in ψ_μ, as shown in the plot. If we "drag" the point (z_μ, z_μ) to the upper-right-hand corner $(1,1)$, and "stretch" the lower-left-hand region of the plot to cover the entire $[0,1] \times [0,1]$ square, what results is the plot of the p.g.f. ψ_{ξ_μ}. On the other hand, if we drag the point (z_μ, z_μ) diagonally towards the origin $(0,0)$, and stretch the upper-right-hand region of the plot, the result is the plot of ψ_{ρ_μ}.

We can verify algebraically that the "stretching" described above does not affect first derivatives, and therefore

$$\psi'_{\xi_\mu}(1) = \psi'_{\rho_\mu}(0) = \psi'_\mu(z_\mu)$$

and

$$\psi'_{\rho_\mu}(1) = \psi'_\mu(1).$$

Now, let us define

$$M^*(\mu) = \psi'_\mu(z_\mu), \tag{71}$$

and by equations (63) and (64), we have the following proposition.

Proposition A.3 *For any distribution μ satisfying $M_1(\mu) > 1$,*

1. $M_1(\xi_\mu) = \rho_\mu(1) = M^*(\mu)$;
2. $M_1(\rho_\mu) = M_1(\mu)$.

This proposition can be interpreted intuitively as follows. Recall that, if μ is the limiting residual degree distribution of a random graph $\mathbf{G}_D$, then ρ_μ gives the limiting residual distribution of the 2-core of $\mathbf{G}_D$, and ξ_μ gives the limiting residual distribution of the subgraph of $\mathbf{G}_D$ with the giant component removed. Hence, statement 1 of proposition A.3 shows that the fraction of endpoints of residual degree 1 in the 2-core of $\mathbf{G}_D$ is asymptotically equal to the average residual degree of the subgraph with the giant component removed. Statement 2 of proposition A.3 shows that the average residual degree of the 2-core of $\mathbf{G}_D$ is asymptotically equal to the average residual degree of $\mathbf{G}_D$ itself.

Finally, we note that the distributions ρ_μ and ξ_μ also arise in a certain decomposition of a supercritical μ-branching process; this decomposition, as well as a more comprehensive discussion of probability generating functions, can be found in [2].

References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 171–180, 2000.

[2] K. B. Athreya and P. E. Ney. *Branching Processes.* Springer, 1972.

[3] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316, 1980.

[4] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[5] F. Chung and L. Lu. The diameter of random sparse graphs. *Advances in Applied Math.*, 26:257–279, 2001.

[6] C. Cooper. The cores of random hypergraphs with a given degree sequence. Manuscript, 2003.

[7] D. Fernholz and V. Ramachandran. The giant k-core of a random graph with a specified degree sequence. Manuscript, 2003.

[8] D. Fernholz and V. Ramachandran. Cores and connectivity in sparse random graphs. Technical Report UTCS TR04-13, Department of Computer Sciences, University of Texas, Austin TX 78712, 2004. `http://www.cs.utexas.edu/ftp/pub/techreports/INDEX/html/Index.2004.html`.

[9] O. Kallenberg. *Foundations of Modern Probability.* Springer-Verlag, 2002.

[10] L. Lu. The diameter of random massive graphs. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 912–921, 2001.

[11] B. D. McKay and N. C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11:369–382, 1991.

[12] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6:161–179, 1995.

[13] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7:295–305, 1998.

[14] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k-core in a random graph. *J. of Combinatorial Theory, Series B*, 67:111–151, 1996.