General expression for the component size distribution in infinite configuration networks

Ivan Kryven

University of Amsterdam, PO box 94214, 1090 GE, Amsterdam, The Netherlands

In the infinite configuration network the links between nodes are assigned randomly with the only restriction that the degree distribution has to match a predefined function. This work presents a simple equation that gives for an arbitrary degree distribution the corresponding size distribution of connected components. This equation is suitable for fast and stable numerical computations up to the machine precision. The analytical analysis reveals that the asymptote of the component size distribution is completely defined by only a few parameters of the degree distribution: the first three moments, scale and exponent (if applicable). When the degree distribution features a heavy tail, multiple asymptotic modes are observed in the component size distribution that, in turn, may or may not feature a heavy tail.

PACS numbers: 64.60.aq, 02.10.Ox, 89.75.Da

Keywords: networks; degree distribution; connected components; heavy tail; power law; giant component

I. INTRODUCTION

Random graphs provide models for complex networks, and in many cases, real-world networks has been accurately described by such models[1][4]. Within the scope of random graph models one finds: Erdős-Rényi model, Barabási-Albert model[1], node copying model[5], small world network[6], configuration network[7][8] and many others. In the configuration network N nodes are assigned pre-defined degrees. The edges connecting these nodes are then considered to be random, and every distinct configuration of edges that satisfies the given degree sequence is treated as a new instance of the network in the sense of random graphs. Interesting properties emerge when the number of nodes, N, approaches infinity, or at the so-called thermodynamic limit[9][10]. In this case, the infinite degree sequence, which provides the only input information for the model, is equivalent to the frequency distribution of degrees, u(k), k = 1, 2, …, i.e. the probability that a randomly chosen node has degree k.

Component-size distribution, w(n), denotes probability that a randomly chosen node is part of a connected component of finite size n. Connected components in the infinite configuration network can be of finite or infinite size. Molloy and Reed[11] showed that if an infinite component exists then it is the only infinite component with probability 1. Hence, the infinite component is referred to as the giant component.

Depending upon a specific context behind the network, the component size distribution may summarise an important feature of the modelled system. In polymer chemistry, for example, the infinite configuration network is used as a toy model for hyper-branched and cross-linked polymers. In this context, the component size distribution predicts viscoelastic properties of the material while the emergence of the giant component is interpreted as a phase transition from liquid to solid state of the soft matter[11][12]. Since connected components are closely related to clusters in bond percolation processes, the distribution of component sizes can be used to model outbreaks for SIR epidemiological processes[2]. In linguistics, component size distribution of the sentence similarity graph is an important tool when studying structure of natural languages[13]. This brief list of application cases is far from being exhaustive. Despite the vast applications, the empirical component size distribution is hard to measure precisely unless the whole topology of
the network is known. On another hand, empirical observations on the degree distribution, \( u(k) \), are much easier to perform.

Ref. [11] provides an elegant criterion that connects moments of \( u(k) \) to the fact that the network contains the giant component. A somewhat deeper question further in this direction reads: providing \( u(k) \) is given, what is the component size distribution, \( w(n) \)? In Ref. [11] Newman et al. showed that the component size distribution can be recovered by a numerical algorithm that involves solving a fixed point problem followed by an inversion of a generation function. Such algorithm demonstrates that indeed \( u(k) \) and \( w(n) \) can be put into a correspondence, however, it becomes computationally infeasible for large values of \( n \). This numerical issues aries due to ill-posedness of the numerical generating function inversion.

On another hand, the component size distribution has been analytically resolved only for a limited number of partial cases of \( u(k) \) [15]. Within the scope of analytically solvable cases, only the Yule-Simon degree distribution features a heavy tail, that is to say it decays proportionally to an algebraic function, \( n^{-\beta} \), \( \beta > 0 \) at large \( n \). At the same time, the heavy-tailed (or scale-free) distributions are commonly observed in the empirical data collected from many real-world networks [16–19]. Empirically observed exponents vary in a broad range. Some studies report degree exponents that are as small as \( \beta = 0.81 \) in the case of the Internet topology [20] and \( \beta = 1 \) in social networks [21]. On the opposite side of this spectrum, one finds exponent \( \beta = 5 \) in the generalisation of preferential attachment model [22].

The only asymptotic analysis available for component size distribution in the configuration network states that for large \( n \), \( w(n) \) is either proportional to \( n^{-3/2} \) or to \( e^{-Cn} \), where \( C > 0 \) is a constant [14]. The current paper uncovers new asymptotic modes for \( w(n) \) that emerge only when the degree distribution features a heavy tail.

The paper shows that for an arbitrary \( u(k) \), \( w(n) \) can be expressed as a finite sum. In practice, this sum can be stably computed up to the machine precision in the cost of \( O(n^2) \) multiplicative operations. Finally, the paper discusses how a finite cutoff introduced in the degree distribution reflects on the distribution of component sizes.

II. COMPONENT SIZE DISTRIBUTION BY LAGRANGE INVERSION

It has been noticed that all components in the infinite configuration model are locally tree-like. Using this fact as a departure point, Newman et al. [14] showed that the degree distribution can be put into a correspondence to the component size distribution by applying the generating-function (GF) formalism. Here, by a GF of \( u(k) \), \( \sum_{k=0}^{\infty} u(k) = 1 \), we refer to the series,

\[
U(x) = \sum_{k=0}^{\infty} u(k)x^k, \ x \in \mathbb{C}, \ |x| \leq 1.
\] (1)

According to the approach presented in Ref. [11], the generating function for the component size distribution, \( W(x) \) is found as a solution of the following system of functional equations,

\[
W(x) = xU'[W_1(x)],
\] (2)

\[
W_1(x) = xU_1[W_1(x)].
\] (2)

where \( U(x) \) is the GF of \( u(k) \), and \( U_1(x) \) is the GF for the excess degree distribution

\[
u_1(k) = \frac{k+1}{\mu_1} u(k+1),
\] (3)

where \( \mu_1 = \sum_{k=1}^{\infty} ku(k) \). Similarly to combinatorial treecounting problems, Eq. (2) can be solved by applying the Lagrange inversion formula [23]. The original formulation of the Lagrange inversion principle is as follows. Suppose, \( A(X), R(x) \), are such formal power series that \( A(x) = xR[A(x)] \) then for an arbitrary formal power series \( F(x) \), the coefficient of power series \( F[A(x)] \) at \( x^n \) reads as,

\[ [x^n]F[A(x)] = \frac{1}{n} [x^{n-1}]F'(t)R^n(t), \ n > 0. \] (4)

Here \( [t^{n-1}] \), as being the inverse operation to the GF transform [1], refers to the coefficient at \( t^{n-1} \) of the corresponding power series. By substituting \( A(x) = W_1(x) \), \( F(x) = U(x) \) and applying Eq. (2) one transforms the left hand side of (1), \( [x^n]F[A(x)] = [x^n]U[W_1(x)] = [x^{n-1}]W(x) = w(n-1) \). Further on, the right hand side
of \([4]\) is transformed by substituting \(R(x) = U_1(x)\) and realising that, according to the definition \([9]\), \(U'(x) = \mu_1U_1(x)\). Now, we are ready to write an expression for \(w(n)\), even though we have no explicit expression for generating function \(W(x)\) itself,

\[
   w(n) = \frac{1}{n-1}[t^{n-2}]U'(x)U_1(x)^{n-1} = \frac{\mu_1}{n-1}[t^{n-2}]U_1(x)^n, \quad n > 1.
\]

(5)

A similar equation was also derived in Ref.\([15]\) by means of different reasoning. In principle, Eq. \([5]\) provides enough information to analytically recover the component size distribution for a few special cases of the degree distribution\([15]\). In practice, however, the main difficulty when applying Eq. \([5]\) is that the equation employs the inverse GF transform, \([t^n]\), which limits the choices one has when searching for an exact solution or performing numerical computations. With this in mind, one may rewrite \([3]\) so that the new expression does not involve the GF concept at all. It turns out that the only reason why Eq. \([5]\) utilises the GF formalism is that it provides means for convolution power.

The convolution of two distributions, \(f(k) \ast g(k)\), \(k > 0\) is defined as a binary multiplicative operation,

\[
   f(k) \ast g(k) = \sum_{i+j=k} f(i)g(j),
\]

where the summation is performed over all non-negative ordered couples \(i, j\) that sum up to \(k\). This sum contains exactly \(k+1\) of such couples. In this paper, the order of operations is chosen in such a way that the pointwise multiplication precedes convolution, for instance, \(f(k) \ast kg(k) = f(k) \ast [kg(k)]\). The convolution can be inductively extended to the \(n\)-fold convolution, or the convolution power,

\[
   f(k)^n = f(k)^{n-1} \ast f(k),
\]

(6)

where \(f(k)^n \equiv 1\) by the definition. It can be shown that the convolution power can be expanded into a sum of products,

\[
   f(k)^n = \sum_{\sum_{k_i \geq 0} k_i = n} \prod_{i=1}^{n} f(k_i).
\]

(7)

The convolution has a peculiar property in respect to the GF transform. If \(F(x), G(x), U(x)\) are GFs for \(f(k), g(k),\) and \(u(k) = f(k) \ast g(k)\) then \(U(x) = F(x)G(x)\). Furthermore, if \(U(x)\) is GF for \(u(k)\) then \(U(x)^n\) generates \(u(k)^n\). By exploiting this relation one immediately reduces Eq. \([6]\) to,

\[
   w(n) = \begin{cases} 
   \frac{\mu_1}{n-1}u_1^n(n-2), & n > 1, \\
   u(0) & n = 1.
   \end{cases}
\]

(8)

Here, the value of \(w(0)\) is derived directly from the formulation of the problem: nodes with degree zero are also components of size one. This simple equation is ready to be used: by combining \([8]\) and the definitions \([3],[7]\), one may directly expresses the values of the component size distribution in terms of \(u(k)\) : for \(n > 1,

\[
   w(n) = [ku(k)]^n(2n-2), \quad n > 1.
\]

(9)

For example, first five values of \(w(n)\) read as,

\[
   w(1) = u(0),
   w(2) = \frac{1}{\mu_1}u(1)^2,
   w(3) = \frac{3}{\mu_1}u(1)^2u(2),
   w(4) = \frac{4}{\mu_1}u(1)^2[2u(2)^2 + u(1)u(3)],
   w(5) = \frac{5}{\mu_1}u(1)^2[4u(2)^3 + 6u(1)u(2)u(3) + u(1)^2u(4)].
\]

The number of terms in this expansion increases rapidly with \(n\). That said, the formula \([8]\) can be easily readjusted for numerical computations. Namely, one can use Fast Fourier Transform (FFT) to compute the convolution powers, \(u_1^n(k) = F^{-1}[F[u_1(k)]^n]\). In this case, \(O(N^2)\) multiplicative operations is sufficient to compute all values of \(w(n), n \leq N\). Alternatively, if \(w(n-1)\) is known \(w(n)\) can be found in the cost of \(O(n \log n)\).

Besides FFT, there are algorithms that are specifically designed for fast approximation of convolution powers, such as projection onto basis functions that are invariant under convolution\([24]\).

Analytic formulas for convolution powers (sometimes also referred to as compositas \([25]\)), were covered by literature for many elementary functions\([26],[27]\). Convolution powers of \(u_1(k)\) can also be found analytically by
applying discrete functional transforms, for instance Z-transform and discrete Fourier transform. A few examples of such results are given in Table I. Focusing on one of them, the first curve in Figure 1 demonstrates that both analytical and numerical results for the exponential degree distribution coincide.

| Degree Distribution | Component Size Distribution |
|---------------------|----------------------------|
| Exponential distribution | $Ce^{-\lambda k}$ |
| Geometric distribution | $(1-p)^{k-1}p$ |
| Binomial distribution | $(\binom{k_{\text{max}}}{k})(1-c)^{k_{\text{max}}-k}c^k$ |

TABLE I. Exact expressions for component size distributions in configuration network as evaluated with Eq. (8) via Z-Transform.

![Graph](image)

FIG. 1. Examples of component size distributions (solid lines) that feature fast (1) and slow (2) convergence to their asymptotes (dashed lines). Both asymptotes are covered by Case A, Table I) $u(k) = Ce^{-1.05k}$, all three: the analytical expression (see Table I), numerical values (according to Eq. (8)) and the asymptote practically coincide. 2) $u(k)$ is non-zero in three points $u(1) = 0.97$, $u(2) = 0.015$, $u(10) = 0.015$; the component size distribution features oscillations before it converges to the asymptote.

### III. ASYMPTOTIC ANALYSIS

The format of Eq. (8), naturally suggests a straightforward way to perform an asymptotic analysis for $n \to \infty$. One may view $u(k)$ as a probability mass function PMF (or alternatively discrete probability density function) of some discrete random variables $k$. Recall the following property of convolution powers: if i.i.d. random variables $k_i$ have PMF $u_1(k)$ then $u_{1}^{n}(k)$ gives the PMF for the sum $k_1 + k_2 + \cdots + k_n$. The central limit theorem (CLT) gives an estimate for this sum as $n \to \infty$, and the idea is now to obtain the asymptotes of $w(n)$ by applying CLT to the definition (8).

#### A. Light-tailed degree distributions

First, let us assume that distribution $u(k)$ decays faster than algebraically, that is,

$$u(k) = o(k^{-\beta}), \quad \beta > 2, \quad k \to \infty,$$

which is also equivalent to $u_1(k) = o(k^{-\beta+1})$. Then according to CLT, $u_1^{n}(k)$ approaches the normal distribution, $u_1^{n}(k) \xrightarrow{d} (\sqrt{n}\sigma)^{-1} N\left(\frac{k-M}{\sqrt{n}\sigma},0,1\right)$, when $n \to \infty$, where $M = \sum_{k=1}^{\infty} k u_1(k)$ and $\sigma^2 = \sum_{k=1}^{\infty} (k-M)^2 u_1(k) < \infty$ denote the mean value and variance of $u_1(k)$. The normal distribution can now replace $u_1^{n}(k)$ in (8), which yields the asymptote for the component size distribution,

$$w(n) \sim \frac{\mu_1 e^{-[(n-M)-2]^2/2n\sigma^2}}{(n-1) \sqrt{2\pi n\sigma^2}} \text{ as } n \to \infty. \quad (11)$$

Quantities $M, \sigma^2$ are directly expressible in terms of moments of degree distribution $u(k)$,

$$M = \sum_{k=1}^{\infty} k u_1(k) = \frac{1}{\mu_1} \sum_{k=1}^{\infty} (k^2 - k) u(k) = \frac{\mu_2 - \mu_1}{\mu_1},$$

$$\sigma^2 = \sum_{k=1}^{\infty} (k-M)^2 u_1(k) = \frac{1}{\mu_1} \sum_{k=1}^{\infty} k(k-M-1)^2 u(k) = \frac{\mu_3 \mu_1 - \mu_2^2}{\mu_1^2},$$

where

$$\mu_i = \mathbb{E}[k^i] = \sum_{k=1}^{\infty} k^i u(k), i = 1, 2, \ldots$$
Finally, substituting the expressions \( \mu_1 \) into (11) and gives the final version of the asymptote,

\[
w(n) \sim \frac{\mu_1^2 n^{-3/2} e^{(\mu_2 - 2\mu_1)^2 n}}{\sqrt{2\pi(\mu_1\mu_3 - \mu_2^2)}}, \quad n \to \infty. \tag{13}
\]

Two examples of component size distributions that converge with various rates to their asymptotes are given in Figure 1. Peculiarly, the only information on \( u(k) \) that is contained in the asymptote definition (13) is the first three moments \( \mu_1, \mu_2, \mu_3 \). Furthermore, depending upon the value of \( \theta = \mu_2 - 2\mu_1 \), the asymptotic expression (13) switches between the two modes: it either decays exponentially as \( O(e^{-Kn}) \), when \( \theta \neq 0 \), or it decays as an algebraic function, \( O(n^{-3/2}) \), when \( \theta = 0 \) (see also Table 1 for \( \mu_2 = 0 \) and the case \( \theta = 0 \) (see also Table 1 Case A). The last equality is the well-known giant component criterion,

\[
\mu_2 - 2\mu_1 = 0. \tag{14}
\]

The criterion (14) was obtained by Molloy and Reed\[11\] by means of a different reasoning. In Ref \[11\], the authors prove that \( \theta > 0 \) implies existence of the giant component in the configuration network, whereas \( \theta < 0 \) implies non-existence of this component. In Ref. \[14\], it was hypothesised that the \(-3/2\) exponent is universal and must hold for all degree distributions at the critical point \( \theta = 0 \). We will see now that when the condition (14) fails to hold, distinct from \(-3/2\) exponents may also appear in the asymptotic of \( w(n) \).

### B. Heavy-tailed degree distributions

Suppose that, on the contrary to the condition (14), degree distribution \( u(k) \) features a heavy tail,

\[
u(k) \sim sk^{-\beta}, \quad \beta > 2, \quad k \to \infty, \tag{15}\]

which is equivalent to \( u_1(k) \sim sk^{-\alpha-1}, \quad \alpha = \beta - 2 > 0, \quad k \to \infty. \) It turns out that exponent \( \alpha \) and the scale \( s \), together with the moments \( \mu_1, \mu_2, \mu_3 \) provide enough information to generalise the asymptote (13) for the case of heavy-tailed degree distributions. Suppose \( 0 < \alpha \leq 2 \). In terms of \( u(k) \) moments this condition casts out as \( \mu_3 = \infty \). As follows from Gnedenko and Kolmogorov’s generalisation of CLT\[28\] the mass density distribution for \( u_1^n(k) \) approaches the stable law,

\[
u_1^n(k) \to \frac{1}{\gamma(n)} G^A \left( \frac{k - \mu(n)}{\gamma(n)}, \alpha, 1 \right), \quad n \to \infty. \tag{16}\]

Here, we use the notation of Uchaikin & Zolotarev\[29\] which includes: exponent parameter \( \alpha \), the location parameter

\[
\mu(n) = \begin{cases} \frac{n^{2-\mu_1}}{\mu_1}, & \alpha > 1, \\ sn \ln n, & \alpha = 1, \\ 0, & 0 < \alpha < 1, \end{cases} \tag{17}
\]

and the scale parameter

\[
\gamma(n) = \begin{cases} \sqrt{s n \ln n}, & \alpha = 2, \\ \sqrt{\pi s} \left[ 2 \Gamma(\alpha) \sin \frac{\pi \alpha}{2} \right]^{-1/\alpha} n^{1/\alpha}, & \alpha \in (0,1) \cup (1,2), \\ \frac{sn}{2}, & \alpha = 1. \end{cases} \tag{18}
\]

No general analytical expression is known for \( G^A(x, \alpha, 1) \), and the stable law is defined via its Fourier transform

\[
\mathcal{F}[G^A(x, \alpha, 1)] = \begin{cases} e^{-x^{\alpha} - it \tan \frac{\pi \alpha}{2}}, & \alpha \in (0,1) \cup (1,2), \\ e^{-x^{\alpha} + it \frac{\pi \alpha}{2}}, & \alpha = 1. \end{cases} \tag{19}
\]

Consider the case when \( 1 < \alpha < 2 \). According to (16) the point in which the stable law is evaluated, \( x(n) = \frac{n-\mu(n)}{\gamma(n)} \), approaches positive or negative infinities depending upon the sign of \( \theta = \mu_2 - 2\mu_1 \). Indeed, as \( n \to \infty \),

\[
x(n) \to \begin{cases} +\infty, & \theta < 0, \\ 0, & \theta = 0, \\ -\infty, & \theta > 0. \end{cases} \tag{20}
\]

For these values of \( \alpha \), function \( G^A(x, \alpha, 1) \) is non-zero on \((0,\infty)\). If \( x(n) \to \infty \), the function features an algebraic decay, whereas if \( x(n) \to -\infty \) the decay is exponential. Therefore, the limiting value switching that takes place in (20) may reflect on the asymptotic behaviour of \( u_1^n(n) \). To give a precise answer one has to consider series expansions of \( G^A(x, \alpha, 1) \) around the points of interest, \( x \in \{-\infty, 0, +\infty\} \). We use here the leading terms
of these series[29],

\[ G^A(x, \alpha, 1) = \begin{cases} 
\Gamma\left(1 + \frac{1}{\alpha}\right) \sin \frac{\pi}{\alpha} + O(x) , & x \to 0 \\
\Gamma(\alpha + 1) x^{-\alpha - 1} + O(x^{-2\alpha - 1}) , & x \to \infty ,
\end{cases} \]

\[ e^{-(\alpha - 1)(\frac{x}{\pi})^\alpha (\frac{\alpha}{\pi})^{1/2} (1 + \frac{1}{\alpha})} \sqrt{2\pi \alpha (1 - \alpha)} \left[1 + O(x^{1-\alpha})\right] , 
\]

By replacing the expression for the limiting distribution [16] with the leading terms given in (21) one obtains the asymptotes for (8). This time, the asymptote has three modes: depending upon the value of \( \theta \), it either features a heavy tail with exponent \(-\alpha - 1\), a heavy tail with exponent \( -\frac{1}{\alpha} - 1 \), or an exponential decay, as shown in Table II, Case D. A few examples of such asymptotic modes for a heavy-tailed degree distribution

\[ u(k) = \begin{cases} 
C & k = 1 , \\
s(\beta - 2) k^{-\beta} & k > 1 
\end{cases} \]

are given in Fig. 2. The degree distribution (22) is defined by two parameters: exponent \( \beta \) and scale \( s \); whereas the constant \( C \) is such that the total probability is normalised, \( \sum_k u(k) = 1 \).

When \( \alpha = 2 \), the behaviour of \( \frac{n - \mu(n)}{\gamma(n)} \), \( n \to \infty \) is identical to [20], but the expression for \( \gamma(n) \) is different and the series expansions [21] lead to somewhat different asymptotes, see Table II Case C.

According to the definition [17], the location parameter vanishes, \( \mu(n) = 0 \), when \( \alpha < 1 \). In this case, \( x(n) = \frac{n}{\gamma(n)} \to 0 \) as \( n \to \infty \), and only one asymptotic mode is possible for \( w(n) \). Stable law \( G^A(x, \alpha, 1) \) is supported on \( (0, \infty) \), and we make use of the series expansion around \( x \to 0^+ \),

\[ G^A(x, \alpha, 1) = e^{-(1-\alpha)\left(\frac{x}{\pi}\right)^\alpha \left(\frac{\alpha}{\pi}\right)^{1/2} (1 + \frac{1}{\alpha})} \sqrt{2\pi \alpha (1 - \alpha)} \left[1 + O(x^{1-\alpha})\right] , \]

which when plugged in [16] yields faster than algebraic decay of the component size distribution, see Table II, Case F. Due to the parametrisation scheme for the stable law, the point \( \alpha = 1 \) needs to be considered separately. In this case, \( \frac{n - \mu(n)}{\gamma(n)} \to \infty \) when \( n \to \infty \), and we utilise the leading term of the series expansion,

\[ G^A(x, \alpha, 1) = \frac{1}{\sqrt{2\pi e}} e^{-\frac{x^2}{2}} e^{\gamma(n)} \left[1 + O(e^{1-x})\right] , \]

which admits one sub-algebraic asymptotic mode for \( w(n) \) as shown in Table II Case E. This case is special in that the stable law \( G^A(x, \alpha, 1) \) is supported on \( x \in (-\infty, \infty) \), but asymptotically, \( \frac{1}{\gamma(n)} G^A\left(\frac{n - \mu(n)}{\gamma(n)}, \alpha, 1\right) \) always tends to \(-\infty\) for large \( n \). At the same time, if for small \( n \) the point \( x(n) = \frac{n - \mu(n)}{\gamma(n)} \) stays on the positive half-axis where (24) does not provide correct description for \( G^A(x, \alpha, 1) \), the convergence to the asymptote will be slow. In other words, there is an intermediate asymptote that the component size distribution can be approximated with, before it eventually switches to Eq. (24). This switching point is given by such \( n_0 \) that \( x(n) \) changes the sign from \( '+1' \) to \( '-1' \), i.e. when \( n \) becomes greater then \( n_0 \). By solving \( x(n_0) = 0 \), one obtains \( n_0 = e^{\pi} \), which means that in principle, the switching between the intermediate and the final asymptotes may be indefinitely postponed if \( s \) is small enough. The intermediate asymptote itself is deduced from the
leading term of the stable law expansion at ∞, that is
\[ G^A(x, 1, 1) = \frac{2}{\pi} x^{-2} + O(x^{-3}). \]
After the substitutions one obtains,
\[ w(n) \simeq \frac{\mu_1 s}{(s \log n - 1)^2} n^{-2}, \quad \frac{1}{s} \gg 0, \ n < e^{\frac{1}{s}}, \ \alpha = 1. \]
As illustrated in Fig. 3, similar considerations are also valid for the case \( 0 < \alpha < 1 \), where
\[ w(n) \simeq \mu_1 s \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)} n^{-\alpha - 1}, \quad \frac{1}{s} \gg 0, \ 0 < \alpha < 1. \]
When occurs, such switching has a practical importance when dealing with empirically observed component size data. Indeed, it may happen that one observes only the intermediate asymptote and not the final one due a small number of samples at the tail of the component size distribution. For instance, the second curve in Fig. 3 does feature an exponential decay at infinity, but if one limits the data points to \( n < 10^6 \), the component size distribution will seem to be a heavy-tailed one.

![FIG. 3. Component-size distributions (solid line) corresponding to degree distributions with exponent \( \beta = 2.6 \) (\( \alpha = 0.6 \)). In this case, the component size distributions can not feature a heavy tail, however, depending upon the scale parameter \( s \) a transient asymptote with exponent \( -1.6 \) (dashed line) emerges: 1) \( s = 8.3 \cdot 10^{-2} \), fast convergence to the exponential asymptote. 2) \( s = 8.3 \cdot 10^{-5} \), the distribution transiently follows what seems to be a heavy tail for \( n < 10^6 \), whereas for larger \( n \) the theory predicts no heavy tail.](image)

![FIG. 4. Examples of component size distributions (solid lines) that are associated with heavy-tailed degree distributions with \( \beta = 6 \) (\( \alpha = 4 \)). The dashed lines represent the asymptotes in accordance with Case B in Table II. Depending on the sign of \( \theta \), three asymptotic modes are distinguished: 1) \( s = 1.93 : \theta = -0.8, 1a \) \( s = 9.42 : \theta = -0.027, 2) s = 9.69 : \theta = 1.6 \cdot 10^{-7}, 3) s = 10.05 : \theta = 0.038. \) When \( \theta \) is a small negative number (curve 1a), \( w(n) \) first decays as \( n^{-3/2} \) but eventually switches to asymptote \( n^{-\alpha - 1} = n^{-5}. \)](image)

Finally, we consider the case when the condition (15) holds for \( \beta > 4 \) : even though \( u_1(k) \) has finite mean \( M \) and variance \( \sigma^2 \) it also features a heavy tail. Again, as \( n \to \infty \), \( x(n) = \frac{n - \mu(n)}{\sigma(n)} \) features the limiting values that are defined by the sign of \( \theta \), see Eq. (20). One would expect that since \( \sigma^2 \) is finite, this case should be also well approximated with (13). This is indeed the case for \( x(n) \approx 0 \). However, large deviations from zero \( x(n) \gg 0 \) do not follow Gaussian statistics (30) (31), and we approximate \( u_1^n(k) \) with the Pareto stable law \( u_1^n(k) \to \frac{1}{\sigma(n)} G^P\left(\frac{k - \mu(n)}{\sigma(n)}, \alpha\right), n \to \infty. \) It turns out that \( G^P(x, \alpha) \) behaves as the normal distribution for \( x < C \), where \( C \) is a finite positive constant, but features a heavy tail with the same exponent as \( u_1(k) \) when \( x \to \infty \), see Ref. (30). Thus, when \( \theta \geq 0 \) the component size distribution features asymptotic modes as in (13), while when \( \theta < 0 \) it features a heavy tail with exponent \( -\alpha - 1 \), see Table II Case B. Interestingly, when \( \theta \) is a small negative number, \( w(n) \) transiently follows one asymptote and then switches to the other as demonstrated in Figure 3. If there is a process that continuously changes the degree distribution so that \( \theta \) progresses from being negative to positive, the exponent of the associated component size distribution will jump from the sub-critical branch, at \( \theta < 0 \), to the critical one at \( \theta = 0 \). An example
of such transition between two power-law modes is given in Figure 4 where a component size distribution switches between power laws with exponents $\eta = 3$ and $\eta = 1.5$.

IV. DISCUSSION AND CONCLUSIONS

The broad generality of the results obtained in the previous section is achieved due to the fact that the configuration networks are locally tree-like and have vanishing probability of clustering in the thermodynamic limit, which allows one to benefit from the available in analytic combinatorics tools. Eq. (8), that was analysed in the previous section, connects the degree distribution in a configuration network to the distribution of sizes for connected components. The main conclusion one may draw from this equation is that the convolution power provides a smoothing effect. This means that all points of $u(k), k = 1, \ldots, \infty$ have a significant contribution to the definition of $w(n)$, but as $n$ increases, the system ‘forgets’ the exact shape of the degree distribution and the component size distribution tends to the asymptote, that is defined by only a few parameters. The only information that is still preserved at the limit $n \to \infty$ is the first three moments of the degree distribution if such does not feature a heavy tail, see for example Fig. 1. If $u(k)$ does feature a heavy tail then the information that characterises the tail becomes also important: that is the scale parameter $s$ and the exponent $\beta$. Depending upon the values of these parameters, many asymptotical modes exist.

The expression for the asymptote is framed in terms of small deviation statistics for a sum of random variables and in some cases can be used as a good approximation for the component size distribution. Table II contains the analytical expressions for the asymptotes. Additionally, supporting code computing the component size distribution and the corresponding asymptotes is provided[32]. When using the asymptotical expressions to approximate $w(n)$, one should pay attention to two factors that follow from central limits: firstly $n$ should be large, secondly the approximation is best for $\theta$ close to zero. Finally, small deviations or a cutoff in a heavy-tailed degree distribution can trigger considerable and non-trivial changes in $w(n)$, for instance, the change of the asymptotical mode of the latter.

A. Degree distributions with a cutoff

In practice, no empirical degree distribution is a heavy-tailed one. Most of the ‘real-world’ degree distributions feature a cutoff, $u(k) = 0, k > k_{\text{cut}}$, and therefore fail to be heavy-tailed in the strict sense of the definition [15]. It turns out that if a cutoff is featured at large enough $k_{\text{cut}}$, the above-provided asymptotic analysis still has a relevant meaning. This situation can be compared to how we commonly attribute the fractal dimension to real-world geometric objects that fail to be fractals on infinitesimal scales.

Suppose one applies a cutoff at $k_{\text{cut}}$ to a degree distribution, $u(k)$, that features a heavy tail. Since $u(k)$ has a finite support, the asymptote of associated $w(n)$ is covered by Case A (Table II), however, if $k_{\text{cut}}$ is large, $w(n)$ may also transiently follow the original asymptote. Instead of an analytical investigation, we demonstrate the influence of the cutoff with numerical examples obtained by computing (8). This influence strongly depends on how the sign of $\theta$ is affected by the introduction of the cutoff. For example, if $\theta > 0$ even after the cutoff, the cutoff will cause more nodes to appear in finite-size components, and thus the component size distribution will shift towards larger sizes. The opposite case is valid when $\theta \leq 0$ before (and after) the cutoff, then the cutoff causes the component size distribution to shift towards smaller sizes. The third option is when the cutoff changes the sign of $\theta$ form ‘+’ to ‘−’. In this case, both shifts are possible. Fig. 5 shows how a component size distribution that corresponds to degree exponent $\beta = 3.3$ is affected by a cutoff with various values of $k_{\text{cut}}$. 

probability, \( w(n) \)

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Component size, } n & 10^0 & 10^1 & 10^2 & 10^3 & 10^4 & 10^5 \\
\text{Probability, } w(n) & 10^{-7} & 10^{-6} & 10^{-5} & 10^{-4} & 10^{-3} & 10^{-2} \\
\end{array}
\]

1. 2. 3. 4. 5.

The asymptote for \( k \) is covered by Case A.

**B. Excess degree distribution with no mean value**

In principle, the excess degree distributions that do not have a mean value, i.e. \( \beta < 2 \), do not fall within any of the above categories. However, if one introduces a cut-off, \( u(k) \) will feature finite moments including, \( \mu_3 < \infty \), hence this case should be treated according to Case A of Table [I]. Fig. [5] shows how cutoffs at \( k = k_{\text{cut}} \) influence an instance of component size distribution with \( \beta = 1 \). Unlike as in the previous example, in which \( u(k) \) with no cutoff generates a valid \( w(n) \), here the increase of \( k_{\text{cut}} \) results in vanishing probability of finding a finite-size component at all: for any \( n, w(n) \to 0 \) when \( k_{\text{cut}} \to \infty \).

This illustrates the fact that finite-size components do not exist for \( \beta \leq 1 \) also follows from the fact that in this case \( \mu_1 \) diverges and the point values of \( w(n) \), as given below the the definition [9], tend to zero.

Suppose the cutoff in the empirical, heavy-tailed degree distribution is due to the fact that the network sample has a finite size, \( k_{\text{cut}} = N, N \neq \infty \), then one may approximate the expected number of edges in this sample as

\[
n_e = \frac{N \mu_1}{2} = \frac{N}{2} \sum_{k=1}^{N} k u(k) \approx \frac{N}{2} \sum_{k=1}^{N} k^{-\beta+1}, \quad N >> 1,
\]

so that

\[
n_e \approx \begin{cases} \quad N(1 - N^{2-\beta}), & \beta \neq 2, \\ N \log N, & \beta = 2. 
\end{cases}
\]

Subsequently, three scenarios are possible here:

i) **Sparse network**, \( n_e = CN, C > 0, \beta > 2 \) : the asymptotic modes are given in Table [II].

ii) **Semi-dense network**, either \( n_e = CN \log N, \beta = 2 \) or \( n_e = CN^{3-\beta}, 1 < \beta < 2 \) : the mean value of excess distribution diverges; there are finite components but no power law in the distribution of component sizes.

iii) **Dense network**, \( n_e = CN^{3-\beta}, \beta \leq 1 \) : the mean value of degree distribution \( \mu_1 \to \infty \), and finite components vanish as \( N \to \infty \).

**C. The role of the giant component**

All the cases presented in Table [II] depend in some way on the value of \( \theta \). This is not a coincidence as the sign of \( \theta \) is the indicator for the giant component existence. If the degree distribution features a heavy tail with exponent \( \beta \geq 3 \), depending upon the value of \( \theta \), there are two possible heavy-tail exponents for the component size distribution: subcritical branch \( \eta = \beta - 1 \).
\[ \eta = \frac{\beta - 1}{\beta - 2} \]

when \( \theta < 0 \), and critical branch \( \eta = \min\{\frac{3}{2}, \frac{\beta - 1}{\beta - 2}\} \) when \( \theta = 0 \). This relation is illustrated in Fig. 7 where the component size distribution exponent \( \eta \) is plotted versus the degree-distribution exponent \( \beta \). We can see that if the giant component exists, \( \theta > 0 \), then irrespectively of what is the degree distribution, the component size distribution always decays faster then the power law. Therefore it can be concluded that the giant component is not compatible with a heavy-tailed component size distribution. Any degree distribution with \( \beta < 3 \), leads to a giant component since \( \theta \) can only be positive in this case. Furthermore, if \( \beta \leq 1 \) the giant component is also the only component: with probability 1 the configuration network is fully connected.

**ACKNOWLEDGMENTS**

This work is part of the research program Veni with project number 639.071.511, which is financed by the Netherlands Organisation for Scientific Research (NWO).

[1] A.-L. Barabási and R. Albert, science 286, 509 (1999).
[2] M. E. Newman, D. J. Watts, and S. H. Strogatz, PNAS 99, 2566 (2002).
[3] A. Vázquez, Physical Review E 67, 056104 (2003).
[4] I. Kryven, J. Duivenvoorden, J. Hermans, and P. D. Iedema, Macromolecular Theory and Simulations 25, 449 (2016).
[5] U. Bhat, P. L. Krapivsky, R. Lambiotte, and S. Redner, Physical Review E 94, 062302 (2016).
[6] M. E. J. Newman and D. J. Watts, Physical Review E 60, 7332 (1999).
[7] M. Molloy and B. Reed, Combinatorics, probability and computing 7, 295 (1998).
[8] M. Newman, *Networks: an introduction* (Oxford university press, Oxford, 2010).
[9] F. Chung, L. Lu, and V. Vu, PNAS 100, 6313 (2003).
[10] I. Kryven, Phys. Rev. E 94, 012315 (2016).
[11] M. Molloy and B. Reed, Random structures & algorithms 6, 161 (1995).
[12] I. Kryven and P. D. Iedema, Chemical Engineering Science 126, 296 (2014).
[13] C. Biemann and A. van den Bosch, *Structure discovery in natural language* (Springer Science & Business Media, 2011).
[14] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Physical review E 64, 026118 (2001).
[15] M. E. J. Newman, Phys. Rev. E 76, 045101 (2007).
[16] S.-H. Yook, H. Jeong, and A.-L. Barabási, PNAS 99, 13382 (2002).
[17] E. Ravasz and A.-L. Barabási, Physical Review E 67, 026112 (2003).
[18] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, Physical review letters 94, 018102 (2005).
[19] F. Fu, L. Liu, and L. Wang, Physica A: Statistical Mechanics and its Applications 387, 675 (2008).
[20] L. A. Adamic and B. A. Huberman, Glottometrics 3, 143 (2002).
[21] G. Timár, S. N. Dorogovtsev, and J. F. F. Mendes, Physical Review E 94, 022302 (2016).
[22] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Physical review letters 85, 4633 (2000).
TABLE II. Asymptotic behaviour of component sizes \( w(n) \), in terms of degree distribution parameters: the first three moments \( \mu_1, \mu_2, \mu_3 \), scale parameter \( s \), and exponent \( \beta \). Supporting source code available in Ref. [32].

| Finite moments of \( u(k) \) | \( u(k), k \to \infty \) | \( \theta = \mu_2 - 2\mu_1 \) | Asymptote of \( w(n) \) |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| \( \mu_3 < \infty \) | A. \( o(k^{-\beta}), \beta > 4 \) | \( \theta \neq 0 \) | \( C_1 e^{-C_2 n^{\theta - 2}} \) |
| | B. \( O(k^{-\beta}), \beta > 4 \) | \( \theta = 0 \) | \( C_1 n^{-\theta / 2} \) |
| | C. \( O(k^{-\beta}), \beta = 4 \) | \( \theta < 0 \) | \( C_1 n^{-\theta / 2} \) |
| | D. \( O(k^{-\beta}), 3 < \beta < 4 \) | \( \theta > 0 \) | \( C_1 n^{-\theta / 2} \) |
| \( \mu_3 = \infty, \mu_2 < \infty \) | E. \( O(k^{-\beta}), \beta = 3 \) | \( \theta > 0 \) | \( C_2 e^{-C_3 n^{\theta - 2}} \) |
| | F. \( O(k^{-\beta}), 2 < \beta < 3 \) | \( \theta > 0 \) | \( C_3 e^{-C_4 n^{\theta - 2}} \) |

\[
\begin{align*}
C_1 &= \frac{e^{\alpha^2}}{\sqrt{2\pi}(\mu_1^2 + \mu_2^2)} , \\
C_2' &= \frac{\mu_1}{\sqrt{2\pi \alpha^2}} , \\
C_3 &= \frac{(\mu_2 - 2\mu_1)^2}{2(2\mu_1^2 - 2\mu_2^2)} , \\
C_4 &= \sqrt{2\pi}(\alpha) \frac{\sin \left( \frac{\pi \alpha}{\alpha + 1} \right)}{\alpha + 1} \Gamma(\alpha) \Gamma(\alpha + 1) , \\
C_6 &= (1 - \alpha) \left( \frac{2(\alpha^2 - 1)}{\alpha^2 - 1} \frac{\pi \sin \left( \frac{\pi \alpha}{\alpha + 1} \right)}{\alpha + 1} \right) , \\
C_7 &= \frac{e^{\alpha^2}}{\sqrt{2\pi \alpha^2}} , \\
C_8 &= \frac{1}{\pi s} + \frac{1}{2} , \\
C_9 &= e^{-1 - \frac{\alpha^2}{\alpha^2}} , \\
C_{10} &= \frac{\Gamma(\alpha) \sin \left( \frac{\pi \alpha}{\alpha + 1} \right)}{\alpha + 1} , \\
C_{11} &= (1 - \alpha) \left( \frac{2(\alpha^2 - 1)}{\alpha^2 - 1} \frac{\pi \sin \left( \frac{\pi \alpha}{\alpha + 1} \right)}{\alpha + 1} \right) .
\end{align*}
\]

[23] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures* (Cambridge University Press, Cambridge, 1998).
[24] I. Kryven and P. Iedema, Macromolecular Theory and Simulations 22, 89 (2013).
[25] D. V. Kruchinin, Advances in Difference Equations 2015, 1 (2015).
[26] N.-Y. Ma and F. Liu, Applied mathematics and computation 158, 225 (2004).
[27] N.-Y. Ma and F. Liu, Applied Mathematics and Computation 133, 83 (2002).
[28] B. V. Gnedenko and A. Kolmogorov, *Limit distributions for sums of independent variables* (Addison-Wesley, Cambridge, 1954).
[29] V. V. Uchaikin and V. M. Zolotarev, *Chance and stability: stable distributions and their applications* (Walter de Gruyter, Utrecht, 1999).
[30] C. M. Ramsay, Communications in Statistics—Theory and Methods 35, 395 (2006).
[31] J. Nacher and T. Akutsu, Physica A: Statistical Mechanics and its Applications 390, 4636 (2011).
[32] “Matlab/GNU Octave source code for calculating component size distribution in configuration network,” https://github.com/ikryven/PhysRevE_2017_GECS