

**Condensation of degrees emerging through a first-order phase transition in classical random graphs**

Fernando L. Metz

*Institute of Physics, Federal University of Rio Grande do Sul, 91501-970 Porto Alegre, Brazil;  
Physics Department, Federal University of Santa Maria, 97105-900 Santa Maria, Brazil;  
and London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, United Kingdom*

Isaac Pérez Castillo

*Department of Quantum Physics and Photonics, Institute of Physics, UNAM, P.O. Box 20-364, 01000 Mexico City, Mexico,  
and London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, United Kingdom*

(Received 17 April 2019; published 15 July 2019)

Due to their conceptual and mathematical simplicity, Erdős-Rényi or classical random graphs remain as a fundamental paradigm to model complex interacting systems in several areas. Although condensation phenomena have been widely considered in complex network theory, the condensation of degrees has hitherto eluded a careful study. Here we show that the degree statistics of the classical random graph model undergoes a first-order phase transition between a Poisson-like distribution and a condensed phase, the latter characterized by a large fraction of nodes having degrees in a limited sector of their configuration space. The mechanism underlying the first-order transition is discussed in light of standard concepts in statistical physics. We uncover the phase diagram characterizing the ensemble space of the model, and we evaluate the rate function governing the probability to observe a condensed state, which shows that condensation of degrees is a rare statistical event akin to similar condensation phenomena recently observed in several other systems. Monte Carlo simulations confirm the exactness of our theoretical results.

DOI: [10.1103/PhysRevE.100.012305](https://doi.org/10.1103/PhysRevE.100.012305)**I. INTRODUCTION**

Condensation occurs when a macroscopic number of degrees of freedom occupy a small region of the configuration space. This is an ubiquitous phenomenon with manifestations in physics, biology, and economics [1]. Classical examples in physics are the familiar transition from a gas to a liquid state of matter, the formation of a Bose-Einstein condensate in large systems of noninteracting bosons [2], or the emergence of a staggered magnetization in mean-field spin systems [3,4]. A crucial ingredient to observe condensation is the presence of global constraints, since they introduce correlations among the microscopic constituents of the system, and a condensed state can be formed even in the absence of interactions. In this context, Bose-Einstein condensation, where the total number of particles is conserved, constitutes the prototypical example.

In the physical examples mentioned above, condensation represents the typical or average behavior of a large system in statistical equilibrium. More recently, it has been realized that condensation may take place in a broader variety of random systems when the ensemble space is probed away from the typical fluctuations around the average. The term *condensation of fluctuations* has been coined to describe such condensed states that are triggered by large deviations of an extensive random variable but whose typical behavior does not necessarily show any sign of condensation. Examples of random systems, where condensation emerges as a rare event, include the Gaussian model [5], the Urn model [6], and models of mass transport [7,8], to name just a few. Condensation of fluctuations usually brings about a rich phe-

nomenology including phase transitions, giant responses to small perturbations, and singularities in the full probability distribution [9].

Here we study condensation of degree fluctuations in classical random graphs, introduced a long time ago by Solomonoff and Rapoport [10], and popularized a decade later by the seminal works of Erdős and Rényi [11,12]. Due to their mathematical and conceptual simplicity, Erdős-Rényi (ER) random graphs have an enormous number of applications in areas such as complex networks, optimization problems, spin glasses, and information theory [13,14]. An instance drawn from the ER random graph ensemble consists of a simple undirected graph of  $N$  nodes, where each pair of nodes is connected by an edge with probability  $p \sim O(1/N)$  [14]. The degree  $k_i$  is an integer random variable that counts the number of edges attached to node  $i$ , with  $i = 1, \dots, N$ . Since the edges are drawn independently, the degree distribution along the graph follows a Poisson law in the large  $N$  limit. Here we are precisely interested in condensation phenomena triggered by rare configurations of the random variables  $\{k_i\}_{i=1, \dots, N}$ .

The study of condensation in the topological structure of complex networks or random graphs has attracted a lot of interest during the last two decades [15]. In this context, condensation refers to an aggregation phenomenon where a macroscopic number of elementary structures or motifs (edges, triangles, etc.) become mutually interconnected to form a compact subgraph. Condensation of edges occurs when a finite fraction of links becomes attached to an

infinitely small fraction of nodes [16–20], producing one or more hubs—densely connected nodes—in the graph structure. Condensation has also been observed in exponential random graphs [21–28], which are sampled from a statistical weight with a Boltzmann form. The Hamiltonian of the exponential model is built in order to incorporate certain graph features, such as the total number of triangles [21–23,27] or two-stars [24,25,28], i.e., paths of length two. Exponential random graphs also display a condensed state characterized by a large number of elementary structures (triangles or two stars) clumped together into a highly interconnected cluster. These different manifestations of the condensed phase are typically characterized by large subgraphs with densely connected nodes, which translates in a subset of degrees scaling with  $N$ .

Here we take a more elementary path and study the condensation of the degree sequence  $\{k_i\}_{i=1,\dots,N}$  in a limited domain of its available configuration space. Considering random graph instances drawn from the classical ER ensemble, we ask what is the fraction of nodes having degrees inside an arbitrary interval  $[a, b]$ , with  $b > a > 0$ . We provide an exact solution to this problem by computing the full probability distribution of the random variable  $F_N[a, b]$  that counts the fraction of degrees lying in  $[a, b]$ . Being more precise, by calculating the cumulant generating function of  $F_N[a, b]$ , we show that the ensemble of ER random graphs undergoes a first-order phase transition between a Poisson-like phase, where the degree distribution is closer to its typical behavior, and a condensed phase, where the degree distribution exhibits a prominent peak. We elucidate the physical and mathematical mechanisms underlying the transition by using standard ideas from statistical mechanics. We derive the phase diagram in the ensemble parameter space, and we show it exhibits two critical lines, each one surrounded by a metastable region and terminating at a critical point. The critical lines define the set of points in the parameter space at which the degree distribution changes abruptly. We also compute the rate function characterizing the large deviation probability of  $F_N[a, b]$ , whose striking property is the nonanalytic behavior. The calculation of the rate function shows that condensation of degrees is a rare statistical event in line with the condensation of fluctuations exhibited by other random systems [9]. The theoretical results for the rate function exhibit an excellent agreement with Monte Carlo simulations.

In the next section we introduce the classical ensemble of random graphs and define the counting random variable  $F_N[a, b]$ . Section III explains the calculation of the cumulant generating function of  $F_N[a, b]$  using standard techniques of statistical mechanics. In Sec. IV we present the results for the first-order condensation transition, the phase diagram, and the rate function. We summarize our results and conclusions in Sec. V.

## II. THE CLASSICAL RANDOM GRAPH MODEL

The binary elements  $c_{ij} \in \{0, 1\}$  of the  $N \times N$  adjacency matrix defining the ensemble of Erdős-Rényi (ER) random graphs control whether there is an edge between pairs of nodes [14]: if  $c_{ij} = 1$ , nodes  $i$  and  $j$  are connected, while  $c_{ij} = 0$  means there is no link between  $i$  and  $j$ . Each instance of the ER ensemble is a simple and undirected graph ( $c_{ij} = c_{ji}$ )

without self-edges ( $c_{ii} = 0$ ). The ensemble of ER random graphs can be defined by the following probability distribution for the adjacency matrix:

$$P_{\text{ER}}(\{c_{ij}\}) = \prod_{i < j} \left[ \frac{c}{N} \delta_{c_{ij},1} + \left(1 - \frac{c}{N}\right) \delta_{c_{ij},0} \right], \quad (1)$$

where the product  $\prod_{i < j}$  runs over all distinct pairs of nodes. The degree of node  $i$ , defined as

$$k_i^{(N)} = \sum_{j=1(\neq i)}^N c_{ij}, \quad (2)$$

gives the number of edges connected to  $i$  in a single graph realization. The random variable  $k_i^{(N)}$  fluctuates from node to node, and the average degree reads

$$c = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \langle k_i^{(N)} \rangle, \quad (3)$$

where  $\langle \dots \rangle$  represents the ensemble average over  $\{c_{ij}\}$  with the distribution in Eq. (1). Here we consider the sparse regime, where  $c$  is finite and independent of  $N$ . Note that, in the above definition of the ER ensemble,  $c$  is a control parameter that plays the role of a soft constraint on the degrees, changing dramatically the topological structure of the random graphs generated from Eq. (1). Indeed, the ER model undergoes a second-order percolation transition: for  $c < 1$  the graph is composed of many finite clusters, each one containing a total number of  $O(1)$  nodes, while for  $c \geq 1$  a giant cluster with  $O(N)$  nodes emerges continuously as a function of  $c$  [14].

Here we explore the condensation of degrees through the random variable

$$F_N[a, b] = \frac{1}{N} \sum_{i=1}^N [\Theta(b - k_i^{(N)}) - \Theta(a - k_i^{(N)})], \quad (4)$$

where  $\Theta(x)$  is the Heaviside step function. Clearly,  $F_N[a, b]$  counts the fraction of nodes having degrees within the interval  $[a, b]$ . In the limit  $N \rightarrow \infty$ , its typical value becomes

$$f_{\text{typ}} \equiv \lim_{N \rightarrow \infty} \langle F_N[a, b] \rangle = \sum_{k=0}^{\infty} p_c(k) I_{[a,b]}(k), \quad (5)$$

with  $I_{[a,b]}(k) \equiv \Theta(b - k) - \Theta(a - k)$  an indicator function. The quantity  $p_c(k)$  is the well-known  $N \rightarrow \infty$  limit of the degree distribution of ER random graphs, given by a Poisson law with average  $c$  [14],

$$p_c(k) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \langle \delta_{k, k_i^{(N)}} \rangle = \frac{e^{-c} c^k}{k!}, \quad (6)$$

where the symbol  $\delta$  denotes the Kronecker delta function. It is straightforward to check that  $f_{\text{typ}}$  vanishes for  $c \rightarrow \infty$  and  $c \rightarrow 0$ , whereas it has a maximum at some value of  $c \in [a, b]$ . Thus, when ER random graphs with  $c \gg a, b$  are sampled from Eq. (1), the fraction  $f_{\text{typ}}$  is typically very small. However, in this particular situation, it is natural to ask what is the probability that Eq. (1) generates samples with a large subset of nodes with degrees in  $[a, b]$ , in spite of  $c$  lying far outside  $[a, b]$ . Here we tackle this problem by computing exactly the full probability distribution of  $F_N[a, b]$ , which

allows us to probe atypical, large fluctuations in the degree statistics of ER random graphs. We show that the ensemble space of ER random graphs have a surprisingly rich structure, displaying a first-order transition to a condensed configuration of  $\{k_i\}_{i=1,\dots,N}$  caused by rare fluctuations of  $F_N[a, b]$  around its typical value.

### III. CALCULATION OF THE CUMULANT GENERATING FUNCTION

The full statistics of  $F_N[a, b]$  for large  $N$  is captured by the cumulant generating function (CGF)

$$\mathcal{G}_{[a,b]}(y) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathcal{Z}_{[a,b]}^{(N)}(y), \quad (7)$$

where

$$\mathcal{Z}_{[a,b]}^{(N)}(y) \equiv \langle e^{yN F_N[a,b]} \rangle. \quad (8)$$

All cumulants of  $F_N[a, b]$  are obtained by taking derivatives of Eq. (7) with respect to  $y$ . The leading contribution to the probability  $\mathcal{P}_{[a,b]}^{(N)}(f)$  of observing a fraction  $0 \leq f \leq 1$  of nodes with degrees in  $[a, b]$  decays, for large  $N$ , according to the large deviation principle [29]

$$\mathcal{P}_{[a,b]}^{(N)}(f) \simeq \exp[-N\Psi_{[a,b]}(f)], \quad (9)$$

with the rate function  $\Psi_{[a,b]}(f)$  given by the Legendre-Fenchel transform of  $\mathcal{G}_{[a,b]}(y)$

$$\Psi_{[a,b]}(f) = \sup_{y \in \mathbb{R}} [yf - \mathcal{G}_{[a,b]}(y)], \quad (10)$$

where we have considered the Legendre-Fenchel transform to be invertible. There is a natural analogy between the canonical ensemble of statistical mechanics and the framework introduced above. The control parameter  $y$  plays the role of inverse temperature,  $y\mathcal{G}_{[a,b]}(y)$  is the free-energy per degree of freedom, and  $\mathcal{Z}_{[a,b]}^{(N)}(y)$  is the partition function. As we will show below, the behavior of  $\mathcal{G}_{[a,b]}(y)$  and its derivative allows us to clearly identify a first-order phase transition. As can be noted from Eq. (8),  $y$  is responsible for biasing the configurations of the ER ensemble: positive (negative) values of  $y$  favor configurations corresponding to large (small) values of  $F_N[a, b]$ . Thus, for a fixed value of  $y$ , we have a measure from which to derive the relevant statistical properties of the biased ensemble describing atypical graph realizations. In particular, the degree distribution of the constrained ensemble is simply given by

$$p_y(k) = \lim_{N \rightarrow \infty} \frac{\langle \frac{1}{N} \sum_{i=1}^N \delta_{k,k_i} e^{yN F_N[a,b]} \rangle}{\langle e^{yN F_N[a,b]} \rangle}. \quad (11)$$

Let us proceed to the calculation of the CGF. The partition function can be rewritten as

$$\mathcal{Z}_{[a,b]}^{(N)}(y) = \sum_{k_1, \dots, k_N=0}^{N-1} e^{y \sum_{i=1}^N I_{[a,b]}(k_i)} \mathcal{P}_N(k_1, \dots, k_N), \quad (12)$$

where

$$\mathcal{P}_N(k_1, \dots, k_N) \equiv \left\langle \prod_{i=1}^N \delta_{k_i, \sum_{j=1}^N c_{ij}} \right\rangle \quad (13)$$

is the joint distribution of degrees, which are correlated random variables. In Appendix A we explain how to recast the

CGF in the following integral form:

$$\mathcal{G}_{[a,b]}(y) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left[ \int_{-\infty}^{\infty} d\mu e^{N \mathcal{F}_{[a,b]}(\mu|y)} \right], \quad (14)$$

where

$$\mathcal{F}_{[a,b]}(\mu|y) = \frac{c}{2} - \frac{1}{2}c\mu^2 + \ln \left[ \sum_{k=0}^{\infty} p_c(k) e^{y I_{[a,b]}(k)} \mu^k \right]. \quad (15)$$

In the limit  $N \rightarrow \infty$ , the integral in Eq. (14) is dominated by the *global maximum* of  $\mathcal{F}_{[a,b]}(\mu|y)$  with respect to the order parameter  $\mu$ , and the integral can be solved through the Laplace method. Defining  $\mu_g$  as the global maximum of  $\mathcal{F}_{[a,b]}(\mu|y)$ , we obtain

$$\mathcal{G}_{[a,b]}(y) = \mathcal{F}_{[a,b]}(\mu_g|y), \quad (16)$$

where  $\mu_g$  is determined from the solution of the fixed-point equation

$$\mu = \frac{\sum_{k=0}^{\infty} p_c(k) e^{y I_{[a,b]}(k+1)} \mu^k}{\sum_{k=0}^{\infty} p_c(k) e^{y I_{[a,b]}(k)} \mu^k} \quad (17)$$

that corresponds to the global maximum of the function  $\mathcal{F}_{[a,b]}(\mu|y)$  given by Eq. (15). The fixed-point equation (17) is derived by requiring that  $\mathcal{F}_{[a,b]}(\mu|y)$  is stationary with respect to  $\mu$ , i.e.,  $\frac{d\mathcal{F}_{[a,b]}(\mu|y)}{d\mu} = 0$ . In principle,  $\mathcal{F}_{[a,b]}(\mu|y)$  may have more than a single maximum depending on the control parameters  $(c, y)$ . The rate function follows from Eq. (10),

$$\Psi_{[a,b]}(f) = yf - \mathcal{F}_{[a,b]}(\mu_g|y), \quad (18)$$

where the fraction  $0 \leq f \leq 1$  is obtained from the first derivative of the CGF  $f = \frac{\partial \mathcal{F}_{[a,b]}(\mu_g|y)}{\partial y}$ , namely,

$$f = \frac{\sum_{k=0}^{\infty} p_c(k) I_{[a,b]}(k) e^{y I_{[a,b]}(k)} \mu_g^k}{\sum_{k=0}^{\infty} p_c(k) e^{y I_{[a,b]}(k)} \mu_g^k}. \quad (19)$$

In a similar manner, one can show that the degree distribution of the constrained ensemble, defined in Eq. (11), takes the following form:

$$p_y(k) = \frac{p_c(k) e^{y I_{[a,b]}(k)} \mu_g^k}{\sum_{k=0}^{\infty} p_c(k) e^{y I_{[a,b]}(k)} \mu_g^k}. \quad (20)$$

By combining Eqs. (17) and (20), we obtain

$$\mu^2 = \frac{1}{c} \sum_{k=0}^{\infty} k p_y(k), \quad (21)$$

which provides the physical meaning of the order parameter  $\mu$ .

For  $y = 0$ , we obtain  $\mu = 1$ , which implies that  $f$  and  $p_y(k)$  reduce to their standard expressions arising from typical fluctuations of ER random graphs [see Eqs. (5) and (6)]. Equations (16)–(21) constitute the main analytical results of this work, as they determine completely the statistical properties of the random variable  $F_N[a, b]$  for ER random graphs.

### IV. FIRST-ORDER TRANSITION AND CONDENSATION OF DEGREES

In order to understand the mechanism underlying the first-order transition, we start investigating the behavior of

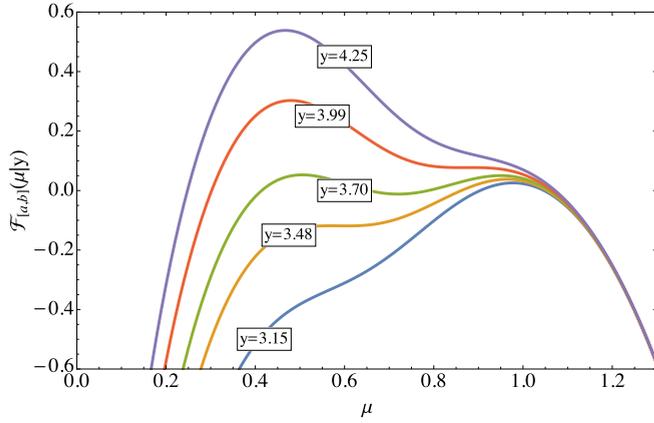


FIG. 1. The quantity  $\mathcal{F}_{[a,b]}(\mu|y)$  as a function of the order parameter  $\mu$  for average degree  $c = 13$ ,  $a = 1$ ,  $b = 3$ , and increasing values of  $y$ . The global maximum of  $\mathcal{F}_{[a,b]}(\mu|y)$  with respect to  $\mu$  provides the cumulant generating function [see Eqs. (14)–(16)]. The behavior of  $\mathcal{F}_{[a,b]}(\mu|y)$  characterizes the emergence of a first-order transition: for  $y$  approximately in the range (3.48, 3.99), the function  $\mathcal{F}_{[a,b]}(\mu|y)$  exhibits two maxima, which have the same height only at the critical value  $y_{FT} \simeq 3.70$ . The function  $\mathcal{F}_{[a,b]}(\mu|y)$  has a single maximum for  $y \lesssim 3.48$  and  $y \gtrsim 3.99$ .

$\mathcal{F}_{[a,b]}(\mu|y)$  as a function of the order parameter  $\mu$ . Here the global maximum  $\mathcal{F}_{[a,b]}(\mu_g|y)$  yields the CGF, which is analogous to the (equilibrium) free energy in statistical mechanics. Figure 1 depicts the typical functional form of  $\mathcal{F}_{[a,b]}(\mu|y)$  for high  $c$  and increasing values of  $y$ . For small  $y$ , closer to the typical case  $y = 0$ ,  $\mathcal{F}_{[a,b]}(\mu|y)$  has a single maximum. Increasing the value of  $y$  leads to the emergence of a second maximum, which means that Eq. (17) admits three fixed-point solutions: two maxima and one minimum. The portion of the phase diagram where  $\mathcal{F}_{[a,b]}(\mu|y)$  has two maxima defines a metastable region, within which the global maximum yields the leading contribution to the integral in Eq. (14). Finally, the heights of  $\mathcal{F}_{[a,b]}(\mu|y)$  corresponding to the two different maxima become even only at  $y_{FT}$ , and both fixed-point solutions for  $\mu$  contribute equally to the saddle-point integral in Eq. (14). The set of critical values  $(c_{FT}, y_{FT})$  in the parameter space, where both solutions contribute equally to the CGF, defines a first-order transition line.

Figure 2 illustrates  $\mathcal{F}_{[a,b]}(\mu|y)$  as a function of  $\mu$  for some values of  $(c_{FT}, y_{FT})$ . As we move along the critical line in the parameter space, the two maxima of  $\mathcal{F}_{[a,b]}(\mu|y)$  gradually approach each other, until they finally merge into a single maximum at a critical point, marking the end of the critical line. This physical picture is analogous to the Van der Waals liquid-gas phase transition or the ferromagnetic mean-field transition in the presence of an external field, with the proviso that our current exact analysis is not a mean field theory for a more complex underlying model. As a consequence, the metastable region cannot be promoted to a coexistence region.

In light of the aforementioned discussion, the CGF  $\mathcal{G}_{[a,b]}(y)$  is necessarily a continuous function of  $(c, y)$ . However, the derivative of  $\mathcal{G}_{[a,b]}(y)$  with respect to  $y$ , which renders the fraction  $f$ , exhibits a jump when crossing the transition line. Figure 3 shows the discontinuous behavior of  $f$  as a function of  $y$  which emerges at sufficiently large values of  $c$  in comparison to the interval  $[a, b]$ . Such discontinuity of the first

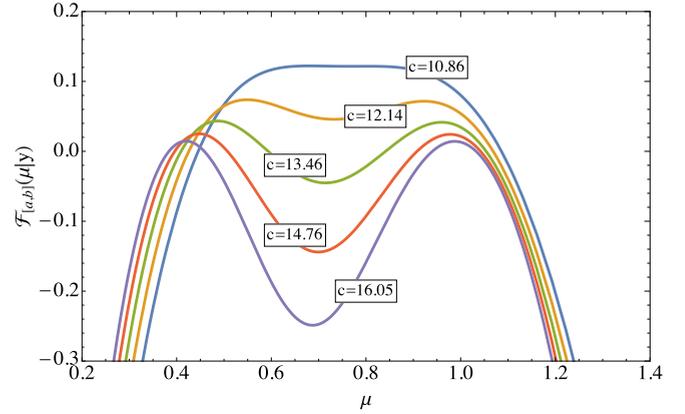


FIG. 2. The quantity  $\mathcal{F}_{[a,b]}(\mu|y)$  as a function of the order parameter  $\mu$  for  $a = 1$ ,  $b = 3$ , and different values of the average degree  $c$  along the upper critical line (see Fig. 5). The two maxima have the same height and thus contribute equally to the cumulant generating function, which features the coexistence of phases. The two maxima merge into a single maximum right at the critical point.

derivative of  $\mathcal{G}_{[a,b]}(y)$  is the hallmark of a *first-order phase transition*. As illustrated in Fig. 3, the discontinuity becomes more prominent for increasing  $c \gg 1$ , whereas below a certain value of  $c$  the fraction  $f$  increases smoothly with  $y$ .

Let us now characterize the different phases by studying the behavior of the degree distribution  $p_y(k)$  across the transition line. Figure 4 shows the average  $\langle k \rangle_y$  and the variance  $\sigma_k^2 = \langle k^2 \rangle_y - \langle k \rangle_y^2$  of  $p_y(k)$  for  $c = 13$  and increasing values of  $y$ . Clearly, the degree statistics changes abruptly at the critical point  $y_{FT}$ , with both  $\langle k \rangle_y$  and  $\sigma_k^2$  exhibiting a discontinuous behavior. For  $y < y_{FT}$ , the random graph instances have a wider range of degrees, and the distribution  $p_y(k)$  is closer to a Poisson law with average  $c$ . For  $y > y_{FT}$ , the random graph instances become more homogeneous, since the majority of nodes have degrees within  $[a, b]$ . Accordingly, the variance  $\sigma_k^2$  becomes smaller than  $\langle k \rangle_y$  and the distribution  $p_y(k)$  has a peak at a certain  $k \in [a, b]$ , closer to the degree distribution of a random regular graph. In fact, if we choose

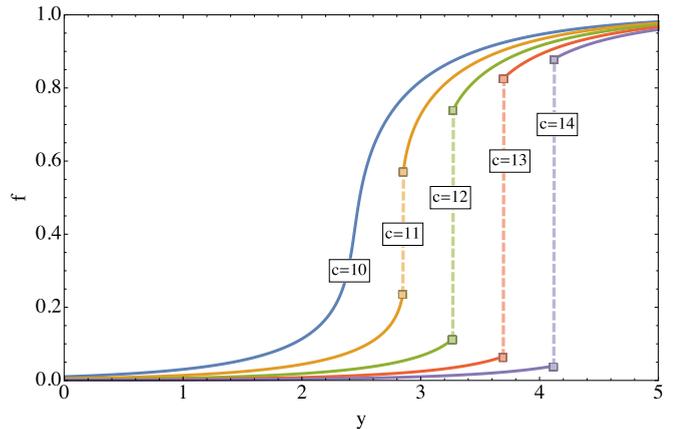


FIG. 3. Fraction  $f$  of degrees within the interval  $[1,3]$  as a function of  $y$  for different values of the average degree  $c$ . For large values of  $c$ ,  $f$  has a discontinuous behavior at a critical value  $y_{FT}$ , marking the first-order phase transition.

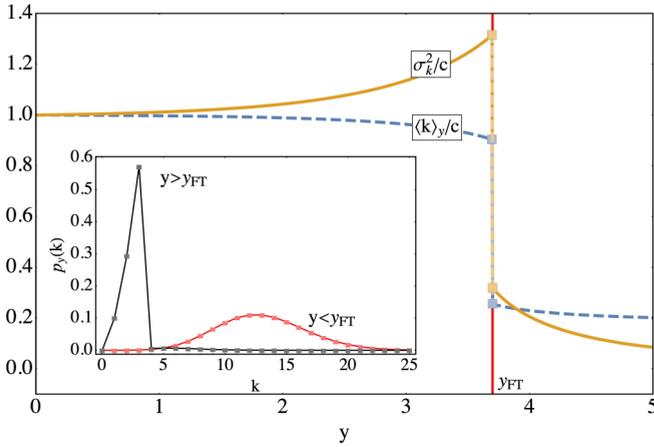


FIG. 4. Average  $\langle k \rangle_y$  (dashed line) and variance  $\sigma_k^2$  (solid line) of the degree distribution  $p_y(k)$  [see Eq. (11)] characterizing rare graph samples generated from Eq. (1), with  $c = 13$ , and conditioned to have a certain fraction  $f$  of degrees inside the interval  $[1,3]$ . The concurrent behavior of  $f$  as a function of  $y$  is presented in Fig. 3. The inset shows the typical profile of the degree distribution in each phase. The quantities  $\langle k \rangle_y$  and  $\sigma_k^2$  have a discontinuous behavior that features the abrupt change of the degree statistics along the first-order phase transition.

an interval  $[a, b]$  such that it contains a single degree  $K$ , Eq. (11) converges to  $p_y(k) = \delta_{k,K}$  for  $y \rightarrow \infty$ . The degree distribution  $p_y(k)$  in each phase is shown as an inset in Fig. 4. In summary, ER random graphs undergo a topological first-order transition between an heterogeneous phase, identified by a broader spectrum of degrees, to an homogeneous phase, where the degrees condensate in the interval  $[a, b]$ .

The above results are summarized in the phase diagram of Fig. 5, where a density plot for the fraction  $f$  on the parameter space  $(c, y)$  is presented in a logarithmic color scale. There exist two first-order critical lines, indicated by dashed lines, for small and large values of the average degree  $c$ . Each critical line terminates at a critical point (solid yellow circles). The black solid curves surrounding each critical line delimit the metastable regions around each first-order transition line, within which Eq. (17) has three fixed-point solutions. As we can appreciate from Fig. 5, by fixing  $y > 0$  sufficiently large, the first-order transition appears for  $c$  large in comparison to  $[a, b]$ , when we simultaneously require that samples have a large average degree and a large fraction  $f$ . By setting  $y < 0$ , with  $|y|$  sufficiently large, the first-order transition appears for  $c \in [a, b]$ , which is incompatible with a small fraction  $f$ . Thus, the appearance of two first-order phase transitions is due to the existence of two distinct situations where conflicting constraints are imposed on the generation of random graph samples.

Finally, we compute the rate function  $\Psi_{[a,b]}(f)$  controlling the large deviation probability of having values of  $f$  away from its typical value  $f_{typ}$  [see Eq. (9)]. According to Eq. (10), the rate function is the Legendre-Fenchel transform of the CGF with respect to  $y$ . The fact that the CGF has a nonanalytic behavior in its first derivative  $\frac{d\mathcal{G}_{[a,b]}(y)}{dy}$  results in an affine Legendre-Fenchel transform. Consequently, the rate function has a linear behavior between two values of  $f$ , whose slope is

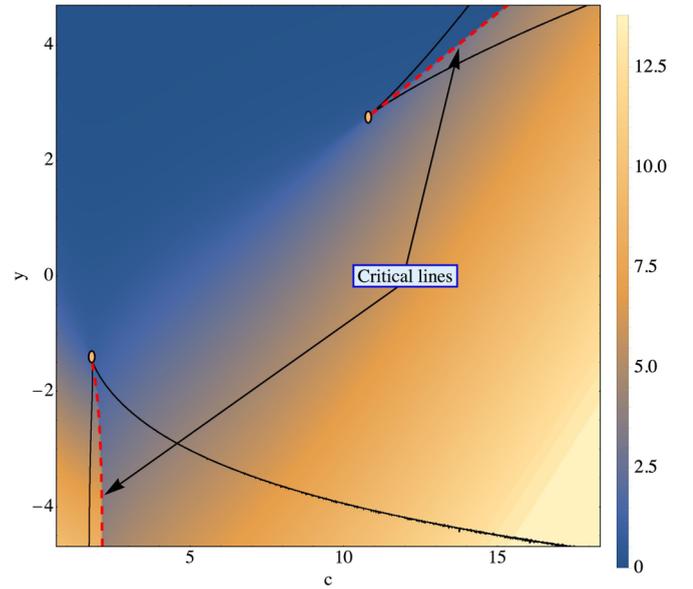


FIG. 5. Phase diagram showing the fraction  $f$  of degrees lying in  $[1,3]$  for each combination of parameters  $(c, y)$ . The values of  $-\log(f)$  are displayed in a density plot according to the color scale. The red dashed curves denote first-order critical lines. The black solid lines surrounding each critical line delimit the regions of the phase diagram where the saddle-point Eq. (17) has three fixed-point solutions. The solid yellow circles, located at the end of each critical line, represent critical points at which the two maxima of  $\mathcal{F}_{[a,b]}(\mu|y)$  merge into a single one (see Fig. 2).

equal to the value of  $y$  at which the CGF is nondifferentiable [30]. This is confirmed by Fig. 6, where we show  $\Psi_{[a,b]}(f)$  as a function of  $f$  for different values of  $c$ , in particular the appearance of the linear behavior of  $\Psi_{[a,b]}(f)$  for  $c = 2$ , which corresponds to the point where the CGF is nondifferentiable. Since the condensed phase is characterized by a very small fraction  $f$  in the case of  $c = 2$ , Fig. 6 clearly shows that condensation of degrees is a rare statistical event. By expanding the rate function around its minimum  $\Psi_{[a,b]}(f_{typ}) = 0$ , we obtain

$$\Psi_{[a,b]}(f) = \frac{1}{2} \frac{(f - f_{typ})^2}{\sigma_f^2}, \quad (22)$$

with

$$\sigma_f^2 = c \left[ \sum_{k=0}^{\infty} p_c(k) I_{[a,b]}(k+1) - f_{typ} \right]^2 + f_{typ}(1 - f_{typ}). \quad (23)$$

This implies that the typical fluctuations of  $f$  around  $f_{typ}$  are described by a Gaussian distribution with variance  $\sigma_f^2$ .

In order to confirm the exactness of our theoretical findings, we have performed a reweighted Monte Carlo method to estimate the rate function [31]. The algorithm employed to generate rare graph samples from our biased ensemble is explained in Appendix B. In our particular case, one must be careful when estimating the rate function for parameters  $(c, y)$  within the metastable region, as the Monte Carlo simulation

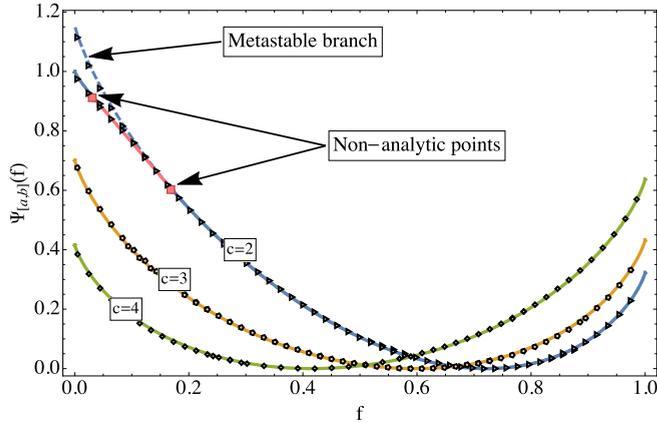


FIG. 6. Rate function  $\Psi_{[a,b]}(f)$ , characterizing the large deviation probability of Eq. (9), as a function of the fraction  $f$  of nodes with degrees within the interval  $[1,3]$ . We show the rate function for different values of the average connectivity  $c$ . The function  $\Psi_{[a,b]}(f)$  has a minimum at the typical value  $f = f_{typ}$  [see Eq. (5)]. For the case  $c = 2$  (solid blue line with triangles), we also show the presence of two nonanalytic points (light red squares) at which the second derivatives of the rate function are discontinuous. By the construction of the Legendre-Fenchel transform [30], these points are connected by a straight line (solid red line with triangles). The dashed blue line is the rate function obtained from choosing the metastable solution for the order parameter  $\mu$  when calculating the CGF. The different symbols (triangles, pentagons, and rhombuses) are simulation results obtained through a reweighting Monte Carlo method of graphs with a total of  $N = 100$  nodes (see Appendix B).

may be trapped in the metastable solution. However, this problem is easily surmounted by choosing the appropriate initial conditions. The simulation results for  $N = 100$  and their comparison with our theoretical findings are shown in Fig. 6, where we have also included the estimation of the rate function obtained from the metastable branch. The comparison between our theory and simulations shows a very good agreement.

V. CONCLUSIONS

In this work we have studied the large deviation properties of the degree sequence characterizing the Erdős-Rényi ensemble of random graphs. By studying the fluctuations of an elementary observable, namely, the fraction  $f$  of degrees lying in a given interval, we have shown that the ensemble space of ER random graphs exhibit rich critical phenomena, with the

presence of two first-order critical lines marking a topological transition in the degree statistics. As the transition lines are crossed, we have shown that the degree distribution changes abruptly from a Poisson-like profile, characteristic of typical samples from the ER ensemble, to a peaked distribution. The latter degree distribution identifies a phase where degrees are condensed or concentrated in a limited sector of their configuration space and, consequently, random graph samples corresponding to this phase have a rather homogeneous structure, similar to regular random graphs. Nevertheless, condensation of degrees in the ER ensemble is an extremely rare event, as confirmed by our computation of the rate function describing the large deviation probability for the fraction  $f$ . Such rate function exhibits two nonanalytic points akin to the presence of the first-order transition. We point out that similar properties have been identified in the probability distributions describing the condensation of fluctuations in other disordered systems [5,9]. Our theoretical results for the rate function are fully confirmed by Monte Carlo simulations.

There are an interesting number of questions to be explored. First, one wonders whether the metastable region might become a coexistence phase for a different system for which our solution would correspond to its mean-field approximation. In this regard studies on Euclidean graphs [32] seem to be the most natural candidate. Second, in the same manner that the percolation transition is inherited in the magnetic properties of the Ising model on random graphs, it is pertinent to ask what is the impact on the magnetic properties of the topological first-order transition we have observed here. In a similar context, it would be also interesting to study whether large deviations in the degree sequence can trigger the decay of the metastable states found in coupled ER random graphs [33]. Finally, in light of recent advances in the study of large deviations on diluted random matrices [34–36], we should consider the spectral properties corresponding to the constrained graph ensemble studied here. Some of these questions are currently under consideration. As a last remark, we point out that the techniques introduced here can be also useful to study exponential random graphs [14,24,28], since the approach of Sec. III can be readily generalized to the case where the indicator function is replaced by any other arbitrary function of a single degree.

ACKNOWLEDGMENTS

F.L.M. and I.P.C. thank London Mathematical Laboratory for financial support. F.L.M. also acknowledges a fellowship and financial support from CNPq/Brazil (Edital Universal 406116/2016-4).

APPENDIX A: CALCULATION OF THE CUMULANT GENERATING FUNCTION

Here we explain how to derive the expression (14) for the CGF. By introducing the integral representation of the Kronecker delta function

$$\delta_{k_i, \sum_{j=1}^N c_{ij}} = \int_0^{2\pi} \frac{du_i}{2\pi} \exp\left(iu_i k_i - iu_i \sum_{j=1(\neq i)}^N c_{ij}\right), \tag{A1}$$

Eq. (12) is rewritten as

$$\mathcal{Z}_{[a,b]}^{(N)}(y) = \sum_{k_1, \dots, k_N=0}^{N-1} \int \left( \prod_{i=1}^N \frac{du_i}{2\pi} \right) \exp \left( i \sum_{i=1}^N u_i k_i \right) \exp \left[ y \sum_{i=1}^N I_{[a,b]}(k_i) \right] \left\langle \exp \left[ -i \sum_{i < j} c_{ij} (u_i + u_j) \right] \right\rangle. \quad (\text{A2})$$

The ensemble average in Eq. (A2) is readily computed, leading to the following expression for large  $N$ :

$$\mathcal{Z}_{[a,b]}^{(N)}(y) = \sum_{k_1, \dots, k_N=0}^{N-1} \int \left( \prod_{i=1}^N \frac{du_i}{2\pi} \right) \exp \left( i \sum_{i=1}^N u_i k_i \right) \exp \left[ y \sum_{i=1}^N I_{[a,b]}(k_i) - \frac{cN}{2} + \frac{c}{2N} \left( \sum_{i=1}^N e^{-iu_i} \right)^2 \right], \quad (\text{A3})$$

where we have retained only the leading terms of  $O(N)$  in the exponent. The above equation couples the variables  $u_i$  on different nodes, which prevents the calculation of the integrals over  $u_i$ . However, by employing the Hubbard-Stratonovich transformation

$$\int_{-\infty}^{\infty} d\mu \exp \left( -\frac{1}{2} N c \mu^2 + c J \mu \right) = \sqrt{\frac{2\pi}{cN}} \exp \left( \frac{c}{2N} J^2 \right) \quad (\text{A4})$$

in Eq. (A3), with  $J = \sum_{i=1}^N e^{-iu_i}$ , we are able to integrate over  $u_1, \dots, u_N$  and obtain Eq. (14).

## APPENDIX B: THE REWEIGHTING MONTE CARLO METHOD

In this section we discuss the algorithm used to obtain the rate function using Monte Carlo simulations [31]. Let us consider the distribution  $P_{\text{ER}}(\{c_{ij}\})$ , from which we want to estimate the distribution  $\rho(f) = \text{Prob}[F_N[a, b] = f]$  of the fraction  $F_N[a, b] \equiv F_N[a, b](\{c_{ij}\})$ . The latter is given by

$$\rho(f) = \sum_{\{c_{ij}\}} P_{\text{ER}}(\{c_{ij}\}) \delta_{F_N[a,b],f}. \quad (\text{B1})$$

This distribution is concentrated around the typical value  $f_{\text{typ}}$ , and, by generating graph samples from  $P_{\text{ER}}(\{c_{ij}\})$ , we are able only to observe statistical events that are close to the center of the distribution  $\rho(f)$  in a reasonable computation time.

In order to induce the simulation to naturally explore rare events of the random variable  $F_N[a, b]$ , we must bias the graph configurations, so that rare events become typical ones. This is mathematically achieved by introducing the following auxiliary distribution:

$$P_{\text{ER}}^{(y)}(\{c_{ij}\}) = \frac{1}{Z(y)} P_{\text{ER}}(\{c_{ij}\}) e^{y F_N[a,b](\{c_{ij}\})}, \quad (\text{B2})$$

where  $y$  plays a role analogous to a fictitious inverse temperature. We can use this auxiliary distribution to derive the distribution of  $F_N[a, b]$  constrained to a fixed  $y$ :

$$\rho^{(y)}(f) = \sum_{\{c_{ij}\}} P_{\text{ER}}^{(y)}(\{c_{ij}\}) \delta_{F_N[a,b],f}. \quad (\text{B3})$$

If we perform a Monte Carlo simulation on  $P_{\text{ER}}^{(y)}(\{c_{ij}\})$  and estimate  $\rho^{(y)}(f)$ , the latter distribution will be concentrated around a different typical value  $f_{\text{typ}}(y)$ , which will be larger

(smaller) than the original value  $f_{\text{typ}}$  when  $y > 0$  ( $y < 0$ ). However, our aim is to estimate the original distribution  $\rho(f)$ . Fortunately, by plugging Eq. (B2) into Eq. (B3) we obtain the equation

$$\rho^{(y)}(f) = \frac{e^{y f}}{Z(y)} \rho(f), \quad (\text{B4})$$

which allows us to obtain  $\rho(f)$  from  $\rho^{(y)}(f)$ .

The idea of a reweighted Monte Carlo simulation works as follows. For a given value of  $y$ , we perform a standard Metropolis Monte Carlo simulation of  $P_{\text{ER}}^{(y)}(\{c_{ij}\})$ , consisting of proposing a change  $\{c_{ij}\} \rightarrow \{c'_{ij}\}$  of the adjacency matrix during each algorithmic step, which is then accepted with rate

$$r = \min \left\{ 1, \frac{P_{\text{ER}}^{(y)}(\{c'_{ij}\})}{P_{\text{ER}}^{(y)}(\{c_{ij}\})} \right\}. \quad (\text{B5})$$

Once this stochastic process has relaxed towards the stationary state, the corresponding Monte Carlo Markov chain can be used to estimate  $\rho^{(y)}(f)$ . Finally, all the histograms obtained for different values of  $y$ , or alternatively their rate functions, are glued together according to Eq. (B4), yielding the histogram for  $\rho(f)$ . It is important that two consecutive values of  $y$  have overlapping histograms, so that we do not need to estimate the normalization factor  $Z(y)$ .

We have used this method to generate the rate function reported in Fig. 6, where we have chosen  $y$  ranging from  $-3$  to  $3$  in steps of  $1$ . In the particular case of  $c = 2$ , for which we have a metastable branch, the correct solution is captured by sampling an initial graph configuration with an average connectivity  $(k)_y$  compatible with the expected value obtained from the theory. This little trick usually renders the Monte Carlo simulation in the appropriate branch.

- [1] M. R. Evans and T. Hanney, *J. Phys. A* **38**, R195 (2005).  
 [2] R. Pathria and P. Beale, *Statistical Mechanics* (Elsevier Science, Burlington, 1996).  
 [3] T. H. Berlin and M. Kac, *Phys. Rev.* **86**, 821 (1952).  
 [4] L. F. Cugliandolo and D. S. Dean, *J Phys. A* **28**, 4213 (1995).

- [5] M. Zannetti, F. Corberi, and G. Gonnella, *Phys. Rev. E* **90**, 012143 (2014).  
 [6] F. Corberi, *J. Phys. A* **48**, 465003 (2015).  
 [7] M. R. Evans, T. Hanney, and S. N. Majumdar, *Phys. Rev. Lett.* **97**, 010602 (2006).

- [8] J. Szavits-Nossan, M. R. Evans, and S. N. Majumdar, *J. Phys. A* **47**, 455004 (2014).
- [9] F. Corberi and A. Sarracino, *Entropy* **21**, 312 (2019).
- [10] R. Solomonoff and A. Rapoport, *Bull. Math. Biophys.* **13**, 107 (1951).
- [11] P. Erdős and A. Rényi, *Publ. Math., Debrecen* **6**, 290 (1959).
- [12] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci., Ser. A* **5**, 17 (1960).
- [13] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford Graduate Texts (Oxford University Press, Oxford, 2009).
- [14] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
- [15] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [16] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [17] G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**, 5632 (2001).
- [18] S. Dorogovtsev, J. Mendes, and A. Samukhin, *Nucl. Phys. B* **666**, 396 (2003).
- [19] S. N. Dorogovtsev, J. F. F. Mendes, A. M. Povolotsky, and A. N. Samukhin, *Phys. Rev. Lett.* **95**, 195701 (2005).
- [20] K. Anand, D. Krioukov, and G. Bianconi, *Phys. Rev. E* **89**, 062807 (2014).
- [21] D. Strauss, *SIAM Rev.* **28**, 513 (1986).
- [22] Z. Burda, J. Jurkiewicz, and A. Krzywicki, *Phys. Rev. E* **69**, 026106 (2004).
- [23] Z. Burda, J. Jurkiewicz, and A. Krzywicki, *Phys. Rev. E* **70**, 026106 (2004).
- [24] J. Park and M. E. J. Newman, *Phys. Rev. E* **70**, 066146 (2004).
- [25] J. Park and M. E. J. Newman, *Phys. Rev. E* **70**, 066117 (2004).
- [26] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Phys. Rev. E* **69**, 046117 (2004).
- [27] J. Park and M. E. J. Newman, *Phys. Rev. E* **72**, 026136 (2005).
- [28] A. Annibale and O. T. Courtney, *J. Phys. A* **48**, 365001 (2015).
- [29] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [30] H. Touchette, Legendre-Fenchel transforms in a nutshell (unpublished).
- [31] A. K. Hartmann and M. Mézard, *Phys. Rev. E* **97**, 032128 (2018).
- [32] E. Gilbert, *J. Soc. Indust. Appl. Math.* **9**, 533 (1961).
- [33] M. Bolfe, L. Nicolao, and F. L. Metz, *J. Stat. Mech.* (2018) 083404.
- [34] F. L. Metz and I. Pérez Castillo, *Phys. Rev. Lett.* **117**, 104101 (2016).
- [35] I. Pérez Castillo and F. L. Metz, *Phys. Rev. E* **98**, 020102(R) (2018).
- [36] I. P. Castillo and F. L. Metz, *Phys. Rev. E* **97**, 032124 (2018).