Journal of Statistical Physics, Vol. 111, Nos. 3/4, May 2003 (© 2003)

A Simple Asymmetric Evolving Random Network

Michel Bauer[1,2] **and Denis Bernard**[1,3]

Received March 15, 2002; accepted September 13, 2002

We introduce a new oriented evolving graph model inspired by biological networks. A node is added at each time step and is connected to the rest of the graph by random oriented edges emerging from older nodes. This leads to a statistical asymmetry between incoming and outgoing edges. We show that the model exhibits a percolation transition and discuss its universality. Below the threshold, the distribution of component sizes decreases algebraically with a continuously varying exponent depending on the average connectivity. We prove that the transition is of infinite order by deriving the exact asymptotic formula for the size of the giant component close to the threshold. We also present a thorough analysis of aging properties. We compute local-in-time profiles for the components of finite size and for the giant component, showing in particular that the giant component is always dense among the oldest nodes but invades only an exponentially small fraction of the young nodes close to the threshold.

KEY WORDS: Evolving random graphs; percolation; K–T transition; biological networks.

1. MOTIVATIONS AND RESULTS

Evolving random graphs have recently attracted attention, see, e.g., refs. 1, 3, and 9 and references therein. This interest is mainly motivated by concrete

[1] Service de Physique Théorique de Saclay, CE Saclay, 91191 Gif sur Yvette, France. (Laboratoire de la Direction des Sciences de la Matière du Commisariat à l'Energie Atomique, URA2306 du CNRS).
[2] To whom correspondence should be addressed; e-mail: bauer@spht.saclay.cea.fr
[3] Member of the CNRS; e-mail: dbernard@spht.saclay.cea.fr

0022-4715/03/0500-0703/0 © 2003 Plenum Publishing Corporation

problems related to the structure of communication or biological networks. Experimental data are now available in many contexts.[1, 7]

In these examples, the asymmetry and the evolving nature of the networks are likely to be important ingredients for deciphering their statistical properties. It is however far from obvious to find solvable cases that would possibly account for some relevant features of, say, the regulating network of a genome. Although biology has strongly influenced our interest in evolving networks, the model we solve is not based on realistic biological facts but it nevertheless incorporates asymmetry and chronological order. Understanding such simple evolving graphs may help understanding biological networks, at least by comparison and opposition.

We were initially motivated by the study of the yeast genetic regulatory network presented in ref. 7. The authors studied in and out degree distributions and discovered a strong asymmetry: a single gene may participate to the regulation of many other genes—the law for out-degrees seems to be large—but each genes is only regulated by a few other genes—the law for in-degrees seems to have finite moments. This is why we consider oriented evolving random graphs in the sequel. A biological interpretation for the asymmetry is that the few promoter-repressor sites for each gene bind only to specific proteins, but that along the genome many promoter-repressor sites are homologous. However, this does not predict the precise laws. An understanding of the same features from a purely probabilistic viewpoint would be desirable as well.

The recent experimental studies dealt with global statistical properties of evolving graphs, i.e., when the evolving network is observed at some fixed time with the ages of different vertices and edges not taken into account. There are simple experimental reasons for that: to keep track of the ages would in many cases dramatically reduce the statistics, and in other cases this information is even not available. Our second motivation is a better understanding of the local-in-time statistical properties of evolving networks. This helps dating or assigning likely ages to different structures of the networks. As we shall later see, the global analysis, which is like a time average, gives a distorted view of the real structure of the networks. We shall present a detailed analysis of local-in-time features in our model.

The model we study is the natural evolving cousin of the famous Erdös-Renyi random graphs.[5] Starting from a single vertex at time 1, a new vertex is created at each time step—so that at time t, the size of the system, i.e., the number of vertices, is t—and new oriented edges are created with specified probabilistic rules. A tunable parameter α ranging from 0 to ∞ describes asymptotically the average number of incoming edges on a vertex. Precise definitions are given in the next section.

Our main results are the following.

From very simple rules, we see an asymmetry emerging. The global in and out degree distributions are different. We also compute the local profiles of in and out degree distributions, and comment on the differences.

We make a detailed global analysis for the structure and sizes of the connected components. We use generating function methods to write down a differential equation that implies recursion relations for the distribution of component sizes, see Eqs. (11) and (14).

A salient global feature of the model is a percolation phase transition at a critical value of the average connectivity. Below this value, no single component contains a finite fraction of the sites in the thermodynamic limit, i.e., in the large t limit. However, a slightly unusual situation occurs in that below the transition the system contains components whose sizes scale like a power of the total size of the graph, see Eq. (26). Correspondingly, the probability distribution for component sizes has an algebraic queue, see Eq. (23), and its number of finite moments jumps at specific values of the average connectivity. Above the transition, this probability distribution becomes defective, but its decrease is exponential, see Eq. (39). The transition is continuous. Close to the threshold, the fraction of sites in the giant component—the percolation cluster—has an essential singularity, see Eq. (37). We argue that this result is universal, with the meaning used in the study of critical phenomena. The essential singularity at the percolation threshold had already been observed numerically by ref. 9 in a different model which we show to be in the same universality class as ours for the percolation transition, and computed analytically for another class of models in ref. 4.

We then turn to the study of local-in-time profiles of connected components. Guided by a direct enumeration based on tree combinatorics, we show that they satisfy recursion relations, and we give the first few profiles (isolated vertices, pairs, triples) explicitly. The profile of the giant component is given by a differential equation, from which we extract the singularity in the far past and the critical singularity in the present—see Eqs. (50) and (51). In particular the giant component invades all the time slices of the graph above the transition. One strange feature of profiles, which would deserve a good explanation, is that in several instances the formal parameter involved in generating functions for global quantities is simply traded for the relative age to obtain interesting local-in-time observables, see Eqs. (48) and (52).

We have compared our analytical results with numerical simulation whenever possible.

While polishing this paper, we became aware of ref. 8, whose goals overlap partly with ours. When they can be compared, the results agree.

2. THE MODEL

We construct evolving random graphs with the following rules:

(i) We consider a triangular array of independent random variables $\ell_{i,j}$, $1 \leqslant i < j$, where $\ell_{i,j}$ takes value 1 with probability $p_j \in [0, 1]$ and value 0 with probability $q_j \equiv 1 - p_j$.

(ii) We start from the graph made of single vertex at initial time $t = 1$. At time t, $t \geqslant 2$, a vertex with label t is added together with the *directed* edges $[j \to t]$ for which $\ell_{j,t} = 1$. We shall often take the viewpoint that the (biased) coin tossings defining $\ell_{i,t}$ are done at time t.

We shall assume that $p_t \simeq \alpha/t$ at large time t, with α a parameter which we shall identify as half the average connectivity. This choice ensures the convergence of various distributions to stationary measures, most of them being independent of the precise values of the early probabilities.

By construction all edges arriving at a given vertex are simultaneously created at the instant of creation of this vertex. As a consequence, these graphs are not only oriented but chronologically oriented—this is unrealistic from the biological viewpoint.

3. EDGE DISTRIBUTIONS

In this section we give the incoming and outgoing edge distributions.

Let $\ell_j^-(t)$ be the number of incoming edges at the vertex j, and $\ell_j^+(t)$ be the number of outgoing edges at this vertex at time t.

Let $v_k^-(t)$ be the number of vertices with k incoming edges, and $v_k^+(t)$ be the number of vertices with k outgoing edges at time t.

We may look either at the edge distributions at a given vertex, or we may look at the edge distributions defined by gathering averaged histograms over whole graphs. The former are specified by their generating functions,

$$\langle z^{\ell_j^{\pm}(t)} \rangle$$

where $\langle \cdots \rangle$ denotes expectation value. It may depend on the specified vertex labeled by j. The latter is defined by the generating functions,

$$\mathscr{V}_t^{\pm}(z) \equiv \frac{1}{t} \sum_{0 \leqslant k \leqslant t} \langle v_k^{\pm}(t) \rangle z^k = \frac{1}{t} \sum_{1 \leqslant j \leqslant t} \langle z^{\ell_j^{\pm}(t)} \rangle \tag{1}$$

We remark that this global histogram distribution is the average of the local-in-time quantity $\langle z^{\ell_j^{\pm}(t)} \rangle$. Since at time t the total number of vertices is

t, $\mathscr{V}_t^{\pm}(z)$ is properly normalized, $\mathscr{V}_t^{\pm}(1)=1$, to define an averaged probability distribution function, independent of the vertices, for the incoming or outgoing edge variables $\ell^{\pm}$:

$$\sum_k z^k \,\mathrm{Prob}_{(\ell^{\pm}=k)} \equiv \mathscr{V}_t^{\pm}(z)$$

Incoming Vertices. The number of incoming edges $\ell_j^-(t)=\sum_i \ell_{i,j}$ at vertex $j \leqslant t$ asymptotically possesses a Poisson distribution since

$$\langle z^{\ell_j^-(t)} \rangle = \prod_{i<j} [q_j + z p_j] \simeq_{j\to\infty} \exp(\alpha(z-1))$$

The convergence of this distribution justifies our choice of asymptotic probabilities $p_j \simeq \alpha/j$. Only the vertices whose ages j scale with the age of the graph, i.e., with $j/t=\sigma$ fixed, $0 \leqslant \sigma \leqslant 1$, give non trivial contributions at large time to the averaged histogram (1) and

$$\mathscr{V}_t^-(z) \simeq_{t\to\infty} \exp(\alpha(z-1)) \tag{2}$$

This expression may also be retrieved by looking at the evolution equation of $\mathscr{V}_t^-(z)$. Indeed, consider adding the new vertex at time t. Since the edges are oriented from older to younger vertices,[4] we have $t\mathscr{V}_t^-(z) = (t-1)\,\mathscr{V}_{t-1}^-(z) + \langle z^{\ell_t^-(t)} \rangle$ from the second definition in Eq. (1). This is equivalent to

$$t\mathscr{V}_t^-(z) = (t-1)\,\mathscr{V}_{t-1}^-(z) + (q_t + z p_t)^{t-1}$$

As $(q_t + z p_t)^{t-1} \simeq e^{\alpha(z-1)}$ at large time, the stationary limit is given by Eq. (2). This yields a Poissonian distribution with probabilities

$$\mathrm{Prob}_{(\ell^-=k)} = e^{-\alpha}\frac{\alpha^k}{k!} \tag{3}$$

Outgoing Vertices. At a given vertex $j \leqslant t$, with $j/t=\sigma$ fixed, the number of outgoing edges $\ell_j^+(t)=\sum_{i\leqslant t} \ell_{j,i}$ at vertex j also have a Poisson distribution at large time,

$$\langle z^{\ell_{\sigma t}^+(t)} \rangle = \prod_{\sigma t = j < i \leqslant t} [q_i + z p_i] \simeq_{t\to\infty} \exp(-\alpha \log \sigma (z-1))$$

[4] A vertex is older than another if it appeared before in the evolution, i.e., if it corresponds to a smaller value of σ.

but with a parameter $\alpha \log(1/\sigma)$ depending on the age of the vertex. Approximating at large time the sum over j in Eq. (1) by an integral over σ gives the histogram distribution:

$$\mathscr{V}_t^+(z) \simeq_{t\to\infty} \int_0^1 d\sigma \, \langle z^{\ell_{\sigma t}^+}(t)\rangle = \frac{1}{1+\alpha(1-z)} \tag{4}$$

As for incoming vertices, this formula follows from the evolution equation for $\mathscr{V}_t^+(z)$. Indeed, since the numbers of outgoing edges $\ell_j^+(t)$ from vertex j at time t and $t-1$ differ by $\ell_{j,t}$ we have $\langle z^{\ell_j^+(t)}\rangle = \langle z^{\ell_j^+(t-1)}\rangle \langle z^{\ell_{j,t}^+}\rangle$. From definition (1) this gives

$$t\mathscr{V}_t^-(z) = 1 + (t-1)\, \mathscr{V}_{t-1}^-(z)\, (q_t + z p_t)$$

where the first term is the contribution of the newly added vertex at time t. The stationary limit is given by Eq. (4). This is a geometric distribution, slightly larger than the Poisson distribution, with probabilities

$$\mathrm{Prob}_{(\ell^+=k)} = \frac{\alpha^k}{(1+\alpha)^{k+1}} \tag{5}$$

Mixed Distribution. Let $v_{k_+,k_-}(t)$ be the number of vertices with k_+ outgoing and k_- incoming edges at time t. As in Eq. (1), the generating function for the mixed histogram distribution is defined by

$$\mathscr{V}_t(z_+, z_-) \equiv \frac{1}{t} \sum_{k_+,k_-} \langle v_{k_+,k_-}(t)\rangle \, z_+^{k_+} z_-^{k_-} = \frac{1}{t} \sum_{1\leqslant j\leqslant t} \langle z_+^{\ell_j^+(t)} z_-^{\ell_j^-(t)}\rangle$$

By construction the outgoing and incoming edges variables $\ell_j^{\pm}(t)$ are statistically independent for j fixed, so that the last expectation values factorize. As above we may derive an evolution equation by evaluating the contribution of the newly added vertex at time t. This yields:

$$t\mathscr{V}_t(z_+, z_-) = (q_t + z_- p_t)^{t-1} + (t-1)\, \mathscr{V}_{t-1}(z_+, z_-)\, (q_t + z_+ p_t)$$

Its stationary limit is factorized:

$$\mathscr{V}(z_+, z_-) = \frac{e^{\alpha(z_- - 1)}}{1+\alpha(1-z_+)}$$

Outgoing and incoming edges are statistically independent at large time.

4. CLUSTER DISTRIBUTIONS

In this section, we present the main relations governing the probability distributions of connected components of the graphs. Two vertices belong to the same connected component if they can be joined by a path made of edges, without any reference to orientation. This definition ensures that the property of being in the same connected component is an equivalence relation, but does not ensure that two points in a connected component can be joined by an oriented path.

To partly avoid repetitions, the term *cluster* is used as a synonymous for *connected component* in the sequel.

Intuitively, the fact that the network is fragmented can be understood as follows: when a vertex t_0 is created, it has a finite probability to be isolated, and the probability that none of the vertices $t_0+1,\ldots, t$ connects to vertex t_0 is $q_{t_0+1}\cdots q_t$ which scales as $(t_0/t)^\alpha$. This quantity remains finite as long as t_0/t does. This argument shows that there are isolated vertices in the system. A small extension of the argument shows that there are also finite components and that young vertices are more likely to be in small components than old ones. This will be made more rigourous in the study of profiles, see Section 8.

Let $N_k(t)$ be the number of connected components with k vertices at time t and let $N_t(z)$ be the generating function,

$$N_t(z) = \sum_{k\geqslant 1} N_k(t)\, z^k$$

By definition, $\sum_k N_k(t)$ is the number of components and $\sum_k N_k(t)\, k$ the total number of vertices, $\sum_k N_k(t)\, k = t$ at any finite time.

Let us write an evolution equation for $N_t(z)$. At time $t+1$, we add the vertex with label $t+1$ which may then be connected to $n_k(t)$ connected components of size k. This creates a new component of size $1+\sum_k n_k(t)\, k$, but also removes $n_k(t)$ components of size k. Thus, at time $t+1$ we have:

$$N_k(t+1) = N_k(t) - n_k(t) + \delta_{k;\,1+\sum_p n_p(t)\, p} \tag{6}$$

with $\delta_{j;k}$ the Kronecker symbol. Alternatively,

$$N_{t+1}(z) = N_t(z) - \sum_{k\geqslant 1} n_k(t)\, z^k + z^{1+\sum_k n_k(t)\, k} \tag{7}$$

As is apparent from this formulation, the transition probability from a given $N_t(z)$ to a given $N_{t+1}(z)$ can be given in closed form. To be precise, the admissible $N_t(z)$'s (describing the accessible distributions of components at time t) are polynomials with integral non-negative coefficients,

whose derivative at $z=1$ have value t. Now suppose $N_t(z)$ and $N_{t+1}(z)$ are admissible. If the difference $N_{t+1}(z)-N_t(z)$ cannot be written as $-\sum_{k\geqslant 1} n_k(t)\, z^k + z^{1+\sum_k n_k(t)\, k}$ for some set of nonnegative integers $n_k(t)$, the transition is forbidden. If it can, then the $n_k(t)$'s are uniquely defined and the transition probability is

$$\mathrm{Prob}(N_t(z) \to N_{t+1}(z)) = \prod_k \binom{N_k(t)}{n_k(t)} q_{t+1}^{k(N_k(t)-n_k(t))} (1-q_{t+1}^k)^{n_k(t)}$$

The meaning of this equation is simple. At time $t+1$, the new vertex is added, and for each of the former t points a (biased) coin is tossed to decide the value of the edge variables $\ell_{j,t+1}$. The tossings are independent with the same law, so the probability that the new point does not attach to a given component of size k is q_{t+1}^k, and distinct components are independent. Hence for each k one makes $N_k(t)$ independent Bernoulli trials with failure probability q_{t+1}^k, and the transition from $N_t(z) \to N_{t+1}(z)$ requires exactly $n_k(t)$ successes. This shows that the graph evolution is a (time inhomogeneous) Markov process on the space of components distributions, a fact that we shall use for the purpose of numerical simulations.

This explicit representation of the transition probability could be used to average Eq. (7). Alternatively, one can represent the number $n_k(t)$ of components of size k which are connected to the new vertex in terms of the edge variables $\ell_{i,j}$ as

$$n_k(t) = \sum_{[k]=1}^{N_k(t)} \left[1 - \prod_{j\in[k]} (1-\ell_{j,t+1}) \right]$$

where $[k]$ runs over connected components of size k. Since the edge variables $\ell_{j,t+1}$ are statistically independent of the earlier edge variables, $\ell_{j,k}$ with $k \leqslant t$, and therefore also independent of the $N_k(t)$'s, we have for any w,

$$\langle w^{n_k(t)} \rangle = \langle [q_{t+1}^k + (1-q_{t+1}^k)\, w]^{N_k(t)} \rangle \tag{8}$$

In particular,

$$\langle n_k(t) \rangle = (1-q_{t+1}^k) \langle N_k(t) \rangle \tag{9}$$

We can now take the average value of Eq. (7) to get

$$\begin{aligned} \langle N_{t+1}(z) \rangle &= \langle N_t(zq_{t+1}) \rangle + z \left\langle \prod_{k\geqslant 1} [q_{t+1}^k + (1-q_{t+1}^k)\, z^k]^{N_k(t)} \right\rangle \\ &= \langle N_t(zq_{t+1}) \rangle + z q_{t+1}^t \left\langle \prod_{k\geqslant 1} [1 + (q_{t+1}^{-k} - 1)\, z^k]^{N_k(t)} \right\rangle \end{aligned} \tag{10}$$

Assume, but this will be justified later using the asymptotic behavior of the p_j's, that at large time and for fixed component size $N_k(t)/t$ is self-averaging, meaning that

$$N_k(t)/t \simeq C_k + o(1), \qquad \text{as} \quad t \to \infty$$

with C_k equals to its averaged value with probability one and with the remaining other $o(1)$ terms random. This in particular implies that for any finite size k the number of connected components of size k scales thermodynamically with the graph size. If both sides of Eq. (10) are expended in powers of z, a given degree involves only components of bounded (t-independent) size. So, order by order in z, self-averaging applies and we conclude that taking $1-q_t \sim \alpha/t$, the following is an accurate approximation for the last term in Eq. (10) at large times:

$$q_{t+1}^t \prod_{k \geqslant 1} [1+(q_{t+1}^{-k}-1)\, z^k]^{N_k(t)} \simeq_{t\to\infty} \exp\left(-\alpha+\alpha \sum_{k\geqslant 1} k z^k C_k\right).$$

The averaged evolution equation (10) then gives a deterministic differential equation

$$0 = -C(z) - \alpha z\, \partial_z C(z) + z \exp(-\alpha + \alpha z\, \partial_z C(z)) \tag{11}$$

for the generating function

$$C(z) \equiv \sum_{k\geqslant 1} z^k C_k.$$

The function

$$z\, \partial_z C(z) = \sum_{k\geqslant 1} z^k P_k, \qquad P_k \equiv k C_k$$

has a direct probabilistic interpretation. Indeed, P_k is the fraction of points in clusters of size k, or equivalently the probability that a randomly chosen vertex belongs to a connected component of size k.

By construction $C(z)$ is a Taylor series with positive coefficients. The series $\sum_k k C_k = \sum_k P_k$ is convergent because it counts the fraction of points in finite clusters which is $\leqslant 1$. As a consequence the radius of convergence of $C(z)$ is at least 1: if we denote by $R(\alpha)$ the radius of convergence $C(z)$, we know that $R(\alpha) \geqslant 1$ for $0 \leqslant \alpha < +\infty$.

We show in Appendix A that analogous methods can be used to describe the connected components of the Erdös-Renyi random graph model.

Most of our subsequent analysis will be based on (11). It recursively determines the C_k's. The first few are:

$$C_1 = \frac{e^{-\alpha}}{(\alpha+1)}, \qquad C_2 = \frac{\alpha e^{-2\alpha}}{(\alpha+1)(2\alpha+1)},$$
$$C_3 = \frac{\alpha^2(6\alpha+5)\, e^{-3\alpha}}{2(\alpha+1)^2\,(2\alpha+1)(3\alpha+1)}, \ldots \tag{12}$$

More generally, the combinations $C_k e^{k\alpha}$ are rational functions in α. They are more efficiently computed using a second order differential equation which follows from derivation of the logarithm of Eq. (11):

$$z\,\partial_z(C+\alpha z\,\partial_z C) = (C+\alpha z\,\partial_z C)(1+\alpha(z\,\partial_z)^2\, C) \tag{13}$$

This leads to

$$(n-1)(1+n\alpha)\, C_n = \sum_{k+l=n} k^2(1+\alpha l)\, C_k C_l \tag{14}$$

A general feature of the recursion relations is that the rational function $C_k e^{k\alpha}$ has no poles except possibly at the points $-1, -1/2, \ldots, -1/k$.

Eq. (10) is not a closed formula for the probability distributions of the number of connected components. But as shown in Appendix B, one may improve slightly the above argument to obtain a closed system (see Eq. (54)) which can be used to prove systematically that the variables C_k are self-averaging, a task we perform for C_1 and C_2 in Appendix B. The proof becomes more and more tedious for large k. But we are confident that the self averaging property is general because of the close relationship with the Erdös-Renyi model presented in Appendix A.

5. PERCOLATION TRANSITION

Up to now, our analysis of connected components has always concentrated on finite components: the thermodynamic limit $t \to \infty$ has been taken while keeping k, the component size, arbitrary but fixed. We have argued that in this regime, the number of components of size k is proportional to t. This has lead to a satisfactory description of C_k. Our arguments did not use any hypothesis on whether or not only finite components play a role in the thermodynamic limit. However, as we observed already, the sum $\sum_k P_k$ measures exactly the fraction of the sites that are either isolated or in components of size 2 or 3 or $\cdots$. To rephrase it more vaguely, $\sum_k P_k$ counts the fraction of vertices that belong to finite connected components.

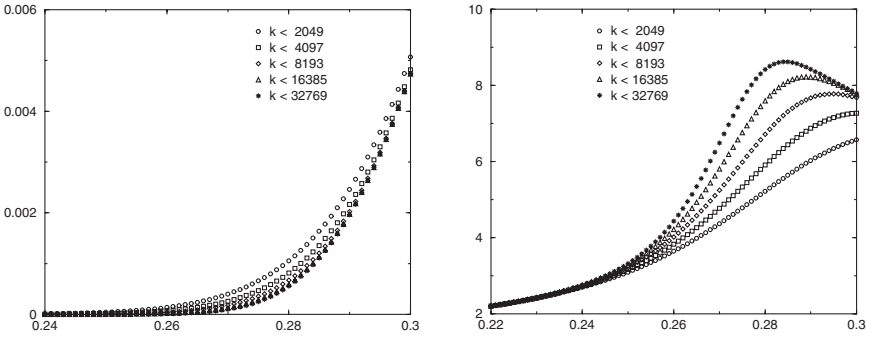

Fig. 1. Partial resummations of P_k and kP_k. Left: The fraction of sites occupied by the giant component. Right: The variation of $\sum kP_k$ close to the transition.

If only clusters of finite size contribute to the thermodynamic limit, then $\sum_k P_k = 1$. Else, by standard physical arguments, a single giant component of size $\sim t(1-\sum_k P_k)$ accounts for the deficit. The giant component is also called the percolating cluster. Its relative size, which we denote by P_∞, is

$$P_\infty \equiv 1 - \sum_{k \geqslant 1} P_k = 1 - \partial_z C(1) \tag{15}$$

To discriminate between the two situations, P_∞ may be computed numerically by evaluating a large number of coefficients P_k using the recursion relation (14). As we shall see later, the convergence of the series is slow for the whole relevant range of α. The result of such a partial summation $\sum_{k \leqslant k_{\max}} P_k$ is plotted in Fig. 1 for $k_{\max} = 2^{11}, \ldots, 2^{15}$. It reveals a phase transition at a value α_c between 0.24 and 0.29, going from a regime where finite components contain all vertices to a regime where they do not. Below 0.24 and above 0.29, the plots corresponding to different values of $k_{\max}$ are hard to distinguish, but in the transition region, large values of k make substantial contributions to $\sum P_k$. The transition is also manifest on an analogous study of $\sum kP_k$. We shall show later that $1/4$ is the exact threshold in this model and that $\sum kP_k$ is discontinuous at the transition.

The growth of P_∞ just above the threshold seems to start with many vanishing derivatives, in strong contrast with what happens in the Erdös-Renyi random graph, for which the growth of the giant component is linear close to the transition. This can be related to the following observations:

– As we have recalled in Appendix A, in the Erdös-Renyi model the components of size k occupy a fraction $\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-k\alpha}$. As a function of α, this

fraction has a single maximum at $\alpha = \frac{k-1}{k}$. These values accumulate at $\alpha = 1^-$, the well-known transition point for the standard random graph. Then for $\alpha \geqslant 1$, the fraction of sites occupied by components of size k decreases with a finite slope for all k's, and so does the sum, so that the growth of the giant component is linear close to the transition.

– In the model studied in this paper, the behavior of $P_k(\alpha)$ as a function of α for generic k is not so easy to get at. However, a simple numerical analysis leads to the following picture: P_1 is a decreasing function of α, but for $k > 1$, P_k has a single maximum, at say α_k. This sequence starts with $\alpha_2 \simeq 0.241$, $\alpha_3 \simeq 0.311$ $\alpha_4 \simeq 0.341$, is maximum for $k = 12$ with $\alpha_{12} \simeq 0.375$ and then decreases very slowly ($\alpha_{100} \simeq 0.338$, $\alpha_{1000} \simeq 0.301$, $\alpha_{10000} \simeq 0.282$), apparently getting closer and closer to $1/4$. At the transition, most $P_k(\alpha)$'s are still increasing, and close enough to the transition a finite but large number of them is still increasing. So subtle compensation mechanisms can take place, leading possibly to the vanishing of (infinitely) many derivatives of P_∞ at $\alpha = 1/4$.

This is confirmed in the following subsections, which are also devoted to a more precise description of the distributions of finite and infinite clusters. To summarize:

$$P_\infty = 0, \qquad \text{if} \quad \alpha \leqslant 1/4,$$
$$P_\infty > 0, \qquad \text{if} \quad \alpha > 1/4.$$

6. BEHAVIOR BELOW THE TRANSITION, $\alpha < \alpha_c$

We turn to the examination of the consequences of Eq. (11), the equation that determines the generating function for the number of clusters of given finite size.

We know that $R(\alpha)$, the radius of convergence of $C(z)$, is at least 1. To analyze the behavior of $C(z)$ around $z = 1$, we define

$$F(\tau) \equiv \alpha - \tau - \alpha\, \partial_\tau Y(\tau), \qquad Y(\tau) = C(e^\tau). \tag{16}$$

From Eq. (11) F satisfies

$$\alpha\,(1 - e^{-F})\, \partial_\tau F = -F - \tau \tag{17}$$

Below the transition, there is no percolating cluster so that $\partial_\tau Y(0) = \sum_k kC_k = 1$ or alternatively,

$$\sum_{k \geqslant 1} P_k = 1, \qquad \alpha < 1/4 \tag{18}$$

This makes clear that the normalized positive numbers P_k are the probabilities for a vertex to be in a connected component of size k. For $F(\tau)$ this translates into the boundary condition

$$F(0) = 0, \qquad \alpha < 1/4$$

6.1. Scaling Laws

We first look for a formal solution $F_{tr}(\tau)$ of Eq. (17) in the form of a Taylor series in τ:

$$F_{tr}(\tau) = -\tau - \alpha \sum_{m \geqslant 1} \frac{\mu_m}{m!} \tau^m \tag{19}$$

The Taylor coefficients are the moments of the measure P_k: $\mu_m = \sum_k k^m P_k$. For example μ_1 is the average proportion of vertices per component. As Eq. (13) the differential equation Eq. (11) may be turned into a second order differential equation for $Y(\tau)$,

$$\partial_\tau Y + \alpha\, \partial_\tau^2 Y = (Y + \alpha\, \partial_\tau Y)\,(1 + \alpha\, \partial_\tau^2 Y) \tag{20}$$

Equation (11) with $\partial_\tau Y(0) = 1$ gives $Y(0) = 1 - \alpha$. Equation (20) then allows us to recursively compute the μ_m's:

$$\mu_1 = \frac{4}{(1+b)^2} = \frac{1 - 2\alpha - \sqrt{1-4\alpha}}{2\alpha^2} \tag{21}$$

$$\frac{1}{2}\,(n - 1 - (n+1)\,b)\,\mu_n = -(1 + \alpha\mu_1)\,\mu_{n-1} - \alpha \sum_{\substack{k+l=n+1 \\ k,l \geqslant 2}} \binom{n}{k} \mu_l (\mu_{k-1} + \alpha\mu_k)$$

with

$$4\alpha = 1 - b^2, \qquad b = \sqrt{1-4\alpha}$$

The square root singularity in the expression of μ_1 indicates that the initial boundary condition (18) becomes pathological at $\alpha = 1/4$ and thus signals the percolation transition. The recursion relation for the higher Taylor coefficients shows that μ_n possesses a pole at $b = (n+1)/(n-1)$. It actually changes sign, from positive to negative, across the pole. As a function of n, the μ_n's have a simple pole at $n_*(b)$,

$$\mu_n \simeq \frac{\text{reg.}}{n - n_*(b)}, \qquad n_*(b) = \frac{1+b}{1-b}, \qquad n \geqslant 2 \tag{22}$$

where the numerator does not vanish at $(n-1)=b(n+1)$. Since $\mu_n = \sum_k k^n P_k$ at least for $n < n_*(b)$, this simple pole codes for the asymptotic behavior of P_k at large k. Indeed, recall that if $P_k \simeq k^{-\nu}$ as $k \to \infty$ then $\sum_k k^x P_k$ diverges as $x \to \nu - 1$ from below with a simple pole at this value. Thus assuming that we may extend the simple pole (22) to non-integers value of n we learn that

$$P_k \simeq_{k\to\infty} \text{const. } k^{-\nu(\alpha)} \tag{23}$$

with

$$\nu(\alpha) = 1 + n_*(b) = \frac{1+\sqrt{1-4\alpha}}{2\alpha} \tag{24}$$

This means in particular that the probability distribution P_k (which describes the proportion of the system occupied by clusters of size k) is large and only the moments μ_n for $n < n_*(b)$, exist.

The fact that the Taylor coefficients μ_n, computed from the recursion relation (21), cannot consistently be interpreted as moments of a probability distribution for $n > n_*(b)$ indicates that the expansion (19) is only up to $o(\tau^{n_*})$ terms. Indeed, the differential equation (17) is compatible with an expansion of $F(\tau)$ for $\tau < 0$, $z = e^\tau < 1$, of the form:

$$F(\tau) = -\tau - \alpha \sum_{q,\, p \geqslant 0} \frac{(-)^q}{q!}\, y_{p,\,q}\, (-\tau)^{q+pn_*(b)},$$

with $y_{0,\,0} = 0$ and $y_{0,\,m} = \mu_m$. As a function of the complex variable z, $\partial_z C(z)$ has thus a branch cut starting at $z = 1$:

$$\partial_z C(z) \simeq \text{const. } (1-z)^{n_*(b)} + \cdots, \qquad \text{around } z = 1 \tag{25}$$

For $n_*(b)$ an integer, this formula should becomes $\partial_z C(z) \simeq \text{const. } (1-z)^{n_*(b)} \log(1-z) + \cdots$. The cut implies the asymptotic behavior (23) for $P_k = \oint_0 \frac{dz}{2i\pi}\, \partial_z C(z)\, z^{-k}$.

At the transition but from below, i.e., $\alpha = 1/4^-$, there are logarithmic correction to the scaling behavior (23), and the branch cut equation is $\partial_z C(z) \simeq \text{const. } (1-z)/\log(1-z) + \cdots$. This ensures that the first moment is finite, and its value, computed from (21), is $\mu_1|_{\alpha = 1/4^-} = 4$

6.2. Scaling Domain, $\alpha < \alpha_c$

The scaling law (23) may be linked to the typical growth rate of large clusters in the system. For concreteness, consider the component of vertex

t' for any given t'. For very large t, the number of arrows emerging from vertex t' grows like $\alpha \log t$. Then we infer that the size of the genealogical tree of t' will grow like t^{α} (under the hypothesis that the genealogical tree is indeed tree-like, a reasonable assumption for small α). This counting of descendants gives a crude lower bound for the size of the connected component of t'. Hence we expect that the system contains components whose sizes grow like a power of t. To estimate this power, we argue as follows.

Consider a given large cluster of size $k(t) \ll t$ at time t:

(i) $k(t+1)-k(t)$ is 0 with probability $q_{t+1}^{k(t)}$ and $1+\sum_p pn_p(t)$ with probability $1-q_{t+1}^{k(t)} \simeq \alpha k(t)/t$ times the probability that vertex $t+1$ connects to $n_p(t)$ clusters of size p apart from the large cluster;

(ii) removing the given large cluster does not change the thermodynamical properties of the graph, so the probability that vertex $t+1$ connects to $n_p(t)$ clusters of size p apart from the large cluster is simply the probability that vertex $t+1$ connects to $n_p(t)$ clusters of size p. Hence for large t, from Eq. (9), $\langle \sum_p pn_p(t) \rangle \simeq \alpha \sum_k kP_k = \alpha\mu_1$;

(iii) suppose we add δt new vertices with $\delta t \ll t$ but $k(t)\,\delta t \gg t$. Between time t and time $t+\delta t$, many new clusters have be connected to the given large components so $1 \ll k(t+\delta t)-k(t) \ll k(t)$, but this has not changed the thermodynamical properties of the graph. Hence we can average the equation in (i) to get a deterministic equation

$$k(t+\delta t)-k(t) \simeq \alpha k(t)(1+\alpha\mu_1)\,\delta t/t.$$

This leads to

$$k(t) \sim t^{1/\nu} \qquad \text{with} \quad 1/\nu = \alpha(1+\alpha\mu_1) = 2\alpha/(1+b). \tag{26}$$

We find that the growth rate of large clusters is universal. As expected, their growth exponent is larger than α, the genealogical tree growth, because it takes not only descendants into account but the whole component. The difference is maximum at the transition, where $\nu = 2$.

The fact that the same exponent, ν, governs the asymptotic behavior of P_k at large k and the size of large clusters for large t can be understood directly as follows. First consider one realization of the random graph for a given t and suppose that there is a single component of maximal size, say K. Observe that $\sum_{k \geqslant l} kN_k(t)$ is by definition the number of points in components of size larger than l. This number is strictly larger than l if $l < K$, but vanishes if $l > K$. So K is characterized by the identity $\sum_{k \geqslant K} kN_k(t) = K$.

From this we infer by taking the average that for large t the relation

$$\sum_{k \geqslant k(t)} k\langle N_k(t)\rangle \approx k(t)$$

gives a sensible characterization for $k(t)$, the order of magnitude of the size of large components in the graph. We write $\sum_{k\geqslant k(t)} k\langle N_k(t)\rangle = t - \sum_{k<k(t)} k\langle N_k(t)\rangle$ and use that for $\alpha < 1/4$, $\sum_k P_k = 1$ to write $\sum_{k\geqslant k(t)} tP_k + \sum_{k<k(t)} tP_k - k\langle N_k(t)\rangle \approx k(t)$. For large $k(t)$, the asymptotics of the first sum is $\sum_{k\geqslant k(t)} tP_k \approx tk(t)^{-\nu+1}$. The second sum is made of finite size corrections. If we assume that these are not too large, we conclude that $tk(t)^{-\nu+1} \approx k(t)$, i.e., that $k(t) \approx t^{1/\nu}$.

In fact, experience from finite size scaling suggests that the two sums give contributions of the same order of magnitude. The idea is that tP_k is the main contribution to $k\langle N_k(t)\rangle$ not only when tP_k is of order t, but even when simply $tP_k \gg 1$ so that self averaging remains valid. This means that $tP_k - k\langle N_k(t)\rangle$ and $k\langle N_k(t)\rangle$ become of the same order of magnitude only when k is so large that $tP_k \approx 1$. So again we see that $k(t)$ is characterized by $tP_{k(t)} \approx 1$, and we conjecture that $tk^{-\nu}$ is a scaling variable.

7. BEHAVIOR ABOVE THE TRANSITION, $\alpha > \alpha_c$

Above the transition there is an giant component. Let P_∞ be its relative size. By definition, $\partial_\tau Y(0) = 1 - P_\infty$ or alternatively

$$\sum_{k\geqslant 1} P_k + P_\infty = 1, \qquad \alpha > 1/4 \tag{27}$$

This makes clear that the P_k's and P_∞ define the probability distribution of vertices among the clusters of different sizes, with P_∞ the probability for a vertex to be in the percolating cluster.

The size of the giant component k_∞ increases linearly with time: $k_\infty \simeq tP_\infty$ for t large. The slope may be evaluated as follows. Imagine adding a new vertex at time $t+1$. It is connected to the percolating cluster with probability $1 - q_{t+1}^{k_\infty}$ so that

$$k_\infty(t+1) \simeq (1 - q_{t+1}^{k_\infty})\left(k_\infty(t) + 1 + \sum_p p n_p(t)\right) + q_{t+1}^{k_\infty} k_\infty(t)$$

where, as in Eq. (6), $n_p(t)$ are the numbers of components of size p connected to the new vertex. This can be rearranged as $k_\infty(t+1) - k_\infty(t) \simeq (1 - q_{t+1}^{k_\infty(t)})(1 + \sum_p p n_p(t))$. The quantity P_∞ is self averaging, so using

$q_t \simeq 1-\alpha/t$ and $k_\infty \simeq tP_\infty$ at large time we infer $P_\infty = (1-e^{-\alpha P_\infty})\langle(1+\sum_p pn_p(t))\rangle$ which by use of Eq. (9) leads to

$$P_\infty = (1-e^{-\alpha P_\infty})(1+\alpha\mu_1) \tag{28}$$

As shown below the above equation is exact, but it does not determine P_∞ as this requires knowing the moment μ_1.

Above the transition, $F(\tau)$ still satisfies the differential equation (17) but with a different boundary condition:

$$F(0) = \alpha P_\infty, \qquad \alpha > 1/4 \tag{29}$$

with P_∞ vanishing at the transition i.e., as $\alpha \to 1/4^+$. This modifies the behavior of its moments. So let us expand $F(\tau)$ in Taylor series around the origin:

$$F(\tau) = \alpha P_\infty - \tau - \alpha \sum_{m\geqslant 1} \frac{\mu_m}{m!}\tau^m$$

with $\mu_m = \sum_k k^m P_k$. Eq. (17) implies Eq. (28) which is thus exact. The second order differential equation (20) does not fix P_∞ but determines recursively the μ_m's which depend parametrically on P_∞.

It remains to decipher what the behavior of the size of the giant component is, at least close to the transition. This will follow from an analysis of the behavior of $F(\tau)$ close to its singularities. As we are going to show, $F(\tau)$ possesses square root branch cut at a point $\tau_c > 0$ but exponentially close to the origin. This means that above the transition the singularity of the function $\partial_z C(z)$ is at $z_c = e^{\tau_c} > 1$, with z_c depending on α, a behavior which has to be compared with the cut at $z = 1$ below the transition, Eq. (25). More precisely, we show that

$$\log P_\infty \sim -\pi/\beta, \qquad \tau_c \sim P_\infty/8e \qquad \text{for} \quad \alpha \to 1/4^+ \tag{30}$$

with β defined by

$$4\alpha \equiv \beta^2 + 1, \qquad \beta \equiv \sqrt{4\alpha - 1}.$$

7.1. Preliminaries

To prove Eq. (30) we shall look at the behavior of F in the neighborhood of three different points:

- at the origin $\tau = 0$ where $F(\tau)$ takes the boundary value (29),
- at the branch point $\tau_c > 0$ with $F(\tau_c) = 0$, $\partial_\tau F(\tau_c) = \infty$ and,
- around the point $\tau_d < 0$ specified by the condition $F(\tau_d) + 2\tau_d = 0$.

Let us first show τ_c and τ_d exist. The function we are interested in is defined by $F = -\tau + \alpha(1 - \sum_{k \geqslant 1} P_k e^{k\tau})$, and satisfies the differential equation Eq. (17). So $\partial_\tau F < 0$ and F decreases from $+\infty$ at $\tau = -\infty$ to 0 at a point τ_c which is non-negative since $F(0) \geqslant 0$. F also satisfies the obvious inequalities $0 < F + \tau < \alpha$ for $\tau < 0$ and $\partial_\tau^2 F < 0$. So there is a single point τ_d, with $-\alpha < \tau_d < 0$, at which $F(\tau_d) + 2\tau_d = 0$ and $F + 2\tau$ has the sign of $\tau - \tau_d$. Both τ_c and τ_d vanish at the transition, $\beta \to 0$.

It turns out that most of these properties could be proved using only the differential equation. More precisely, take any solution E of (17) with an initial condition at $\tau_i < 0$ such that $E(\tau_i) + \tau_i > 0$. The function E can be extended on the right as a positive function on a maximal interval $[\tau_i, \tau_f[$. Then E is strictly decreasing for $\tau > \tau_i$, $\lim_{\tau_f^-} E = 0$, $\tau_f \geqslant 0$, and $\tau_f > 0$ if $\alpha > 1/4$. This discussion emphasizes the intrinsic role played by $\alpha = 1/4$ in this problem.

Before proving Eq. (30) let us get intuition from a simpler, but more universal, version of the differential equation (17). If we look at a region where F is small—and we know that such regions exist generically—we can estimate $1 - e^{-F}$ by $F \simeq f$ with f approximating F. This leads to the simplified, but still non-linear, differential equation

$$\alpha f \partial_\tau f + \tau + f = 0. \tag{31}$$

It turns out that this simpler equation can be solved in closed form. Before doing that, let us further assume that τ is close to τ_c so that f can be neglected compared to τ. Then we get $\alpha f \partial_\tau f \simeq -\tau$, with solution

$$f \simeq \sqrt{(\tau_c^2 - \tau^2)/\alpha}\,,$$

leading to the announced square root singularity. But this does not give informations on the location of the branch point.

Consider now the following functional of an arbitrary function f of τ

$$\frac{1}{2}\log(\alpha f^2 + \tau f + \tau^2) - \frac{1}{\beta}\arctan\left(\frac{\beta f}{f + 2\tau}\right). \tag{32}$$

To fix conventions, we specify the function arc tan by demanding that it is continuous and takes value in $]-\pi/2, \pi/2[$. The total derivative of this functional with respect to τ is

$$\frac{\alpha f \partial_\tau f + \tau + f}{\alpha f^2 + \tau f + \tau^2}.$$

It vanishes if f is a solution of Eq. (31), so we have indeed solved in closed form equation (17) for small F. As this is the domain we are interested in, it is tempting to argue that F and f should exhibit the same singular behaviour. This turns out to be true, but there are some subtleties because the limit $\alpha \to 1/4^+$ is singular: the size of the domain for which f is a uniform approximation to F shrinks to 0.

7.2. Rigourous Estimates

Our strategy is to use the invariant (32) for the approximate equation (31) to derive exact inequalities for F.

Let us observe that the functional (32) is singular at points where $f+2\tau=0$ where it has a jump of amplitude $\pm\pi/\beta$. With this in mind, we define the functional

$$I(\alpha', F(\tau), \tau) \equiv \tfrac{1}{2}\log(\alpha' F^2+\tau F+\tau^2) - \begin{cases} \frac{1}{\beta'} \arctan\left(\frac{\beta' F}{F+2\tau}\right)+\frac{\pi}{\beta'} & \text{if} \quad \tau<\tau_d<0 \\ \frac{\pi}{2\beta'} & \text{if} \quad \tau=\tau_d \\ \frac{1}{\beta'} \arctan\left(\frac{\beta' F}{F+2\tau}\right) & \text{if} \quad \tau_d<\tau<\tau_c \end{cases} \tag{33}$$

In this definition, $\alpha'=(1+\beta'^2)/4>1/4$ is a priori independent of α, the value for which $F(\tau)$ is considered. The functional $I(\alpha', F(\tau), \tau)$ is a smooth function of τ on $]-\infty, \tau_c[$. Using the differential equation for F, its total derivative with respect to τ is found to be

$$\frac{dI}{d\tau}=\frac{\tau+F}{\alpha' F^2+\tau F+\tau^2}\left(1-\frac{\alpha'}{\alpha}\frac{F}{1-e^{-F}}\right). \tag{34}$$

If $\alpha'=\alpha$, the right-hand side is always negative, so I is decreasing, and comparing its values at $\tau<\tau_d$, at τ_d, at 0 and at τ_c we find

$$I(\alpha, F(\tau), \tau)>\log(\beta\,|\tau_d|\,e^{-\pi/2\beta})>\log(\alpha^{3/2}P_\infty)-\frac{1}{\beta}\arctan\beta>\log\tau_c. \tag{35}$$

We know that $\tau_d>-\alpha$, so taking a fixed $\tau<-1/4$, we can take the limit $\alpha\to 1/4^+$. At point τ, F is analytic in α, so $I(\alpha, F(\tau), \tau)+\pi/\beta$ has a finite limit, and we get exponentially small upper-bounds for $|\tau_d|$, P_∞ and τ_c. For instance, $|\tau_d|\,\beta e^{\pi/2\beta}$ is bounded above when $\alpha\to 1/4^+$.

For $\alpha'<\alpha$, $\frac{dI}{d\tau}$ changes sign at a point τ' solution of $1-\frac{\alpha'}{\alpha}\frac{F}{1-e^{-F}}=0$. This point is unique as F is decreasing and goes from $+\infty$ at $-\infty$ to 0 at τ_c. Let us choose α' such that $\tau'\leqslant\tau_d$—and such choices exist. Then $\frac{dI}{d\tau}$ goes from

negative to positive if τ increases so that I increases for τ varying from τ' to τ_c. This leads to lower bounds

$$I(\alpha', F(\tau'), \tau') < \log(\beta' |\tau_d| e^{-\pi/2\beta'}) < \log(\alpha'^{1/2}\alpha P_\infty) - \frac{1}{\beta'} \arctan \beta' < \log \tau_c. \tag{36}$$

Take $\alpha'/\alpha = (e^{2\tau_d} - 1)/2\tau_d$, in such a way that $\frac{dI}{d\tau}$ vanishes exactly at τ_d. Then $\alpha' = \alpha(1 + \tau_d + \cdots)$, and $\beta' = \beta(1 + \tau_d/2\beta^2 + \cdots)$ when $\alpha \to 1/4^+$. Comparing the lower bounds with the upper bounds obtained above, we see that

$$\beta |\tau_d| e^{-\pi/2\beta} \sim P_\infty/8e \sim \tau_c \qquad \text{for} \quad \alpha \to 1/4^+.$$

To get a lower bound for $|\tau_d|$ we take α' such that $\frac{dI}{d\tau}$ vanishes at $\tau' < \tau_d$. Then $-\frac{1}{\beta'} \arctan(\frac{\beta' F}{F + 2\tau}) > 0$ and $\alpha' F^2 + \tau F + \tau^2 = \beta'^2 F^2/4 + (\tau + F/2)^2 \geqslant \beta'^2 F^2/4$ so we have a crude bound $I(\alpha', F(\tau'), \tau') + \pi/\beta' > \log \beta' F(\tau')/2$. Taking for instance $\beta' = \beta - \beta^2$, so that $\beta'/\beta \to 1$ and $F(\tau') \to 0$ as $\beta \to 0^+$. Reporting in $\frac{dI}{d\tau} = 0$ yields $F \sim 4\beta^3$. Thus, from (36), $|\tau_d| \beta^{-3} e^{\pi/2\beta}$ is bounded below when $\alpha \to 1/4^+$.

The above upper and lower bounds then imply

$$2 \log \tau_d \sim \log \tau_c \sim \log P_\infty \sim -\pi/\beta \qquad \text{as} \quad \beta \to 0^+.$$

We restate the physically most important result: up to an algebraic prefactor, the size of the percolation cluster close to percolation is

$$P_\infty \propto e^{-\pi/\sqrt{4\alpha - 1}}. \tag{37}$$

In fact, we have obtained a better estimate. We expect that $\tau_d \beta^{1-\gamma} e^{\pi/2\beta}$ as a finite limit for $\beta \to 0^+$ for a certain γ, or equivalently

$$P_\infty \sim \text{const } \beta^\gamma e^{-\pi/2\beta}, \qquad \beta \to 0^+. \tag{38}$$

We have proved that if γ exists, $0 \leqslant \gamma \leqslant 4$. The asymptotic behavior of the probabilities P_k is not the same below and above the transition as $C(z)$ does not have the same analytical properties on the two sides of the transition.

Since above the transition the branch point is located at $z_c = \exp \tau_c > 1$, the P_k's now decrease exponentially. More precisely, $\partial_z C(z)$ possesses a square root branch point at z_c,

$$\partial_z C(z) = \text{const.} \sqrt{z_c - z} + \cdots$$

so that $P_k = \oint \frac{dz}{2i\pi} z^{-k} \partial_z C(z)$ behave as

$$P_k \simeq_{k\to\infty} \text{const.}\, k^{-3/2} z_c^{-k} \simeq \text{const.}\, k^{-3/2} e^{-k\tau_c}, \tag{39}$$

to be compared with Eq. (23).

7.3. Universality

We give now a universality argument suggesting that $\gamma = 0$ and this is confirmed by solving numerically the differential equation (17).

Our invocation of universality rests on Eq. (39), in which τ_c, or equivalently (in the vicinity of the critical point) $P_\infty/8e$, controls the exponential decrease of the P_k's and plays the role of a mass gap. This mass gap is exponentially small close to the transition, and by analogy we may argue that increasing β is a marginally relevant perturbation of the percolating critical point. Introducing the Wilson-Callan-Symanzik beta function $B(\beta)$, we expect a relation

$$\tau_c \simeq P_\infty/8e \simeq \exp \int^\beta \frac{d\beta'}{B(\beta')}. \tag{40}$$

Comparison with our formula gives $B(\beta) = \beta^2/\pi + B_3\beta^3 + \cdots$. It is known from field theory that the coefficient B_3, which dictates the exponent of the algebraic prefactor, is universal.

Such universal features are controlled by the continuum limit. In our framework, the continuum region is reached at small τ, and small F, and we expect that the continuum limit is governed by (31). Note that both τ_c and $|\tau_d|$ are exponentially small close to the transition, and $F(\tau)$ remains small for τ between these points. This a posteriori justifies looking at the approximate equation (31). Consequently, as far as universal quantities are concerned, the inequalities in (35) can be replaced by equalities. This leads to $\gamma = 0$, or $B_3 = 0$, i.e.,

$$P_\infty \sim \text{const}\, e^{-\pi/2\beta}$$

as announced.

As universality could suggest, Eq. (31) turns out to describe the continuum limit for a larger class of evolving networks than just the specific one we are studying. This is the case for the model studied in ref. 9. We refer to the original paper for the definitions. It suffices to say that Eq. (17) is replaced by

$$2\,\delta S\,\partial_\tau S = -S - (e^\tau - 1).$$

In this equation, $1-S$ is a generating function, the coefficient of $e^{k\tau}$ giving the fraction of points in components of finite k, so $S(0)$ is the fraction of sites occupied by the giant component. The parameter δ is the average number of edges created at each time step, so this is is precisely the equivalent of our α. To study S close to $\tau = 0$, the approximation is $e^{\tau}-1 \sim \tau$, and we retrieve (31), with α replaced by 2δ. The percolation threshold is $\delta = 1/8$ (as expected, percolation thresholds are not universal), and the size of the infinite component behaves like $e^{-\pi/2\sqrt{2}(\delta-1/8)^{1/2}}$. The prefactor $\pi/2\sqrt{2} \simeq 1.111$ compares quite well with the numerical value 1.132 ± 0.008 obtained in the original paper. This fact was already noted in ref. 4.

7.4. Scaling Regime, $\alpha > \alpha_c$

We have accumulated evidence that the percolation phase transition is very similar to the Kosterlitz-Thouless phase transition in the XY model. For instance, the probabilities P_k follow the scaling laws (23) below the transition with scaling exponents varying continuously with the parameter α while they decrease exponentially above the transition.

The discussion of universality in the previous section suggests to look for scaling functions describing the neighborhood of the critical point. For instance, from Eq. (20) it follows that the moments μ_m, $m \geqslant 2$, diverge as $P_\infty \to 0$ with

$$\mu_m \simeq_{\alpha \to 1/4^+} \frac{g_m}{P_\infty^{m-1}}, \qquad g_1 = 12, \quad g_2 = 64, \ldots$$

In particular, μ_1 is discontinuous at the transition, jumping from 4 to 12. These scaling relations imply that the function

$$G(x) \equiv P_\infty^{-1} F(xP_\infty)$$

has a finite limit, denoted by $G_c(x)$, as $\alpha \to 1/4^+$ for any fixed x,

$$G_c(x) = \alpha - x - \alpha \sum_{m \geqslant 1} \frac{g_m}{m!} x^m \tag{41}$$

The differential equation (17) at $\alpha = 1/4^+$ then reduces to $G_c\, \partial_x G_c + 4(G_c + x) = 0$. With the boundary condition $G_c(0) = 1/4$, this is integrated as:

$$\log(4(G_c + 2x)) + \frac{2x}{G_c + 2x} = 0 \tag{42}$$

The left-hand side is very reminiscent of the limit $\alpha \to 1/4$ of Eq. (32). As a consequence $G_c(x)$ has a square root branch point at $x_c = 1/8e$ at which G_c vanishes and $g_m/m! \simeq m^{-3/2}(8e)^m$ for m large. Actually, a more precise computation based on the Lagrange formula presented in Appendix C yields the exact value of the scaling coefficients g_m:

$$g_{m+1} = m^m\, 8^{m+1}, \quad m \geqslant 1. \tag{43}$$

8. CHRONOLOGICAL PROFILES

One can make a direct counting of the average number of connected components in the random graph at time t which are copies of a given finite labeled graph. This will allow us to retrieve the results of Section 4. But this also leads to a detailed local-in-time description of the connected components which illustrates the consequences of the chronological memory of our model.

8.1. Tree Distributions

Let G be a labeled graph with vertices $1, 2, \ldots, k$. We let m_j be the number of edges connecting vertex j to a vertex with smaller label, so that $m = m_1 + \cdots + m_k$ is the number of edges of G. We look for the average number of increasing maps v from $[1,\ldots,k]$ to $[1,\ldots,t]$ such that the vertices $v_1,\ldots,v_k$ span a connected component of the random graph isomorphic to G. This number is the average number of connected components isomorphic to G in the random graph.

By the rules of construction of the random graph, the probability that the vertices $v_1,\ldots,v_k$ with $1 \leqslant v_1 < \cdots < v_k \leqslant t$ span a connected component of the random graph isomorphic to G is

$$\prod_{i=1}^{k} \left((1-q_{v_i})^{m_i}\, q_{v_i}^{v_i - 1 - m_i} \prod_{v_i < w_i < v_{i+1}} q_{w_i}^{i} \right)$$

with the convention $v_{k+1} \equiv t+1$.

The average we look for is obtained by summing this expression over the v_i's. For large t, using the asymptotic behavior $p_j \sim \alpha/j$ for large j, the sum can be formally reinterpreted as a Riemann sum, leading to a contribution

$$t^{k-m} e^{-k\alpha} \alpha^m \int_{0 \leqslant \sigma_1 \leqslant \cdots \leqslant \sigma_k \leqslant 1} d\sigma_1\, \sigma_1^{\alpha - m_1} \cdots d\sigma_k\, \sigma_k^{\alpha - m_k}.$$

In deriving this formula, we have not treated carefully the contribution of small values of the v_i's. This is reflected in the fact that the integral can be divergent if some m_i's are too large. However, the prefactor t^{k-m} is the sign that in this case the sum over the v_i's is nevertheless negligible in the large t limit, as a more careful treatment would show.

The most salient feature of this formula is that only connected graphs with $k = m+1$ give a contribution proportional to t, i.e., contribute to C_k. Since k is the number of vertices and m the number of links, this relation characterizes trees as follows from the Euler formula: only trees contribute to the thermodynamic limit. A given tree on k vertices with incoming degrees m_i, $i = 1,\ldots, k$ gives a contribution

$$\frac{e^{-k\alpha}\alpha^{k-1}}{(\alpha+1-m_1)(2\alpha+2-m_1-m_2)\cdots(k\alpha+k-m_1-\cdots-m_k)} \tag{44}$$

to C_k. Observe that for a tree, $m_1+\cdots+m_i \leqslant i-1$ for $i = 1,\ldots, k$ so that all integrals are well-defined and finite for real non-negative α. It is amusing to note also that the contribution of a single tree of size k can contain poles in α at values that are not in the list $-1, -1/2,\ldots, -1/k$. These poles have to cancel between different trees in the sum over trees of size k, because we know that they are absent in C_k. But we have no simple explanation for this cancellation.

This explicit formula makes it easy to show that $C(z)$ is (complex) analytic in α in a neighborhood of $]0, +\infty[$ for every z such that $|z| < 1$. Indeed, we know that for $\alpha \in]0, +\infty[$, and $|z| < 1$ the series $\sum_k P_k(\alpha)\, z^k$ is absolutely convergent. But $P_k(\alpha)$ is a sum of non-negative contributions, each tree giving a contribution of the form (44). Now suppose that α is complex with positive real part. For each tree contribution and for fixed $\Re\alpha$, the modulus of

$$\frac{e^{-k\alpha}}{(\alpha+1-m_1)(2\alpha+2-m_1-m_2)\cdots(k\alpha+k-m_1-\cdots-m_k)}$$

is maximal when $\Im\alpha = 0$ and then the expression is real and positive. This is because the statement is true for every factor. Taking the sum over trees we infer that $P_k(\alpha) \leqslant P_k(\Re\alpha)(\frac{|\alpha|}{\Re\alpha})^{k-1}$. So the series $\sum_k P_k(\alpha)\, z^k$ is absolutely convergent if $|z\alpha| < \Re\alpha$. This equation defines, for fixed $|z| < 1$, a neighborhood of $\alpha \in]0, +\infty[$ in which we have an absolutely convergent sum of analytic functions analytic of α. Hence the sum is analytic in α as claimed.

To resum more explicitly the contribution of all trees of a given size, we need the generating function for labeled trees with given incoming degrees. Suppose more generally that we give a weight x_i for each edge

leaving vertex i (i.e., connecting i to a $j > i$) and a weight y_i for each edge entering vertex i (i.e., connecting i to a $j < i$). The generating function $\mathscr{T}$ for weighted trees on n vertices factorizes nicely as

$$\begin{aligned}\mathscr{T} = x_1(&y_2x_1 + y_2x_2 + y_3x_2 + \cdots + y_kx_2)\\ &(y_3x_1 + y_3x_2 + y_3x_3 + y_4x_3 + \cdots + y_nx_3)\cdots\\ &(y_{n-1}x_1 + \cdots + y_{n-1}x_{n-1} + y_nx_{n-1})\, y_n.\end{aligned}$$

This generalization of the famous Caley tree formula[5] implies it immediately. It seems to be little known, although it is implicit in the mathematical literature.[2, 6] Gilles Schaeffer[10] provided us with a clean proof using a refined version of one of the standard proofs of the Caley tree formula, putting trees on k vertices in one to one correspondence with applications from $[1,\ldots,k]$ to $[1,\ldots,k]$ fixing 1 and k, see, e.g., ref. 11.

This formula can be specialized to $x_i = 1$ and $y_i = 1/\sigma_i$ for $i = 1,\ldots,k$ to give

$$\begin{aligned}C_k = e^{-k\alpha}\alpha^{k-1}\int_{0\leqslant\sigma_1\leqslant\cdots\leqslant\sigma_k\leqslant 1} &d\,\sigma_1\cdots d\,\sigma_k\,(\sigma_1\cdots\sigma_k)^{\alpha}\\ &\left(\frac{2}{\sigma_2}+\frac{1}{\sigma_3}+\cdots+\frac{1}{\sigma_k}\right)\left(\frac{3}{\sigma_3}+\frac{1}{\sigma_4}+\cdots+\frac{1}{\sigma_k}\right)\cdots\left(\frac{k-1}{\sigma_{k-1}}+\frac{1}{\sigma_k}\right)\frac{1}{\sigma_k}.\end{aligned} \tag{45}$$

If one integrates only over a subset of the σ's, one gets marginal distributions. For instance, for $k = 1$, we get that in the thermodynamic limit the fraction of sites with age close to $t\sigma$ that are isolated is $e^{-\alpha}\sigma^{\alpha}$. For $k = 2$, if we integrate over σ_2, we get that the fraction of sites with age close to $t\sigma$ that are the older vertex of a tree on two vertices is $e^{-2\alpha}(\sigma^{\alpha} - \sigma^{2\alpha})$ while if we integrate over σ_1, we get that the fraction of sites with age close to $t\sigma$ that are the younger vertex of a tree on two vertices is $e^{-2\alpha}\sigma^{2\alpha}\frac{\alpha}{\alpha+1}$. The sum, $e^{-2\alpha}(\sigma^{\alpha} - \sigma^{2\alpha}\frac{1}{\alpha+1})$, gives the probability that a site with age close to $t\sigma$ belongs to a tree on 2 vertices.

Our explicit representation in terms of trees shows that in this model, and at least for questions concerning connected components, the thermodynamic limit applies not only to the full system, but also to slices of fixed relative age $\sigma = t'/t$. In the next section we shall study σ-dependent profiles.

For small k's, we have done all the integrals and checked the agreement with the value of C_k obtained by the recursion relation. But a general proof valid for all k's is lacking.

[5] Which states that there are n^{n-2} labeled trees on n vertices.

8.2. Dating Finite Components

To illustrate the evolving nature of our model, we now determine the local-in-time distribution of the cluster sizes. This means determining for any given age interval what is the proportion of vertices of these ages which belong to clusters of given size.

Define $p_k(t, t')$ to be the probability that vertex t' belongs to a component of size k at time t. Guided by the previous tree representation, we infer that in the thermodynamical limit $\sum_{t' \in [t\sigma, t(\sigma+d\sigma)]} p_k(t, t') \simeq t\rho_k(\sigma)\, d\sigma$, with $\rho_k(\sigma)$ a deterministic function. By construction, $\int_0^1 \rho_k(\sigma)\, d\sigma = kC_k$, the total fraction of points that belong to components of size k.

The reasoning leading to (11) can be generalized: one writes down a recursion relation for $p_k(t+1, t')$ and then takes the average, a step justified by the explicit tree representation. The event that vertex t' belongs to a component of size k at time $t+1$ is the exclusive union of several events.

(i) Vertex t' belonged to a component of size k at time t and this component is not linked to the new vertex $t+1$. This has probability $p_k(t, t')\, q_{t+1}^k$.

(ii) Vertex t' belonged to a component of size $l < k$ at time t, this component is linked to the new vertex $t+1$, and together with the other components linked to $t+1$ (say $n_m(t)$ components of size m), it builds a component of size $k = l+1+\sum_m mn_m(t)$. This has probability

$$p_l(t, t')(1-q_{t+1}^l) \prod_m \binom{N_m(t)-\delta_{m,l}}{n_m(t)} q_{t+1}^{m(N_m(t)-\delta_{m,l}-n_m(t))} (1-q_{t+1}^m)^{n_m(t)}$$

To perform the explicit sum over l and the $n_m(t)$'s, we introduce again generating functions and set $p_{t,t'}(z) = \sum_l p_l(t, t')\, z^l$. This leads to

$$p_{t+1,t'}(z) = p_{t,t'}(zq_{t+1}) + \left(\sum_l p_l(t, t') \frac{z^l(1-q_{t+1}^l)}{q_{t+1}^l + z^l(1-q_{t+1}^l)} \right) \prod_m (q_{t+1}^m + z^m(1-q_{t+1}^m))^{N_m(t)}$$

In the large t limit this complicated formula simplifies if we use again the hypothesis of self-averaging and asymptotic independence. Defining

$$\rho(\sigma, z) = \sum_k \rho_k(\sigma)\, z^k, \tag{46}$$

this leads to

$$\sigma\, \partial_\sigma \rho = \alpha(1 - ze^{-\alpha+\alpha z\, \partial_z C})\, z\, \partial_z \rho.$$

Together with the sum rule relating ρ to C, this fixes completely the profiles $\rho_k(\sigma)$. A relation between $\rho(1, z)$ and $C(z)$ is obtained by integrating (8.2) for σ between 0 and 1. Using the defining equation for $C(z)$, Eq. (11), this leads to $\rho(1, z) = (1+\alpha z\partial_z)\, C(z)$. Thus, we can summarize our knowledge on component time profiles with the four relations:

$$\begin{aligned} \sigma\, \partial_\sigma \rho(\sigma, z) &= \alpha(1-\rho(1, z))\, z\, \partial_z \rho(\sigma, z) \\ \int_0^1 d\sigma\, \rho(\sigma, z) &= z\, \partial_z C(z) \\ \rho(1, z) &= z e^{-\alpha+\alpha\int_0^1 \rho(\sigma, z)\, d\sigma} \\ \rho(1, z) &= (1+\alpha z\, \partial_z)\, C(z). \end{aligned} \tag{47}$$

Expansion in powers of z leads to recursion relations for the $\rho_k(\sigma)$'s. They can be shown to be alternating polynomials of degree k in $x = \sigma^\alpha$ with vanishing constant coefficient. The first few polynomials are

$$\begin{aligned} \rho_1 &= e^{-\alpha} x \\ \rho_2 &= e^{-2\alpha}\left(x - \frac{x^2}{\alpha+1}\right) \\ \rho_3 &= e^{-3\alpha}\left(\frac{3\alpha+2}{2(\alpha+1)}\, x - \frac{2}{\alpha+1}\, x^2 + \frac{3\alpha+2}{2(\alpha+1)^2\,(2\alpha+1)}\, x^3\right) \end{aligned}$$

Again, we have checked that for small k's the values of ρ_k as computed from iterated tree integrals and from the generating function coincide, but we have no general proof.

As an application, let us look at the profile of vertices of relative age close to σ which are the youngest in components of size k. These are vertices that created, when they appeared, a component of size k—this happens with probability $\rho_k(1) = (1+k\alpha)\, C_k$—which was then left untouched for the rest of the evolution—this happens with probability $\sigma^{k\alpha}$. So the distribution of vertices that are the youngest in their connected (finite) component is

$$\sum_k (1+k\alpha)\, C_k \sigma^{k\alpha} = (1+\alpha z\partial_z)\, C(\sigma^\alpha). \tag{48}$$

In this expression, the formal parameter z has been replaced by σ^α.

The giant component, when it exists, also has a well-defined profile to which we turn now.

8.3. Dating the Percolating Cluster

We now determine the profile of the giant component.

So let $\rho_\infty(\sigma)\, t\, d\sigma$ be the fraction of vertices whose ages are between σt and $(\sigma + d\sigma)\, t$ which belong to the giant component. By definition $\int_0^1 d\sigma\, \rho_\infty(\sigma) = P_\infty$. We know from the tree representation that for finite clusters the thermodynamic limit applies not only on the full graph, but also on time slices. The giant component is the complement of finite components, so we expect that the density $\rho_\infty(\sigma)$ is self-averaging as is the size of the percolating cluster. To derive an equation fixing this density we look for the probability for a site of age $j = \sigma\, t$ not to be in the percolating cluster. On the one hand, by definition of the density this probability is

$$\mathrm{Prob}(j = \sigma\, t \notin [k_\infty]) = 1 - \rho_\infty(\sigma)$$

On the other hand, this probability may be evaluated by demanding the vertex j not to be connected to the older and younger vertices of the percolating cluster:

$$\mathrm{Prob}(j = \sigma\, t \notin [k_\infty]) = \prod_{k<j,\, k \in [k_\infty]} \left(1 - \frac{\alpha}{j}\right) \prod_{k>j,\, k \in [k_\infty]} \left(1 - \frac{\alpha}{k}\right)$$

At large time, the first above product converges to $\exp(-\alpha \int_0^\sigma d\zeta\, \frac{\rho_\infty(\zeta)}{\sigma})$ and the second to $\exp(-\alpha \int_\sigma^1 d\zeta\, \frac{\rho_\infty(\zeta)}{\zeta})$. So we get the relation:

$$\begin{aligned} 1 - \rho_\infty(\sigma) &= \exp\left(-\alpha \int_0^\sigma d\zeta\, \frac{\rho_\infty(\zeta)}{\sigma} - \alpha \int_\sigma^1 d\zeta\, \frac{\rho_\infty(\zeta)}{\zeta} \right) \\ &= \exp\left(-\alpha \int_0^1 d\zeta\, \rho_\infty(\zeta) \min\left(\frac{1}{\sigma}, \frac{1}{\zeta}\right) \right) \end{aligned} \tag{49}$$

As ρ_∞ is positive, $\int_0^1 d\zeta\, \rho_\infty(\zeta) \min(\frac{1}{\sigma}, \frac{1}{\zeta})$ is an increasing function of σ. So $\rho_\infty(\sigma)$ is decreasing and has a right limit at 0. Unless $\rho_\infty \equiv 0$ (i.e., $\alpha \leqslant 1/4$), the integral has a logarithmic divergence, and henceforth $\rho_\infty(0) = 1$. More precisely,

$$\rho_\infty(\sigma) = 1 - \sigma^\alpha \exp\left(\alpha \int_0^1 d\zeta\, \frac{1 - \rho_\infty(\zeta)}{\zeta} \right) + \cdots \tag{50}$$

This means that the early vertices belong to the giant component with probability 1.

On the other hand, by definition $\int_0^1 d\zeta\ \rho_\infty(\zeta) = P_\infty$ so taking $\sigma = 1$ in (49) we get that

$$\rho_\infty(1) = 1 - e^{-\alpha P_\infty}. \tag{51}$$

This means that the late vertices always belong to the giant component with a non vanishing probability—although this probability is exponentially small close to the threshold. Hence the giant component invades all time slices above the threshold, and the term percolation transition is appropriate even with this unusual interpretation. Our results are illustrated on Fig. 2.

To conclude this discussion, let us observe that ρ_∞ can be expressed in terms of the profiles of finite components that we studied before:

$$\rho_\infty(\sigma) = 1 - \sum_k \rho_k(\sigma) = 1 - \rho(\sigma, z = 1)$$

with $\rho(\sigma, z)$ defined in Eq. (46). We end up with yet another equation constraining $\rho(\sigma, z)$,

$$\rho(\sigma, z = 1) = \sigma^\alpha \exp\left(-\alpha + \int_0^1 d\zeta\ \rho_\infty(\zeta) \min\left(\frac{1}{\sigma}, \frac{1}{\zeta}\right)\right).$$

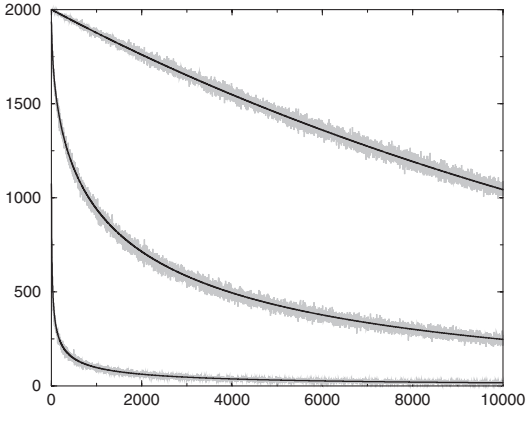

Fig. 2. The analytic result (solid lines) for the profile of the giant component compared to numerical simulations (gray clouds) on 2000 random graphs of size 10000. From top to bottom, the values of α are 1, 1/2 and 1/3.

If we expand $\rho(\sigma, z=1)$ in powers of $\sigma^{n\alpha}$ as $\rho(\sigma, z=1) = \sum_{k\geqslant 1} D_n(1+n\alpha)\,\sigma^{n\alpha}$ and define $D(z) = \sum_{k\geqslant 1} \frac{D_n}{n} z^n$, we can check that

$$z\,\partial_z D(z) + \alpha(z\partial_z)^2\, D(z) = z\exp(-\alpha - D(z) + D(1) + z\,\partial_z D(1)) \qquad (52)$$

In particular, by differentiation this implies that

$$(n-1)(1+n\alpha)\, D_n = -\sum_{k+l=n} k^2(1+\alpha l)\, D_k D_l \qquad (53)$$

The resemblance with Eqs. (11) and (14) for the function $C(z)$ is rather striking, but we have not been able to use this more deeply. One difference is that (11) determines C_1 from scratch, whereas (52) does it in a two step process. If d_n is the solution of this equation satisfying $d_1 = 1$, then $D_n = d_n D_1^n$, and then D_1 has to satisfy

$$(1+\alpha)\, D_1 = \exp\left(-\alpha + \sum_{n\geqslant 1} \frac{1+n\alpha}{n}\, d_n D_1^n\right)$$

Another difference is that the sequence D_n alternates in sign. Again, the formal parameter z receives a simple physical interpretation $z = \sigma^\alpha$ where σ labels the relative date of birth of vertices.

9. CONCLUSIONS

In this study, we have described detailed global and local-in-time features of evolving random graphs with uniform attachment rules.

Concerning global properties, we have shown that the model has a percolation phase transition at $\alpha = 1/4$. Below the transition, the system contains clusters whose sizes scale like $t^{(1-\sqrt{1-4\alpha})/2}$. Above the transition, a single component, the giant component, grows steadily with time. We have shown that, close to the threshold, the fraction of sites in the giant component has an essential singularity and behaves as $e^{-\pi/\sqrt{4\alpha-1}}$. The behaviors below and above the transitions are strongly reminiscent of the two dimensional XY model, so that our model can be interpreted as some algorithmic equivalent of it, but it seems unlikely that a direct connexion exists. This analogy and further scaling properties we present call for an alternative renormalization group approach to the transition.

By describing local-in-time profiles, we have shown that they offer a more accurate vision of the specificities of evolving graphs. It would be desirable to generalize this approach by answering the following question: assume that a procedure to assign ages to the vertices of some evolving

graphs has been given, what informations on the microscopic evolution rules of these graphs can be decoded from the knowledge of local-in-time statistics ?

APPENDIX A

We show how our arguments of Section 4 can be modified to describe the Erdös-Renyi random graph model. In this model, one starts with n points, and any two points are connected by an edge with probability α/n (so in this model all the points are equivalent). Then a limit $n \to \infty$ is taken. This famous model describes a static graph, but it can also be rephrased as an evolving graph in the following way: set $t = n$ and suppose that points are added one by one, from 1 to t, each new point connecting to any previous one with probability α/t. From this point of view, it can be seen that looking only at the first t' vertices ($t' \leqslant t$) amount to look at an Erdös-Renyi random graph with a modified connectivity parameter $\alpha' = \alpha t'/t$. To get a recursion relation, we start from an Erdös-Renyi random graph of size t with connectivity parameter α. We add vertex $t+1$ and connect any older vertex to it with probability $q_{t+1} = \alpha/t$, so that the effective connectivity parameter for the graph on $t+1$ vertices is $\alpha(t+1)/t$. Then the derivation proceeds as before, with the little proviso that in Eq. (10), on the left-hand side $\langle N_{t+1}(z) \rangle$ has an effective connectivity parameter $\alpha(t+1)/t$ instead of α. So when we take the thermodynamic limit, nothing changes on the right-hand side of (10), but an additional term $\alpha \partial_\alpha$ contributes to the left-hand side. If c denotes the analog of C but for the Erdös-Renyi random graph, then

$$\alpha\, \partial_\alpha c = -c - \alpha z\, \partial_z c + z \exp(-\alpha + \alpha z\, \partial_z c).$$

This little modification in the equation has drastic consequences. The new equation has a single solution regular at $\alpha = 0$, namely

$$c = \sum_{k \geqslant 1} \frac{k^{k-2}}{k!} \alpha^{k-1} e^{-k\alpha} z^k$$

which is the well-known result. We conclude that our self-averaging hypothesis is valid for the Erdös-Renyi model. This makes it more plausible that it works for our original model as well, a fact also confirmed by numerical simulations.

APPENDIX B

Our goal is to prove the following formula and to investigate a few of its consequences. We claim that

$$\left\langle \prod_{k\geqslant 1} w_k^{N_k(t+1)} \right\rangle = \oint_0 \frac{d\xi}{2i\pi} \sum_{j\geqslant 1} \frac{w_j}{\xi^j} \left\langle \prod_{m\geqslant 1} [q_{t+1}^m w_m + (1-q_{t+1}^m)\, \xi^m]^{N_m(t)} \right\rangle \tag{54}$$

where the contour integral is around the origin. It is a Fokker-Planck equation for the Markov process formed by the $N_k(t)$'s, which could be used to prove systematically that the variables C_k are self-averaging, a task we perform for C_1 and C_2 at the end of this appendix.

We start from Eq. (6) which may be rewritten as

$$\left\langle \prod_k w_k^{N_k(t+1)} \right\rangle = \left\langle \prod_j w_j^{N_j(t)-n_j(t)}\, w_j^{\delta_{j;\,1+\sum_p pn_p(t)}} \right\rangle$$

To compute the last term we insert the tautological identity

$$1 = \sum_{k\geqslant 1} \delta_{k;\,1+\sum_p pn_p(t)}$$

in the r.h.s. to get

$$\left\langle \prod_k w_k^{N_k(t+1)} \right\rangle = \sum_{k\geqslant 1} w_k \left\langle \prod_j w_j^{N_j(t)-n_j(t)}\, \delta_{k;\,1+\sum_p pn_p(t)} \right\rangle$$

To compute the r.h.s. expectation value we use a contour integral representation of the Kronecker symbol

$$\delta_{k;\,1+\sum_p pn_p(t)} = \oint_0 \frac{d\xi}{2i\pi}\, \xi^{-k+\sum_p pn_p(t)}$$

This yields

$$\left\langle \prod_k w_k^{N_k(t+1)} \right\rangle = \sum_{k\geqslant 1} \oint_0 \frac{d\xi}{2i\pi} \left(\frac{w_k}{\xi^k}\right) \left\langle \prod_p (\xi^p/w_p)^{n_p(t)}\, w_p^{N_p(t)} \right\rangle$$

The r.h.s. can now be computed using Eq. (8) and gives Eq. (54).

This computation has a rather simple combinatorial reinterpretation. We set

$$F_t(w_1, w_2, \ldots) = \left\langle \prod_k w_k^{N_k(t)} \right\rangle,$$

and observe that going from time t to $t+1$, we add vertex $t+1$ and edges from the rest of the graph to $t+1$. Suppose that the component of vertex $t+1$ has size say k. This component was build by "eating" some components of the graph at time t. A component of size j is swallowed with probability $1-q_{t+1}^j$ and survives with probability q_{t+1}^j. On the other hand, if one expands

$$F_t(w_1 q_{t+1} + \xi(1-q_{t+1}),\, w_2 q_{t+1}^2 + \xi^2(1-q_{t+1}^2),\ldots)$$

in powers of ξ, the term of degree $k-1$ enumerates all the possibilities to build the component of vertex $t+1$ with the correct probability. Defining $\hat{w}_l(\xi) \equiv w_l q_{t+1}^l + \xi^l(1-q_{t+1}^l)$, summation over k gives

$$F_{t+1}(w_1, w_2,\ldots) = \sum_k w_k \oint_0 \frac{d\xi}{2i\pi}\, \xi^{-k} F_t(\hat{w}_1(\xi), \hat{w}_2(\xi),\ldots),$$

as obtained previously.

The Fokker-Planck equation (54) may also be formulated as a difference equation, similar to a discrete Schrodinger equation. Let us for instance specify it for $w_k = \sum_j \zeta_j^k$ with ζ_j, $j=0,\ldots,M$, a set of complex numbers and let

$$\Psi_t(\zeta_0,\ldots,\zeta_M) \equiv \left\langle \prod_k [\zeta_0^k + \cdots + \zeta_M^k]^{N_k(t)} \right\rangle$$

This parametrization is similar to that used in matrix theory where w_k may be thought of as the trace of the k^{th} power of matrix whose eigenvalues are the ζ_j's. The contour integral in Eq. (54) can then be explicitly evaluated by deforming the integration contour to pick the simple pole contributions located at the points $\xi = \zeta_j$. This gives:

$$\Psi_{t+1}(\zeta_0,\ldots,\zeta_M) = q_{t+1}^t \sum_j \zeta_j\, \Psi_t(\ldots,\zeta_j/q_{t+1},\ldots) \tag{55}$$

Equation (54) or (55) may be used to prove that the numbers C_k are self-averaging. Let us choose for example two parameters $\zeta_0 = 1$ and $\zeta_1 = z/t$. Then, only the clusters of size 1 give a non trivial contribution to Ψ_t at large time so that

$$\Psi_t^{(1)}(z) \equiv \Psi_t(1, z/t) = \left\langle \prod_k (1+z^k/t^k)^{tC_k} \right\rangle \simeq_{t\to\infty} \langle e^{zC_1} \rangle$$

At large time, the difference equation (55) then reduces to a differential equation for $\Psi_t^{(1)}(z)$,

$$(1+\alpha)\,\partial_z \Psi_t^{(1)}(z) = e^{-\alpha}\,\Psi_t^{(1)}(z)$$

It implies that $\log \Psi_t^{(1)}(z)$ is linear in z at large time which means that C_1 is self-averaging and stationary at large time. More precisely, integrating the above equation gives:

$$\Psi_t^{(1)}(z) = \exp(zC_1(\alpha)), \qquad C_1(\alpha) = \frac{e^{-\alpha}}{\alpha+1}$$

in agreement with Eq. (12).

Similarly, to prove that C_2 possesses a finite self-averaging limit as $t \to \infty$ we choose three parameters $\zeta_0 = 1$, $\zeta_1 = -\zeta_2 = \sqrt{z/t}$ so that only clusters of size two survive in Ψ_t at large time,

$$\Psi_t^{(2)}(z) \equiv \Psi_t(1, \sqrt{z/t}\,, -\sqrt{z/t}) \simeq_{t\to\infty} \langle e^{2zC_2} \rangle$$

Eq. (55) then gives a first order differential equation for $\Psi_t^{(2)}(z)$ which implies that

$$\Psi_t^{(2)}(z) = \exp(2zC_2(\alpha)), \qquad C_2(\alpha) = \frac{e^{-2\alpha}\alpha}{(2\alpha+1)(\alpha+1)}$$

in agreement with Eq. (12). Although we do not have a global argument this proof may clearly be extended to recursively prove self-averageness of any C_k by choosing the parameters $\zeta_j = \omega^j$ with $\omega^k = z/t$.

APPENDIX C

Here we present the proof of Eq. (43). Let us first recall Lagrange formula. Consider a variable X defined by the implicit relation $f(X) = y$ for some given analytic function f. The solution of this equation is supposed to be unique so that X is function of y. Given another analytic function $g(w)$ we look for the Taylor series in y of $g(X(y))$. This composed function may be presented as a contour integral:

$$g(X(y)) = \oint \frac{dz}{2i\pi} \frac{f'(z)}{f(z)-y}\, g(z)$$

Expanding the integrated rational function in Taylor series in y gives Lagrange formula:

$$g(X(y)) = \sum_{n \geqslant 0} y^n \oint \frac{dz}{2i\pi} \frac{f'(z)}{f(z)^{n+1}} g(z)$$

Let $H(x) \equiv 4(G_c(x) + 2x)$. Eq. (42) translates into $H(x) \log H(x) = -8x$ or $W(y)\, e^{W(y)} = y$ for $y = -8x$ and $H(x) = e^{W(y)}$. We now apply Lagrange formula with $f(X) = Xe^X$ and $g(w) = e^w$. This gives

$$H(y) = \sum_{n \geqslant 0} y^n \oint \frac{dz}{2i\pi} \frac{z+1}{z^{n+1}} e^{(1-n)z}$$

$$= 1 + \sum_{n \geqslant 1} \frac{(1-n)^{n-1}}{n!} y^n$$

With $H(x) = 4(G_c(x) + 2x)$ and $y = -8x$, this proves Eq. (43).

AKNOWLEDGMENTS

We thank Gilles Schaeffer for his clarifying help in tree generating functions and Sergei Dorogovtsev for his comments and interest.

REFERENCES

1. R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Modern Phys.* **74**:47 (2002).
2. F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial Species and Tree-like Structures*, Chap. 3, Encyclopedia of Mathematics (Cambridge University Press, 1998).
3. S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, to appear in *Adv. Phys.* **51** (2002).
4. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Anomalous percolation properties of growing networks, *Phys. Rev. E* **64**:066110 (2001).
5. P. Erdös and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**:17–61 (1960).
6. I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Chap. 3 (Wiley, 1983).
7. N. Guelzim, S. Bottani, P. Bourgine, and F. Képès, Topological and causal structure of the yeast genetic network, *Nature Genetics*, in press.
8. J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Evolving protein interaction networks, arXiv:cond-mat/0203167.
9. D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and H. Strogatz, Are randomly grow graphs really random?, *Phys. Rev. E* **64**:041902 (2001).
10. G. Schaeffer, Private Communication.
11. R. M. Wilson and J. H. van Lint, *A Course in Combinatorics*, Chap. 2 (Cambridge University Press, New York, 1992).