Asymmetric evolving random networks

S. Coulomb[a] and M. Bauer[b]

Service de Physique Théorique de Saclay, Laboratoire de la Direction des Sciences de la Matière du Commisariat
à l'Énergie Atomique, URA2306 du CNRS, CEA Saclay, 91191 Gif-sur-Yvette, France

Abstract. We generalize the Poissonian evolving random graph model of M. Bauer and D. Bernard (2003), to deal with arbitrary degree distributions. The motivation comes from biological networks, which are well-known to exhibit non Poissonian degree distributions. A node is added at each time step and is connected to the rest of the graph by oriented edges emerging from older nodes. This leads to a statistical asymmetry between incoming and outgoing edges. The law for the number of new edges at each time step is fixed but arbitrary. Thermodynamical behavior is expected when this law has a large time limit. Although (by construction) the incoming degree distributions depend on this law, this is not the case for most qualitative features concerning the size distribution of connected components, as long as the law has a finite variance. As the variance grows above $1/4$, the average being $< 1/2$, a giant component emerges, which connects a finite fraction of the vertices. Below this threshold, the distribution of component sizes decreases algebraically with a continuously varying exponent. The transition is of infinite order, in sharp contrast with the case of static graphs. The local-in-time profiles for the components of finite size allow to give a refined description of the system.

PACS. 02.10.Ox Combinatorics; graph theory – 02.50.-r Probability theory, stochastic processes, and statistics – 64.60.Ak Renormalization-group, fractal, and percolation studies of phase transitions

1 Introduction

Evolving graphs arise naturally in the modelization of communication networks, but also of social organizations and biological phenomena: brain formation, genetic regulations, etc. Until recently, quantitative data were scarce, but the situation is changing very quickly [1,8,10]. In many important cases the laws governing the evolution of the network are unknown but non deterministic, and the final number of nodes is rather large. This explains why physicists (see the reviews [1,5] and Refs. therein) have been developing recently random graph techniques in the thermodynamic limit to understand some peculiar features, the most salient being large degree distributions, that seem to occur in an unexpected variety of situations.

Many cases of interest lead to oriented graph models, for which the in- and out-degree distributions can be governed by different laws. This was one of the basic observations in the study of the yeast genetic regulatory network presented in [8]: a single gene may participate to the regulation of many other genes – the law for out-degrees seems to be large –, but each gene is only regulated by a few other genes – the law for in-degrees seems to have finite moments. A biological interpretation for the asymme-

try 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. One of our aims is to work with a model that incorporates such an asymmetry.

We shall follow quite closely the philosophy of [2], and in particular pay attention not only to global quantities, but also to local-in-time profiles. We feel that for evolving networks this is a crucial condition to extract relevant information, because as we shall see, global quantities (averaged over time) give a distorted view of the network. On the other hand, it is hard experimentally to access local quantities, either because the ages are not known, or because their consideration would reduce the statistics down to an unacceptable level. But the amount of available information is growing very rapidly, and one can hope that local quantities will become accessible in a near future.

The model we study is the natural evolving cousin of the static maximal entropy model with given in-degree distribution [3,13]. 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. An arbitrary probability generating function T encodes the parameters of the model in the thermodynamic (large t) limit. Precise definitions are given in the next section.

[a] e-mail: `coulomb@spht.saclay.cea.fr`
[b] e-mail: `bauer@spht.saclay.cea.fr`

Our main results are the following:

The global and local in-degree distributions are given by T, see equation (3), whereas the global out-degree distribution is geometric with average $T'(1)$, see equation (5). The local out-degree distribution is Poissonian but age dependent, equation (4). In the case of a static maximal entropy random graph with the same in-degree distribution ([3,13]), the distribution of out-degrees would be Poissonian.

The global structure of connected components is studied *via* a generating function which satisfies a differential equation, leading to recursion relations for the distribution of component sizes, see equation (7). This is illustrated by analytical computations in the case when $T(z)$ is a polynomial of degree 1.

The general criterion for the absence of a component containing a finite fraction of the sites in the thermodynamic limit is that the average α be $\leq 1/2$ and the variance γ be $\leq 1/4$. In that case the system contains components whose sizes scale like a power of the total size of the graph, see equation (14); equivalently, the probability distribution for component sizes has an algebraic queue, see equation (13). Above the threshold (when $\alpha > 1/2$ or $\gamma > 1/4$) this probability distribution is defective but decreases exponentially see equation (16). The boundary separating the percolating and non percolating phases is $\gamma = 1/4$ and $\alpha < 1/2$. In the percolating phase, but close to this boundary, $\gamma - 1/4$ is > 0 but small and the giant component is exponentially small, see equation (15). This situation, somehow reminiscent of the Kosterlitz-Thouless transition, had already been observed in a variety of models [2,4,6].

We compare the percolation criterion with the one that emerges from the study of a static maximal entropy random graph with the same in- and out-degree distributions as our evolving graph ([3,13]). In the static case, the growth of the giant component is generically linear close to the threshold. But we show that if $T(z)$ leads to a percolating static graph, it is automatically percolating for the evolving graph model. The intuitive explanation lies in the inhomogeneities of the evolving graph: the environment of an old vertex is denser than what a static model produces. And indeed, the giant component profile close to threshold is very asymmetric, see equations (23, 24).

In the thermodynamic limit, the finite components are trees, and we derive a direct enumeration formula to count for their abundance equation (11). This can be used to describe all local in time profiles of finite components, a result we also recover for one time quantities *via* a generating function approach equation (8). In Appendix A, we give a proof of the equivalence of the generating function approach and tree enumeration, a question that was left opened in [2].

We have confronted our analytical results with numerical simulations whenever possible, our preferred example being when T is a geometric distribution.

2 The model

The random graphs we consider are constructed according to the following rules:

1. for $t = 1, 2, \cdots$, we denote by $\mathcal{G}_t$ the set of simple graphs with vertex set $V_t = [1, \cdots, t]$, *i.e.* the set of pairs (V_t, E_t) where E_t, the set of edges of the graph, is a subset of $\{(i,j), 1 \leq i < j \leq t\}$. We orient the graph (V_t, E_t) by saying that $(i,j) \in E_t$ is an edge from i to j: an edge always goes from an older to a younger vertex. Note that $\mathcal{G}_1 = \{(\{1\}, \emptyset)\}$ contains a single graph, made of one vertex but no edge.
2. An evolving graph is a sequence $G = (G_t)_{t \geq 1} = ((V_t, E_t))_{t \geq 1}$, where $G_t \in \mathcal{G}_t$ for $t \geq 1$, and for $t \geq 2$, G_{t-1} is the induced subgraph of G_t obtained by removing vertex t and all edges adjacent to it. Stated differently, G_t for $t \geq 2$ is obtained from G_{t-1} by adding vertex t and some edges of the form (i,t), $i < t$. This model of evolution implies that knowing G_t is equivalent to knowing $G_1, \cdots, G_{t-1}, G_t$. We write $\mathcal{G}$ for the set of evolving graphs.
3. In addition to these rules of construction, we put a probability measure p on the set of evolving graphs. If G is an evolving graph, we denote by $\hat{G}_t$ the number of edges arriving at vertex t in G_t (which is the same as the number of edges arriving at t in $G_{t'}$ for any $t' \geq t$). We want to fix the probability distributions for the in-degrees, *i.e.* for the number of edges arriving at vertices $1, 2, \cdots$. So we impose $p(\{G \in \mathcal{G}, \hat{G}_t = k\}) = \tau_{t,k}$ for a given sequence of probability distributions $\{\tau_{t,k}, t \geq 1, 0 \leq k \leq t-1\}, \tau_{t,k} \geq 0, \sum_k \tau_{t,k} = 1$. Since this criterion is far from fixing unambiguously the probability law for the evolving graph, we add the assumption of independence: the number of edges arriving at vertex $1, 2, \cdots, t$ are independent random variables.

The independence assumption deserves a comment. This is clearly the simplest choice to make. It has another simple but interesting property. Suppose that for each $t \geq 1$, π_t is a probability law on $\mathcal{G}_t$ such that

1. the family $\{\pi_t\}_{t \geq 1}$ is compatible: π_t averaged over the edges arriving at vertex t is π_{t-1}.
2. The probability that k edges arrive at vertex t under π_t is $\tau_{t,k}$.

For fixed t, define the entropy of π_t as usual: $S_t \equiv \sum_{G \in \mathcal{G}_t} \pi_t(G) \log \pi_t(G)$. Then a trivial computation shows that for each t, S_t is maximum if under π_t the number of edges arriving at vertices $1, 2, \cdots, t$ are independent random variables. In that sense, the probability law we have chosen has maximum entropy under the constraint that the distribution of in-degrees is fixed.

This means that p is the law that does not introduce any further bias if one is to model a (real) network for which the only observed features are the in-degrees. For instance, possible departures between the (real) network and our model when other quantities are measured would be *the* test to know whether or not the evolution of the

graph is governed solely or mostly by in-degree distributions. This could be generalized to more complicated situations.

Note that we did not worry about the size of the sequence defining an evolving graph. In fact, we shall as often as possible suppose it is infinite.

3 Notations and basic results

3.1 Generating functions

For any vertex $v \geq 1$, let T_v be the generating function for the incoming degree distribution:

$$T_v(z) \equiv \sum_{k=0}^{v-1} \tau_{v,k} z^k.$$

We shall be mostly interested in the thermodynamic limit, *i.e.* in t independent features of the large t behavior of G_t when G is a random element in $(\mathcal{G}, p)$. A typical example is the behavior of $\lim_{t \to \infty} |E_t|/|V_t|$[1]. Obviously, this limit does not exist for any *a priori* choice of the $\tau_{v,k}$'s. The study of thermodynamic convergence for this quantity and others would be of independent interest, but we shall see that a simple assumption ensures that many quantities of interest have a thermodynamic limit: we impose that the sequence of functions $T_v(z)$ converges to a probability generating function

$$T(z) \equiv \sum_{k \geq 0} \frac{\tau_k}{k!} z^k \qquad (1)$$

in such a way that the sequence of averages or of variances converges to that of $T(z)$.

This property implies in particular that, for any k, $\tau_{v,k}$ tends to $\frac{\tau_k}{k!}$ for large v. Moreover, we will often use the fact that $T(1) = 1$. We use the notations $\alpha \equiv T'(1)$ and $\gamma \equiv T''(1) + T'(1) - T'(1)^2$ for the average and variance of T.

As a technical hypothesis, to avoid several pathologies, we shall always implicitly assume that $\tau_0 = T(0) > 0$.

3.2 A preliminary formula

We shall sometimes be lead to consider situations in which a new vertex $t+1$ appears and connects to the rest of the graph avoiding m forbidden vertices. This will happen, for instance, if one is interested in the distribution of connected components in the graph. In this section we would like to give a general formula for such situations. Hence, let $\alpha_{t+1,m}$ be the probability for a new vertex $t+1$ not to connect to m given vertices. This probability is given by the formula:

$$\alpha_{t+1,m} \equiv \sum_{n \geq 0} \tau_{t+1,n} \frac{\binom{t-m}{n}}{\binom{t}{n}}.$$

Defining $\alpha_{t+1,m}(z) \equiv \sum_{n \geq 0} \tau_{t+1,n} z^n \frac{\binom{t-m}{n}}{\binom{t}{n}}$, we notice that

$$\alpha_{t+1,m+1}(z) = \alpha_{t+1,m}(z) - \frac{z}{t-m} \alpha'_{t+1,m}(z).$$

Since $\alpha_{t+1,0}(z) = T_{t+1}(z)$, we have:

$$\alpha_{t+1,m}(z) = T_{t+1}(z) - \frac{\binom{m}{1} z}{t} T'_{t+1}(z)$$
$$+ \frac{\binom{m}{2} z^2}{t(t-1)} T''_{t+1}(z)$$
$$- \cdots (-)^m \frac{\binom{m}{m} z^m}{t \cdots (t-m+1)} T^{(m)}_{t+1}(1)$$
$$\alpha_{t+1,m} = 1 - \frac{\binom{m}{1}}{t} T'_{t+1}(1) + \frac{\binom{m}{2}}{t(t-1)} T''_{t+1}(1)$$
$$- \cdots (-)^m \frac{\binom{m}{m}}{t(t-1) \cdots (t-m+1)} T^{(m)}_{t+1}(1).$$

In the thermodynamic limit, we shall only need the large t finite m approximation

$$\alpha_{t+1,m} \simeq 1 - \frac{\alpha m}{t}. \qquad (2)$$

These simple formulæ may now be used to calculate a few quantities. We'll first describe the degree distributions of incoming and outgoing edges, whereas the two following sections will present the main relations governing the distributions of connected components of the graphs.

4 Degree distribution

4.1 Notations

Let

1. $l_j^-(t)$ (resp. $l_j^+(t)$) be the number of incoming (resp. outgoing) edges at a vertex j at time t
2. $v_k^-(t)$ (resp. $v_k^+(t)$) be the number of vertices with k incoming (resp. outgoing) edges at time t.

With these notations, the edge distributions are described by the generating functions:

$$\nu_t^{\pm}(z) \equiv \frac{1}{t} \sum_{0 \leq k < t} \left\langle v_k^{\pm}(t) \right\rangle z^k = \frac{1}{t} \sum_{1 \leq j \leq t} \left\langle z^{l_j^{\pm}(t)} \right\rangle.$$

In the thermodynamic limit, it is possible to give an expression for these generating functions.

4.2 In-degree distribution

Up to a normalization factor, the probability for vertex j to have k incoming edges is $\tau_{j,k}$. This shows that

$$\left\langle z^{l_j^-(t)} \right\rangle = \sum_{k \geq 0} \tau_{j,k} z^k.$$

In the thermodynamic limit, this becomes

$$\left\langle z^{l_j^-(t)} \right\rangle = T(z). \tag{3}$$

Hence, in the thermodynamic limit, the probability for any vertex to have k incoming edges is given by $\frac{\tau_k}{k!}$, which is correctly normalized.

This implies in particular that, in the thermodynamic limit, the average number of edges coming to a vertex is precisely α.

4.3 Out-degree distribution

Any edge emerging from a vertex j can be seen as an edge arriving at a vertex younger than j. Let $t > j$ be the time at which one observes the number of edges emerging from j. We define $\sigma \equiv j/t$. The quantity $\left\langle z^{l_{\sigma t}(t)} \right\rangle$ is given by the formula:

$$\sum_k z^k \sum_{\sigma t < j_1 < \cdots < j_k \leq t} \prod_{i=1}^k \left(\sum_{k'=0}^{j_i-2} \tau_{j_i,k'+1} \frac{\binom{j_i-2}{k'}}{\binom{j_i-1}{k'+1}} \right)$$
$$\times \prod_{j < j' \leq t; j' \neq j_i} \left(\sum_{k'=0}^{j'-2} \tau_{j',k'} \frac{\binom{j'-2}{k'}}{\binom{j'-1}{k'}} \right).$$

Indeed, the probability for j to have k outgoing edges is obtained by summing over $j_1, \cdots, j_k > j$ the probabilities for j to be linked to these k vertices and to no other vertex. Each of this probabilities can easily be calculated from the following relations:

$$\sum_{0 \leq k' \leq j_i-2} \frac{\tau_{j_i,k'+1}}{\binom{j_i-1}{k'+1}} \binom{j_i-2}{k'} = \frac{T'_{j_i}(1)}{j_i-1}$$
$$\sum_{0 \leq k' \leq j'-2} \frac{\tau_{j',k'}}{\binom{j'-1}{k'}} \binom{j'-2}{k'} = 1 - \frac{T'_{j'}(1)}{j'-1}.$$

For fixed σ and large t, the second product simplifies to σ^α and the average we look for becomes:

$$\left\langle z^{l_{\sigma t}(t)} \right\rangle \simeq \sigma^\alpha \sum_k z^k \sum_{\sigma t < j_1 < \cdots < j_k \leq t} \frac{\alpha^k}{j_1 \cdots j_k}$$
$$\simeq \sigma^\alpha \sum_k z^k \frac{1}{k!} [\ln t - \ln \sigma t]^k \alpha^k.$$

Hence, for fixed σ,
$$\left\langle z^{l_{\sigma t}(t)} \right\rangle \to e^{-(z-1)\alpha \ln \sigma} \tag{4}$$

for large t, so the local out-degree distribution is Poissonian with σ-dependent parameter $\alpha \ln \sigma$. Integrating σ between 0 and 1 yields the asymptotic value for the generating function $\nu_t^+(z)$:

$$\nu_t^+(z) \to \frac{1}{1 + \alpha(1 - z)} \tag{5}$$

at large t.

Identifying the term of degree k in z in the development of this function finally yields the probability for a vertex to have k outgoing edges:

$$p^+(k) = \frac{\alpha^k}{(1 + \alpha)^{k+1}}.$$

We see that the distribution of outgoing edges is geometric and depends on the probability distribution T only through the average number of incoming edges α.

4.4 Mixed distribution

In our model, the number of in and out edges $l_j^+(t)$ and $l_j^-(t)$ at a given vertex are independent by construction. The generating function for the mixed degree distribution is

$$\nu_t(z_+, z_-) \equiv \frac{1}{t} \sum_{k_+, k_-} \left\langle v_{k_+, k_-}(t) \right\rangle z_+^{k_+} z_-^{k_-}$$
$$= \frac{1}{t} \sum_{1 \leq j \leq t} \left\langle z_+^{l_j^+(t)} z_-^{l_j^-(t)} \right\rangle$$

which is easily obtained from equations (3, 5):

$$\nu(z_+, z_-) = \frac{T(z_-)}{1 + \alpha(1 - z_+)}. \tag{6}$$

The local counterpart would easily follow from equations (3, 4).

5 Connected components

In this section we give formulæ for the number of components of size k in the thermodynamic limit (Sect. 5.1), and for the time distribution of these components (Sect. 5.3). This derivation makes some natural assumptions of self averaging. In Section 5.4, we evaluate without such assumptions the contribution of individual graphs to the weights of random graphs, and show that trees dominate the thermodynamic limit. We use this result in the appendix to compute directly the generating function for the number of components of size k, and show that it coincides with the one given in Section 5.1.

5.1 Global-time results

Let $N_k(t)$ be the number of connected components of size k at time t for a given graph, and $N_t(z)$ the corresponding generating function:

$$N_t(z) \equiv \sum_{k \geq 1} N_k(t) z^k.$$

Instead of making an argument at finite t and taking the thermodynamic limit $t \to \infty$ afterwards, we shall assume

that the thermodynamic limit exists. The computations in Appendix A, which yield the same results, contain a fully rigorous argument by using a totally different approach.

So we assume that for large t and fixed k, $N_k(t) \sim tC_k$. In addition, there might be a giant component which occupies about tP_∞ sites. We analyze what happens when the graph grows from size t to size $t + \Delta t$ where $t \gg \Delta t \gg 1$ are such that Δt is already a thermodynamical number but t is so large that the statistical properties of the graph do not change appreciably between t and $t + \Delta t$, and make a quick estimate of the averages.

Neglecting terms which are $o(\Delta t)$ (coming either from fluctuations or from other origins), we first observe that there are about $\Delta t \tau_m/m!$ vertices between t and $t + \Delta t$ which have in-degree m. On the other hand, between t and $t + \Delta t$, $\Delta t \alpha$ edges have been created, they have connected at random young vertices to older connected components, the probability that a new vertex connects more than once to a finite connected component being negligible. So $\Delta t \alpha k C_k$ components that were of size k at time t are now parts of larger components. So the number of components of size k destroyed between t and $t + \Delta t$ is

$$\sim \Delta t \, \alpha k C_k.$$

On the other hand, about $\Delta t \tau_m/m!$ vertices between t and $t + \Delta t$ have m incoming edges. Suppose that such a vertex got connected to components of size $k_1, \cdots, k_m$ to build a component of size $1 + k_1 + \cdots + k_m$. The probability of this event is $k_1 C_{k_1} \cdots k_m C_{k_m}$ so that the total number of components of size k that have been created by such a process is

$$\sim \Delta t \sum_m \frac{\tau_m}{m!} \sum_{k_1, \cdots, k_m} k_1 C_{k_1} \cdots k_m C_{k_m} \delta_{k, 1+k_1+\cdots+k_m}.$$

On the other hand, we know the total variation of the number of components of size k between t and $t + \Delta t$ is about $(t + \Delta t)C_k - tC_k = \Delta t C_k$. To resume, the net balance is

$$C_k = -\alpha k C_k + \sum_m \frac{\tau_m}{m!}$$
$$\times \sum_{k_1, \cdots, k_m} k_1 C_{k_1} \cdots k_m C_{k_m} \delta_{k, 1+k_1+\cdots+k_m}.$$

This is cleanly expressed as a functional equation for the generating function $C(z) \equiv \sum_k C_k z^k$:

$$C(z) = -z\alpha \partial_z C(z) + zT(z\partial_z C(z)) \qquad (7)$$

which describes the number of components of size k at large t.

5.2 A simple example

The case when $T(z) = (1-p) + pz$ leads to tractable equations even at finite t. As we shall see later, this example is pathological from the point of view of the percolation

transition. This is another reason to give a separate treatment.

Though it is not unlikely that this toy model has been solved more than once before, we have found no reference. So we give a sketch of the solution with apologies to the original contributions if any. In particular, we compute the scaling function governing the statistics of large components.

When vertex $t + 1$ is added, it remains isolated with probability $1 - p$ in which case $N_k(t+1) - N_k(t) = \delta_{k,1}$, or is attached to a component of size l to build a component of size $l+1$ with probability $plN_l(t)/t$, in which case $N_k(t+1) - N_k(t) = \delta_{k,l+1} - \delta_{k,l}$. We infer

$$\langle N_{t+1}(z)\rangle - \langle N_t(z)\rangle = (1-p)z$$
$$+ \sum_{k,l} pl \langle N_l(t)\rangle /t(\delta_{k,l+1} - \delta_{k,l})z^k$$
$$= (1-p)z + p\sum_l l \langle N_l(t)\rangle/t(z^{l+1} - z^l)$$
$$= (1-p)z + p/t(z^2 - z)\partial_z \langle N_t(z)\rangle.$$

The initial condition is $\langle N_1(z)\rangle = N_1(z) = z$ We can simplify this equation by the change of variable $z = w/(1+w)$. Setting $Q_t(w) \equiv \langle N_t(z)\rangle$, the equation for Q_t is

$$Q_{t+1} - Q_t = (1-p)\frac{w}{1+w} - p\frac{w}{t}\partial_w Q_t.$$

$N_t(z)$ is a polynomial in z, so Q_{t+1} has a regular series expansion in w, $Q_t \equiv \sum_{k\geq 1} q_k(t)w^k$ which leads to $q_k(t+1) = \frac{t-pk}{t}q_k(t) + (1-p)(-1)^{k+1}$. Direct substitution shows that $q_k(t) = t\frac{1-p}{1+pk}(-1)^{k+1}$ is a particular solution. The general solution $r_k(t)$ of the associated homogeneous equation $r_k(t+1) = \frac{t-pk}{t}r_k(t)$ is $r_k(t) = \frac{\Gamma(t-pk)}{\Gamma(t)\Gamma(1-pk)}r_k(1)$. Taking into account the initial condition $Q_1 = w/(1+w)$ leads to

$$q_k(t) = (-1)^{k+1}\left(t\frac{1-p}{1+pk} + \frac{p(k+1)}{1+pk}\frac{\Gamma(t-pk)}{\Gamma(t)\Gamma(1-pk)}\right).$$

We can now go back to the z variable:

$$\langle N_k(t)\rangle = \sum_{l=1}^k (-1)^{l+1}$$
$$\times \left(t\frac{1-p}{1+pl} + \frac{p(l+1)}{1+pl}\frac{\Gamma(t-pl)}{\Gamma(t)\Gamma(1-pl)}\right)\frac{\Gamma(k)}{\Gamma(l)(k-l)!}.$$

The contribution corresponding to the first term inside the parenthesis, the one which is linear in t, can be resummed. Indeed, if $C(z)$ is a solution of $C = (1-p)z + p(z^2 - z)\partial_z C$, $tC(z)$ solves the original equation for $\langle N_t(z)\rangle$ even at finite t. The solution which is regular at $z = 0$ is

$$C(z) = \sum_{k\geq 1} \frac{1-p}{p}\frac{\Gamma(k)\Gamma(1+1/p)}{\Gamma(k+1+1/p)}z^k.$$

One can check that, for any $p \in [0, 1[$, $C'(z = 1) = 1$. As we shall explain later, that means there is no percolation.

To summarize,

$$\langle N_k(t)\rangle = t\frac{1-p}{p}\frac{\Gamma(k)\Gamma(1+1/p)}{\Gamma(k+1+1/p)}$$
$$+ \sum_{l=1}^{k}(-1)^{l+1}\frac{p(l+1)}{1+pl}\frac{\Gamma(t-pl)}{\Gamma(t)\Gamma(1-pl)}\frac{\Gamma(k)}{\Gamma(l)(k-l)!}.$$

It is now possible to get the exact scaling function governing the size distribution of large components. When both t and k are large, the thermodynamic contribution scales like $\frac{1-p}{p}\Gamma(1+1/p)tk^{-1-1/p}$ while in the other contribution, for fixed l, one finds $(-1)^{l+1}\frac{p(l+1)}{1+pl}\frac{1}{\Gamma(1-pl)\Gamma(l)}t^{-pl}k^{l-1}$. This decreases very fast with l so for large k one can extend the range of l to ∞. The balance between the thermodynamic and finite size contributions shows that the scaling variable is $s = kt^{-p}$. In the scaling limit

$$\langle kN_k(t)\rangle \sim \frac{1-p}{p}\Gamma(1+1/p)s^{-1/p}$$
$$+ \sum_{l\geq 1}(-1)^{l+1}\frac{p(l+1)}{1+pl}\frac{1}{\Gamma(1-pl)\Gamma(l)}s^l.$$

Let $S(s)$ denote the scaling function on the right hand side.

Defining $sA(s) \equiv \sum_{l\geq 1}(-1)^{l+1}\frac{1}{\Gamma(1-pl)\Gamma(l)}s^l$, one can rewrite

$$S(s)/s = \frac{1-p}{p}\Gamma(1+1/p)s^{-1-1/p}$$
$$+ A(s) - \frac{1-p}{p}\int_0^1 d\lambda A(\lambda s)\lambda^{1/p}.$$

This expression of S exhibits clearly its small s behavior. To get control on the large s behavior, one can use the familiar representation $\frac{1}{\Gamma(z)} = \frac{1}{2i\pi}\int_{\mathcal{C}}dw e^w w^{-z}$ for $z = 1 - pl$ and sum over l to obtain

$$A(s) = \frac{1}{2i\pi}\int_{\mathcal{C}}dw e^{w-sw^p}w^{p-1}.$$

Observing that $\Gamma(1+1/p)s^{-1-1/p} = \int_0^\infty d\lambda A(\lambda s)\lambda^{1/p}$, one gets a compact expression

$$S(s)/s = A(s) + \frac{1-p}{p}\int_1^\infty d\lambda A(\lambda s)\lambda^{1/p}.$$

The large s expansion can then be obtained by standard methods (saddle point for $A(s)$ and then analysis at $\lambda = 1$ for the second term). The formulæ are rather cumbersome and we simply quote the leading exponential behavior

$$\log S(s) \sim -\frac{1-p}{p}p^{\frac{1}{1-p}}s^{\frac{1}{1-p}} \qquad s \to +\infty,$$

showing that $S(s)$ decreases very fast at large s. As a
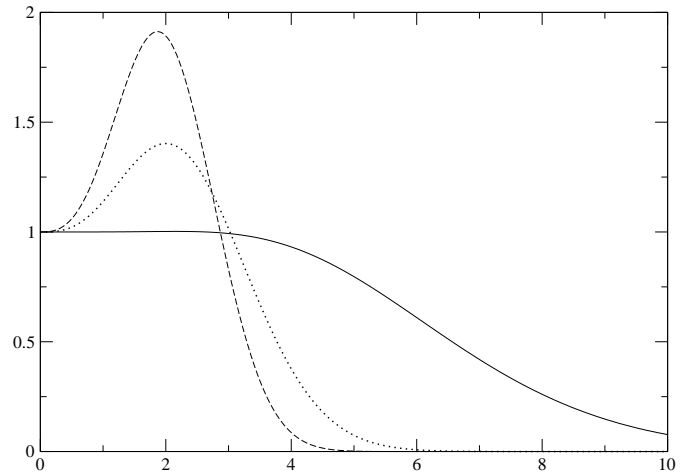
Fig. 1. Plots of $S(s)/S_{thermo}(s)$ for $p = 0.2$ (solid line), $p = 0.5$ (dots) and $p = 0.6$ (dashed line).

simple example, take $p = 1/2$. Then $A(s) = \frac{1}{\sqrt{\pi}}e^{-s^2/4}$ and

$$S(s) = \frac{1}{\sqrt{\pi}}\left(se^{-s^2/4} + \frac{1}{s^2}\int_s^\infty d\lambda\lambda^2 e^{-\lambda^2/4}\right).$$

Returning to arbitrary p, the relation $\langle kN_k(t)\rangle \sim S(k/t^p)$ implies that the average number of components of size $k \geq st^p$ is $\left\langle\sum_{k\geq st^p}N_k(t)\right\rangle \sim \int_s^\infty du\, S(u)/u < \infty$. So the large components have a size of order t^p.

5.3 Local-in-time results

Knowing the global-in-time distribution of connected components, it is natural to wonder whether this distribution is homogeneous in time or not. More precisely, if $j = \sigma t$ with fixed σ and $t \to \infty$, what can we say about the probability for vertex j to be in a component of size k?

For $k \geq 1$ let $\rho_k(\sigma)$ be the probability that, in the thermodynamic limit, vertex σt belongs to a connected component of size k. The purpose of this section is to give an equation governing the ρ_k's.

Local-in-time equation

We use directly the thermodynamic assumption. A rigorous justification could be obtained by a straightforward refinement of the argument in Appendix A.

As in the global case, let Δt be an interval of time such that $\Delta t \gg 1$ but $\frac{\Delta t}{t} \ll 1$. Between times t and $t + \Delta t$, the number of vertices added to the graph is large but much smaller than the size of the graph.

During the proof of equation (7), we have established that the number of new components of size k is

$$\Delta t \sum_m \frac{\tau_m}{m!}\sum_{k_1,\cdots,k_m}k_1 C_{k_1}\cdots k_m C_{k_m}\ \delta_{1+\sum_i k_i,k}.$$

Each of the m old components has average time distribution $\frac{\rho_{k_i}(\sigma)}{C_{k_i}}$ so that, on average, the contribution of the Δt new components to the time distribution of components on k vertices is

$$\Delta t \sum_m \frac{\tau_m}{m!} \sum_{k_1,\cdots,k_m} k_1 C_{k_1} \cdots k_m C_{k_m}$$

$$\times \left(\frac{\rho_{k_1}(\sigma)d\sigma}{C_{k_1}} + \cdots + \frac{\rho_{k_m}(\sigma)d\sigma}{C_{k_m}} \right) \delta_{1+\sum_i k_i, k}.$$

Let us put $\rho(\sigma, z) \equiv \sum_k \rho_k(\sigma) z^k$. Multiplying the expression above by z^k and summing over k yields the term $\Delta t\, z^2 \partial_z \rho(\sigma, z) T'(z\partial_z C) d\sigma$.

During the proof of equation (7), we have already observed that $\alpha \Delta t\, k C_k$ components of size k are destructed between t and $t+\Delta t$. According to the definition of $\rho_k(\sigma)$, the number of vertices of relative age contained between σ and $\sigma + d\sigma$, which belonged to one of these components is $\alpha \Delta t\, k \rho_k(\sigma)$.

After summation over k, the destruction term is $\alpha \Delta t\, z \partial_z \rho(\sigma, z) d\sigma$.

Moreover, a vertex of relative age σ at time t has relative age $\frac{\sigma t}{t+\Delta t}$ at time $t + \Delta t$. Hence, the local-in-time profile verifies, to first order in $\Delta t/t$, the following relation
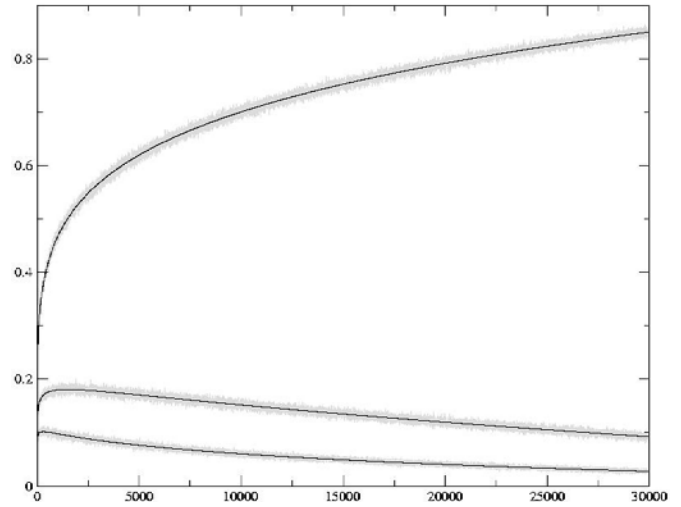
$$t d\sigma \left(\rho \left(\frac{\sigma t}{t + \Delta t}, z \right) - \rho(\sigma, z) \right) = -\alpha \Delta t\, z \partial_z \rho(\sigma, z) d\sigma$$

$$+ \Delta t\, z^2 \partial_z \rho(\sigma, z) T'(z\partial_z C) d\sigma.$$

This leads to the differential equation

$$\sigma \partial_\sigma \rho = \left(\alpha - z T'(z\partial_z C) \right) z \partial_z \rho. \tag{8}$$

As in the case of a Poisson law, this equation leads to recursion relations for the $\rho_k(\sigma)$'s. Differentiating equation (8) k times with respect to z and taking $z = 0$ leads to a first order linear differential equation for ρ_k, in which $\rho_1, \cdots, \rho_{k-1}$ appear. Putting $x = \sigma^\alpha$, the first few distributions are:

$$\rho_1 = \tau_0 x$$

$$\rho_2 = \frac{\tau_0 \tau_1}{\alpha} \left[x - \frac{x^2}{1+\alpha} \right]$$

$$\rho_3 = x \left[\frac{\tau_0^2 \tau_2}{2\alpha(1+\alpha)} + \frac{\tau_1^2 \tau_0}{\alpha^2} \right]$$

$$- x^2 \frac{2\tau_0 \tau_1^2}{\alpha^2(1+\alpha)}$$

$$+ x^3 \left[\frac{\tau_0 \tau_1^2}{\alpha^2(1+2\alpha)} - \frac{\tau_0^2 \tau_2}{2\alpha(1+\alpha)^2} \right].$$

Fig. 2. The analytic result (solid lines) for the profiles of small connected components (from top to bottom $k = 1, 2, 3$) compared to numerical simulations (gray clouds) on 5000 random graphs of size 30000. $T(z) = (1-p)/(1-pz)$ with $p = 0.2$.

Component distribution at $\sigma = 1$

To conclude this paragraph, we shall compute the probability for the youngest vertex to belong to a component of size k. This is not difficult to do because, unlike all the other vertices in the graph, the youngest one does not have any outgoing edge. Hence, as established in the proof of equation (7), it belongs to a component of size k with probability $\sum_m \frac{\tau_m}{m!} \sum_{k_1,\cdots,k_m} k_1 C_{k_1} \cdots k_m C_{k_m}\ \delta_{1+\sum_i k_i, k}$. More compactly:

$$\rho(\sigma = 1, z) = z T(z\partial_z C). \tag{9}$$

Note that equations (8, 9) can be used to recover the global equation (7).

5.4 Tree distributions

Given a connected graph G, one may wonder how many connected components of the random evolving graph are isomorphic to G. In other words, if k is the number of vertices of G, we look for the average number of increasing maps $v : [1, \cdots, k] \to [1, \cdots, t]$ such that the vertices $v_1, \cdots, v_k$ span a connected component of the random graph isomorphic to G.

Let m_i be the number of edges incoming to vertex i in G. The probability that vertices $v_1, \cdots, v_k$ span a connected component of the random graph isomorphic to G is evaluated using the following two rules:

- vertex v_i has m_i incoming edges coming from the given vertices ;
- the vertices w of the random graph which are not in the image of v must not be connected to any of the v_i's.

Hence, putting $v_{k+1} = t+1$, the probability we look for is

$$\prod_{i=1}^{k} \left[\frac{\tau_{v_i,m_i}}{\binom{v_i-1}{m_i}} \prod_{v_i < w_i < v_{i+1}} \sum_{j \geq 0} \frac{\tau_{w_i,j}}{\binom{w_i-1}{j}} \binom{w_i-i-1}{j} \right].$$

The average number of components is then obtained by summing this expression over all increasing maps v

$$\langle n_G \rangle = \sum_{1 \leq v_1 < \cdots < v_k \leq t} \prod_{i=1}^{k}$$
$$\times \left[\frac{\tau_{v_i,m_i}}{\binom{v_i-1}{m_i}} \prod_{v_i < w_i < v_{i+1}} \sum_{j \geq 0} \frac{\tau_{w_i,j}}{\binom{w_i-1}{j}} \binom{w_i-i-1}{j} \right].$$

In the thermodynamic limit, the sum over j is given by formula (2), and its product over w_i tends to $e^{-i\alpha \ln \frac{v_{i+1}}{v_i}}$ provided only large w_i's contribute significantly. Hence, approximating the sum over $v_1, \cdots, v_k$ by an integral yields a contribution of G equal to

$$C_{m_1,\cdots,m_k} = t^{k-m} \left(\prod_{i=1}^{k} \tau_{m_i} \right)$$
$$\times \int_{0 \leq \sigma_1 \leq \cdots \leq \sigma_k \leq 1} d\sigma_1 \sigma_1^{\alpha-m_1} \cdots d\sigma_k \sigma_k^{\alpha-m_k}. \quad (10)$$

This formula looks pretty much like the one proposed in the poissonnian case in [2]. In particular, it shows exactly in the same way that only connected graphs with $k = m+1$ (*i.e. trees* by Euler's formula) give a contribution that scales like t in the thermodynamic limit. Moreover, the contribution of a tree with degree distribution m_i is

$$\prod_{i=1}^{k} \frac{\tau_{m_i}}{i(\alpha+1) - (m_1 + \cdots + m_i)}. \quad (11)$$

This formula was obtained by integration over all relative ages $\sigma_0, \cdots, \sigma_{k-1}$. However, if we only integrate over some of these variables while fixing the others, say $\sigma_{k_1} < \sigma_{k_2} < \cdots < \sigma_{k_p}$, we expect to get the contribution of a given tree with k vertices amongst which p vertices have imposed age. For instance, taking $k = 2$ and integrating over σ_1 while imposing $\sigma_2 = \sigma$ gives the contribution of trees of size 2 with younger vertex of age σ. Explicit integration yields a contribution $\tau_0 \tau_1 \sigma^{2\alpha}/(\alpha+1)$. On the other hand, if we integrate over σ_2 and fix $\sigma_1 = \sigma$, we get $\tau_0 \tau_1 (\sigma^\alpha/\alpha - \sigma^{2\alpha}/\alpha)$ for the contribution of trees of size 2 with older vertex of age σ. The sum of these two quantities, which is expected to be the probability that a site of relative age σ belongs to a tree with two vertices is, indeed, equal to $\rho_2(\sigma)$ as calculated from equation (8).

In fact, this result is not really surprising, since the derivation of equation (7) relies on the fact that, in the thermodynamic limit, a new vertex connects with vanishing probability to several vertices in the same component. In other words, this equation of evolution takes only trees into account so equation (11) should imply it. The full proof is instructive but tedious, and we relegate it to the appendix.

6 The percolation transition

Formula (7) gives a relation between the C_k's, which represent the asymptotic number of connected components of size k. This equation involves in particular the function $z \partial_z C(z) = \sum_k k C_k z^k$.

Note that $k C_k$ is the fraction of sites belonging to components of size k. This means that, if $\sum_k k C_k = 1$, no single component in the graph can have size $O(t)$. On the other hand, if $\sum_k k C_k < 1$, the possibility exists that a giant component contains a finite fraction of the sites. The possibility that several very large components coexist is usually ruled out because under a rearrangement of a number $o(t)$ of edges these components would merge with finite probability. Though we have not tried to build a formal argument in the case of evolving graphs, the intuition remains the same and is confirmed by numerical simulations. So we take for granted that if $\sum_k k C_k < 1$ a single giant component contains a fraction $1 - \sum_k k C_k$ of the sites.

6.1 Main results

The arguments that lead to the main qualitative and quantitative features of the percolation transition are of technical nature. So we postpone them to the next section.

Percolation criterion

Unless $T(z)$ is a polynomial of degree 1, the system contains a giant component if either the variance γ of T is $> 1/4$ or the average α of T is $> 1/2$. The boundary separating the non percolating phase from the percolating phase is given by the two conditions $\gamma = 1/4$ and $\alpha < 1/2$. The emergence of a giant component is purely due to an increase of the variance of T above the threshold $1/4$.

Behaviour below threshold

If $\gamma \leq 1/4$ and $\alpha \leq 1/2$, the distribution of component sizes in the system is critical, characterized by the following three equivalent properties.

The dominant singularity of $C(z)$ at $z = 1$ is

$$C^{sing}(z) \propto (1-z)^{\frac{2}{1-\sqrt{1-4\gamma}}}. \quad (12)$$

For large k, the fraction of sites belonging to components of size k decreases like

$$k C_k \propto k^{-\frac{2}{1-\sqrt{1-4\gamma}}}. \quad (13)$$

For a finite system of size $t \to \infty$, the large components have a size of order

$$k(t) \propto t^{\frac{1-\sqrt{1-4\gamma}}{2}}. \quad (14)$$

Behaviour above but close to threshold

The percolation transition is of infinite order when γ crosses the value $1/4$ while keeping $\alpha < 1/2$. If we denote by P_∞ the fraction of sites occupied by the giant component, then when $\gamma - 1/4 \to 0^+$, P_∞ is exponentially small:

$$\log P_\infty \sim -\pi/\sqrt{4\gamma - 1} \quad \text{for } \gamma \to 1/4^+. \qquad (15)$$

For large k, the fraction of sites belonging to components of size k decreases like

$$kC_k \propto k^{-\frac{3}{2}} e^{-kP_\infty}. \qquad (16)$$

6.2 Discussion

We start our discussion by analyzing the behavior of $C(z)$ close to $z = 1$. As before, we assume that $T(z)$ is not a polynomial of degree 1, *i.e.* that $T''(1) \neq 0$.

We start from equation (7) and apply the operator $z\partial_z$ to get

$$(z\partial_z C) + \alpha z\partial_z(z\partial_z C) =$$
$$zT(z\partial_z C) + zT'(z\partial_z C)z\partial_z(z\partial_z C) \quad (17)$$

which involves only $z\partial_z C$.

Set $z = e^\tau$ and $Y(\tau) \equiv 1 - z\partial_z C$. Equation (17) can be rewritten for Y:

$$0 = Y(\tau) - 1 + \alpha\dot{Y}(\tau)$$
$$+ e^\tau \left(T(1 - Y(\tau)) - T'(1 - Y(\tau))\dot{Y}(\tau) \right).$$

We now make two assumptions.

i)   There is no giant component: $\sum_k kC_k = 1$ or equivalently $Y(0) = 0$,
ii) the size distribution of clusters has a first moment: $\mu_1 = \sum_k k^2 C_k$ is finite.

Note that ii) implies that $\dot{Y}(0^-)$ exists and has value $-\mu_1$.

We differentiate the equation for Y with respect to τ and put $\tau = 0$, yielding:

$$\mu_1^2 T''(1) + \mu_1(2\alpha - 1) + 1 = 0.$$

The discriminant of this equation for μ_1 is $1 - 4\gamma$ where $\gamma = T''(1) + \alpha - \alpha^2$ is the variance of the distribution T.

The case when $T''(1) = 0$, *i.e.* when $T(z)$ is affine, has some pathologies, but it has already been treated in detail.

If $T''(1) > 0$, the quadratic equation has two roots,

a) both real and positive if $1 - 2\alpha$ and $1 - 4\gamma$ are positive,
b) both real and negative if $1 - 2\alpha$ is negative but $1 - 4\gamma$ is positive,
c) both complex if $1 - 4\gamma$ is negative.

Clearly, only case a) is compatible with our two assumptions i) and ii). In the sequel we shall take for granted that in this case, the assumptions i) and ii) are indeed true.

In cases b) and c), at least one of the assumptions must fail. We show that it is i), the absence of giant component. To do that we need a more precise analysis.

Let us first give some properties of $Y(\tau)$ for $\tau \leq 0$. By construction, $Y(\tau)$ has a convergent expansion in powers of e^τ with negative coefficients (except the first) and is bounded by 0. So Y is continuous decreasing on $]-\infty, 0]$.

To obtain the large order behavior of C_k we simplify equation (17) assuming that τ and $Y(\tau)$ are small[2]. This is certainly a good approximation to describe the small τ behavior of $Y(\tau)$ when there is no percolation cluster because in that case $Y(\tau)$ is continuous and vanishes at $\tau = 0$. It is also true close to a continuous phase transition because $Y(0)$ is small.

Keeping only the dominant contributions yields

$$\dot{Y}(T''(1)Y - \alpha\tau) + (1 - \alpha)Y + \tau \simeq 0.$$

Setting $\gamma F = T''(1)Y - \alpha\tau$, we derive the limit equation

$$\gamma F\dot{F} + F + \tau = 0. \qquad (18)$$

This equation is the same as the one found and studied in [2,6]. As the presentation in [2] is closer in spirit to this one, this is the one we refer to in the sequel for details. In the Poissonian case, the authors showed carefully that equation (18) indeed contains the quantitative universal features of the exact cluster generating function. We take for granted that this remains true for general T.

When $\gamma < 1/4$, we write $\beta = \sqrt{1 - 4\gamma}$. The general integral of equation (18) is

$$\left(1 + \frac{1 - \beta}{2}\frac{F}{\tau}\right)^{\frac{1+\beta}{2\beta}} \left(1 + \frac{1 + \beta}{2}\frac{F}{\tau}\right)^{\frac{\beta-1}{2\beta}} = \frac{\mathrm{C}^{st}}{\tau}. \qquad (19)$$

Suppose $F(0) = 0$, but $F(\tau)/\tau$ is not bounded close to $\tau = 0^-$. Then at a point where $F(\tau)/\tau$ is large, equation (19) implies that $F(\tau)$ is of order one, a contradiction. For analogous reasons, if $F(0) = 0$ and $F(\tau)/\tau$ is bounded, $\lim_{\tau \to 0^-} 1 + \frac{1+\beta}{2}\frac{F}{\tau} = 0$ and then

$$F(\tau) + \frac{2}{1 + \beta}\tau \propto \tau^{\frac{1+\beta}{1-\beta}} \ll \tau. \qquad (20)$$

To summarize, if $\gamma < 1/4$ and $F(0) = 0$, $F'(0)$ exists (and then $\mu_1 = \sum_k k^2 C_k$ is finite) so that i) implies ii): this means that if $\gamma < 1/4$ but $\alpha > 1/2$, i) has to be wrong, and $\sum_k kC_k < 1$.

A word of caution is needed here. To get the limiting equation (18), we have neglected terms of order τ^2 in (17). So strictly speaking, equation (20) is correct only if $\frac{1+\beta}{1-\beta} < 2$, *i.e.* $2/9 < \gamma < 1/4$. A more careful analysis, analogous to that sketched in [2], would show that in general, in the absence of a giant component, the small τ expansion of

---

[2] Note that we do not assume that they are of the same order of magnitude.
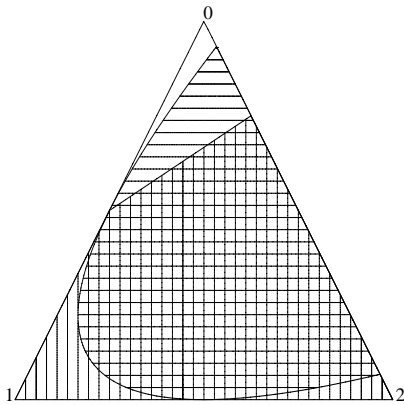
Fig. 3. Phase structure of the evolving graph in the subsimplex $0 = p_3 = p_4 = \cdots,\ p_0 + p_1 + p_2 = 1$. The unshaded area corresponds to the non percolating phase. Horizontal lines fill the region $\gamma \geq 1/4$ and vertical lines the region $\alpha \geq 1/2$.

$F(\tau)$ starts with a standard Taylor series in τ up to order $\lfloor \frac{1+\beta}{1-\beta} \rfloor$ and then a leading singularity proportional to $\tau^{\frac{1+\beta}{1-\beta}}$.

When $\gamma < 1/4$, we write $\beta = \sqrt{4\gamma - 1}$. Equation (18) implies that

$$\frac{1}{2}\log(\gamma F^2 + \tau F + \tau^2) - \frac{1}{\beta}\arctan\left(\frac{\beta F}{F + 2\tau}\right) \qquad (21)$$

is locally constant. As F is continuous, the above quantity jumps by $\pm\pi$ when $F + 2\tau$ changes sign. To fix conventions, we specify the function arctan by demanding that it is continuous and takes value in $]-\pi/2, \pi/2[$. We argue by contradiction that F cannot vanish at $\tau = 0$: if F and τ are small, the argument of the log is a small positive number so the first term gets large and negative, while the term involving arctan remains bounded. So again i) fails.

We have established the percolation criterion announced in the previous section.

Furthermore, the relation $\gamma - 1/4 = T''(1) - (2\alpha - 1)^2/4$ shows that the region separating the percolating phase from the non percolating phase is $\gamma = 1/4$ and $2\alpha - 1 \leq 0$: the emergence of a giant component is purely due to an increase of the variance of T above the threshold $1/4$. This is a bit counterintuitive, because it implies that there are cases with arbitrary small α and a giant component.

To illustrate that point, we present in Figure 3 the case when T is a quadratic polynomial.

The properties that describe the system when there is no giant component or close to the threshold (when the giant component is small) only rely on equation equation (18), and we refer to [2] for the detailed analysis that leads to equations (12–16).

6.3 Comparison with static graphs

Maximum entropy static graphs with fixed in- and/or out-degree distributions have been studied in [3,13]. If only the in-degree is fixed, the out-degree is always Poissonian, in contrast with the global geometric out-degree distribution of the evolving cousin model. Let us compare briefly the percolation criteria of the evolving case with two static maximum entropy situations. If $T_-(z) = T(z)$ is the in-degree and $T_+(z)$ the out-degree distribution of the static model, we read from [3,13] that the condition for percolation is $T''(1)T''_+(1) - (T'(1) - T'(1)^2)^2 > 0$[3] When only the in-degree is fixed in the static case, T_+ is Poissonian, one finds $\gamma + \alpha > 1$. When moreover the out-degree is fixed to be independent of the in-degree and follow a geometric law, so that the static graph has the same (global) degree distribution as the evolving graph, one finds $2\gamma + \alpha^2 > 1$. The thresholds are different in the static and evolving cases. Moreover, the percolating region of the evolving case always contains strictly the static percolating region, because when $\gamma < 1/4$ and $\alpha < 1/2$ both $\gamma + \alpha$ and $2\gamma + \alpha^2$ are less than $3/4$: the inhomogeneities of the evolving graph, in which old vertices have an effective high coordinacy, favor the emergence of a giant component. However, this giant component starts with a very tiny size, in contrast with the static case, when its growth is generically linear.

6.4 Comments on the profile of the infinite cluster in the percolating phase

For a Poissonian in-degree distribution, the authors of [2] obtained a closed equation which fitted perfectly with numerical simulations. The naive adaptation of their argument to the general case is straightforward, but gives an incorrect result.

Though we have not been able to derive a closed equation for the profile of the infinite component in the case of an arbitrary degree distribution, the successive derivatives of this profile at $\sigma = 1$ can be computed in a systematic way as follows.

Defining $D \equiv z\partial_z C$, we can derive from the previous results that

$$\sigma\partial_\sigma\rho = (\alpha - zT'(D))\, z\partial_z\rho$$
$$zT(D) - D = (\alpha - zT'(D))z\partial_z D$$
$$\rho(\sigma = 1, z) = zT(D).$$

The first equation is just a rewriting of equation (8), and the second one was obtained by applying the operator $z\partial_z$ to equation (7). The last equation gives the density at $\sigma = 1$. Then the first equation can be used to get the

---

[3] This is, as should be the case for a static graph, symmetric in T and T_+ because $T'(1) = T'_+(1)$ is automatic: any edge is *in* at one end and *out* at the other.

derivative of the density at $\sigma = 1$:

$$
\begin{aligned}
\sigma\partial_\sigma\rho(\sigma = 1, z) &= (\alpha - zT'(D))\, z\partial_z\rho(\sigma = 1, z)\\
&= (\alpha - zT'(D))\, z\partial_z(zT(D))\\
&= (\alpha - zT'(D))\,(zT(D) + zT'(D)z\partial_z D)\\
&= (\alpha - zT'(D))\, zT(D)\\
&\quad + zT'(D)(zT(D) - D)\\
&= \alpha zT(D) - zDT'(D).
\end{aligned}
$$

If $D(z = 1) = 1$, $\rho(\sigma = 1, z = 1) = 1$, but $\sigma\partial_\sigma\rho(\sigma = 1, z = 1) = 0$. The same kind of algebra can be used to compute the successive derivatives $((\sigma\partial_\sigma)^n\rho)(\sigma = 1, z)$ for $n = 2, \cdots$. Again, one can check that this vanishes if $D(z = 1) = 1$. This is of course natural in the non percolating phase.

However, in the percolating phase, $D(z = 1) = 1 - P_\infty$ and one finds

$$
\rho(\sigma = 1, z = 1) = T(1 - P_\infty)
$$
$$
\sigma\partial_\sigma\rho(\sigma = 1, z = 1) = \alpha T(1 - P_\infty) - (1 - P_\infty)T'(1 - P_\infty).
$$

Formally, we can write

$$
\rho(\sigma, z) = e^{\log\sigma[(\alpha - zT'(D))z\partial_z + (zT(D) - D)\partial_D]}zT(D), \quad (22)
$$

where now z and D are independent variables, which is true order by order in an expansion in powers of $\log\sigma$. Writing $\rho_\infty(\sigma) = 1 - \rho(\sigma, z = 1)$ for the profile of the giant component, equation (22) can be used to obtain systematically, for small values of n, a (rather complicated) formula for $((\sigma\partial_\sigma)^n\rho_\infty)(\sigma = 1)$ as a polynomial in $P_\infty, T(1 - P_\infty), T'(1 - P_\infty)\cdots, T^{(n)}(1 - P_\infty)$.

In particular, as a trivial example,

$$
\rho_\infty(\sigma = 1) = 1 - T(1 - P_\infty) \simeq \alpha P_\infty. \quad (23)
$$

On the other hand, for σ close to 0,

$$
1 - \rho_\infty(\sigma) \propto \sigma^\alpha, \quad (24)
$$

so the giant component contains all the old vertices, but only a fraction (which is exponentially small close to the threshold) of the young vertices. That means the percolation transition is very inhomogeneous and takes mainly place in the part of the graph where it is denser than average.

7 Conclusions

In this study, we have solved a model of evolving random graph which, albeit simple, involves an arbitrary in-degree distribution.

We have described the degree distributions and their local-in-time profiles. By construction, in-degree and out-degree at each vertex are independent. The local out-degree distribution follows an age dependent Poisson law, which after integration over ages leads to a geometric global out-degree distribution.

We have also made a detailed analysis of the distribution of component sizes, again at the global and the local-in-time level. We have shown the validity of the self averaging hypothesis by proving the equivalence with a direct exact tree enumeration.

The parameter controlling the percolation transition has been found, quite surprisingly, to be simply the variance of the in-degree distribution, in contrast with the analogous static models.

Below the transition, the large components have a size which scales like a power of the total size of the graph. The size of the giant component close to the threshold has been computed. It is exponentially small. The unusual fluctuation induced percolation mechanism might be the reason why the critical behavior of this models is so different from what is observed in the case of static graphs (see [3,11,13]). There, the generic behavior close to the transition is a linearly growing giant component.

We have shown how all these differences could be used to discriminate in certain cases between a static and an evolving random graph even when the in and out degree distributions are the same for both.

Among the unanswered questions is a direct description of the profile of the giant component. For a Poissonian in-degree distribution, the authors of [2] obtained a closed equation which fitted perfectly with numerical simulations. The naive adaptation of their argument to the general case is straightforward, but gives an incorrect result. It would be desirable to find a valid argument for general T, or more modestly to understand why the argument in fact works for the Poissonian case.

Appendix A: A combinatorial identity

We show that the tree distribution, equations (10, 11), leads to the generating function formula, equation (7).
If V is a finite (nonempty) set, we denote by $\mathcal{T}_V$ the set of trees with vertex set V.

If moreover V is totally ordered , we denote its supremum by s_V. If moreover $|V| \geq 2$, we define $\mathcal{T}_V^n$ for $n = 1, 2, \cdots, |V| - 1$ as the set of trees with vertex set V such that vertex s_V has n neighbors. Note that $\mathcal{T}_V^n$ is non empty.

If $V = [1, \cdots, k]$ (with the usual order) we denote $\mathcal{T}_V$ by $\mathcal{T}_k$ and if $k \geq 2$ we denote $\mathcal{T}_V^n$ by $\mathcal{T}_k^n$ for $n = 1, \cdots, k-1$. We shall sometimes use the shorthand notation $\mathcal{T}_{k_1, \cdots, k_n} \equiv \mathcal{T}_{k_1} \times \cdots \times \mathcal{T}_{k_n}$.

If V is a totally ordered finite (nonempty) set, there is a unique order preserving map from V to $[1, \cdots, |V|]$, so that there is a canonical bijection between $\mathcal{T}_V$ and $\mathcal{T}_{|V|}$. For the same reason, if moreover $|V| \geq 2$ there is a canonical bijection between $\mathcal{T}_V^n$ and $\mathcal{T}_{|V|}^n$ for any $n = 1, 2, \cdots, |V| - 1$.

Let V be a totally ordered finite (nonempty) set V with $|V| \geq 2$ and let $n \in [1, \cdots, |V| - 1]$. Take a tree $\Upsilon \in \mathcal{T}_V^n$.

To Υ we associate the following data.

[i] An n-tuple of positive integers $(k_1, \cdots, k_n)$ such that

$$k_1 + \cdots + k_n = |V| - 1.$$

[ii] A sequence $(\Upsilon_1, \cdots, \Upsilon_n) \in \mathcal{T}_{k_1, \cdots, k_n}$.

[iii] A sequence $(V_1, \cdots, V_n)$ of disjoint subsets of V, each endowed with the order induced from that of V, such that

[a] $(|V_1|, \cdots, |V_n|) = (k_1, \cdots, k_n)$,

[b] $V_1 \cup \cdots \cup V_n = V \backslash \{s_V\}$,

[c] $s_{V_1} < \cdots < s_{V_n}$ for the order in V.

[iv] A sequence $(v_1, \cdots, v_n) \in [1, \cdots, k_1] \times \cdots \times [1, \cdots, k_n]$, as follows.

Remove from Υ the vertex s_V and the edges incident to it. What remains is a forest made of n components. There is a single way to label the corresponding n vertex sets $V_1, \cdots, V_n$ so as to satisfy [c], and then we define $(k_1, \cdots, k_n)$ by [a] so we have obtained [i] and [iii]. For $l \in [1, \cdots, n]$ the connected component with vertex set V_l is a tree. We define Υ_l as its canonical representative in $\mathcal{T}_{k_l}$ and by v_l the vertex of Υ_l whose preimage in V_l is connected to s_V in Υ; this gives [ii] and [iv].

Conversely, one can recover Υ from the data by reversing the procedure.

We let $\mathcal{P}_{k_1, \cdots, k_n}^V$ be the set of sequences $(V_1, \cdots, V_n)$ of disjoint subsets of V, each endowed with the order induced from that of V, satisfying conditions [a] and [b] above. The set obtained when moreover [c] is taken into account is denoted by $\mathcal{P}_{k_1, \cdots, k_n}^{restr\ V}$. There is a 1 to $n!$ correspondence between $\mathcal{P}_{k_1, \cdots, k_n}^{restr\ V}$ and $\mathcal{P}_{k_1, \cdots, k_n}^V$.

To summarize what we have found, we have put $\mathcal{T}_V^n$ in
– 1 to 1 correspondence with the disjoint union

$$\bigcup \mathcal{T}_{k_1, \cdots, k_n} \times \mathcal{P}_{k_1, \cdots, k_n}^{restr\ V} \times [1, \cdots, k_1] \times \cdots \times [1, \cdots, k_n],$$

– 1 to $n!$ correspondence with the disjoint union

$$\bigcup \mathcal{T}_{k_1, \cdots, k_n} \times \mathcal{P}_{k_1, \cdots, k_n}^V \times [1, \cdots, k_1] \times \cdots \times [1, \cdots, k_n],$$

where in both cases the union is taken over n-tuples of positive integers $(k_1, \cdots, k_n)$ such that $k_1 + \cdots + k_n = |V| - 1$.

From now on, we fix $V = [1, \cdots, k]$ for some integer $k \geq 2$. We assign to each tree $\Upsilon \in \mathcal{T}_k$ a weight as follows. Write m_i for be the number of edges of Υ connecting i to $[1, \cdots, i - 1]$. Then

$$\mathcal{W}_\Upsilon = \tau_{m_1} \cdots \tau_{m_k} \int_{0 \leq \sigma_1 \leq \cdots \leq \sigma_k \leq 1} \times d\sigma_1 \cdots d\sigma_k \sigma_1^{\alpha - m_1} \cdots \sigma_k^{\alpha - m_k}.$$

Write $\mathcal{H}_\Upsilon$ for the first factor and $\mathcal{I}_\Upsilon$ for the integral, so that $\mathcal{W}_\Upsilon = \mathcal{H}_\Upsilon \mathcal{I}_\Upsilon$. We change variables and set $\sigma_k = \tilde{\sigma}_k$ and $\sigma_i = \tilde{\sigma}_i \tilde{\sigma}_k$ for $i = 1, \cdots, k - 1$. The power of $\tilde{\sigma}_k$ in

the new integrand is $k - 1 + k\alpha - m_1 - \cdots - m_k = k\alpha$ so integration over $\tilde{\sigma}_k$ leads to $\mathcal{I}_\Upsilon = \frac{1}{1 + k\alpha} \tilde{\mathcal{I}}_\Upsilon$ with

$$\tilde{\mathcal{I}}_\Upsilon = \int_{0 \leq \sigma_1 \leq \cdots \leq \sigma_{k-1} \leq 1} d\sigma_1 \cdots d\sigma_{k-1} \sigma_1^{\alpha - m_1} \cdots \sigma_{k-1}^{\alpha - m_{k-1}}.$$

To avoid ambiguities when several trees are used at the same time, we shall sometimes write m_i^Υ instead of m_i.

Suppose that $\Upsilon \in \mathcal{T}_k^n$, which amounts to set $n = m_k$. We want to express $\mathcal{W}_\Upsilon$ in terms of its decomposition, in fact in term of the n trees $(\Upsilon_1, \cdots, \Upsilon_n) \in \mathcal{T}_{k_1} \times \cdots \times \mathcal{T}_{k_n}$ with $k_1 + \cdots + k_n = |V| - 1$, and of the partition $V_1 \cup \cdots \cup V_n$ (it turns out that $\mathcal{W}_\Upsilon$ does not depend on the choice of one vertex in each V_i).

The decomposition procedure associates to each $i \in [1, \cdots, k - 1]$ one of the trees $\Upsilon_1, \cdots, \Upsilon_n$, say Υ_l and a vertex $i' \in [1, \cdots, k_l]$ in Υ_l. By construction of the decomposition, if $j \in [1, \cdots, i - 1]$ is such that (i, j) is an edge of Υ then i and j have the same Υ_l, $j' \in [1, \cdots, i' - 1]$ and (i', j') is an edge of Υ_l. Hence $m_i^\Upsilon = m_{i'}^{\Upsilon_l}$ and $\mathcal{H}_\Upsilon = \tau_n \mathcal{H}_{\Upsilon_1} \cdots \mathcal{H}_{\Upsilon_n}$ has a simple multiplicative behavior.

Our aim is now to show that when $(\Upsilon_1, \cdots, \Upsilon_n)$ (and then automatically $(k_1, \cdots, k_n)$) are fixed $\sum_{\mathcal{P}_{k_1, \cdots, k_n}^{[1, \cdots, k]}} \tilde{\mathcal{I}}_\Upsilon = \mathcal{I}_{\Upsilon_1} \cdots \mathcal{I}_{\Upsilon_n}$.

We introduce another tree, $\dot{\Upsilon}$, whose decomposition is made of the same trees $(\Upsilon_1, \cdots, \Upsilon_n)$ as Υ, but with $\dot{V}_1 = [1, \cdots, k_1], \dot{V}_2 = [k_1 + 1, \cdots, k_1 + k_2], \cdots$ and $\dot{v}_1 = k_1, \dot{v}_2 = k_1 + k_2, \cdots$. Write $\dot{m}_i$ for the number of edges of $\dot{\Upsilon}$ connecting i to $[1, \cdots, i - 1]$.

There is a unique permutation, say λ, of $[1, \cdots, k - 1]$ which maps V_1 into $\dot{V}_1$, V_2 into $\dot{V}_2$, $\cdots$, and is strictly increasing in each. Then $m_i = \dot{m}_{\lambda(i)}$ for each $i \in [1, \cdots, k - 1]$. If we set $\sigma_i = \dot{\sigma}_{\lambda(i)}$ we obtain

$$\tilde{\mathcal{I}}_\Upsilon = \int_{0 \leq \dot{\sigma}_{\lambda(1)} \leq \cdots \leq \dot{\sigma}_{\lambda(k-1)} \leq 1} \times d\dot{\sigma}_1 \cdots d\dot{\sigma}_{k-1} \dot{\sigma}_1^{\alpha - m_1} \cdots \dot{\sigma}_{k-1}^{\alpha - m_{k-1}}.$$

Write R_λ for the region of integration $0 \leq \dot{\sigma}_{\lambda(1)} \leq \cdots \leq \dot{\sigma}_{\lambda(k-1)} \leq 1$.

Conversely, $(k_1, \cdots, k_n)$ being kept fixed, if $(V_1, \cdots, V_n)$ describes $\mathcal{P}_{k_1, \cdots, k_n}^{[1, \cdots, k]}$ each permutation λ of $[1, \cdots, k - 1]$ such that λ^{-1} is strictly increasing when restricted to $\dot{V}_1, \cdots, \dot{V}_n$ appears exactly once. If R be the union of all such R_λ's (the intersection of different R_λ's is of measure 0), one checks that $(\dot{\sigma}_1, \cdots, \dot{\sigma}_{k-1})$ is in R if and only if $0 \leq \dot{\sigma}_1 \leq \cdots \leq \dot{\sigma}_{k_1} \leq 1$, $0 \leq \dot{\sigma}_{k_1+1} \leq \cdots \leq \dot{\sigma}_{k_1+k_2} \leq 1, \cdots$.

This shows that when $(\Upsilon_1, \cdots, \Upsilon_n)$ are fixed $\sum_{\mathcal{P}_{k_1, \cdots, k_n}^{[1, \cdots, k]}} \tilde{\mathcal{I}}_\Upsilon = \mathcal{I}_{\Upsilon_1} \cdots \mathcal{I}_{\Upsilon_n}$.

If we recall moreover that $\mathcal{H}_\Upsilon = \tau_n \mathcal{H}_{\Upsilon_1} \cdots \mathcal{H}_{\Upsilon_n}$ and that $\mathcal{T}_V^n$ is in 1 to $n!$ correspondence with the disjoint union

$$\cup_{(k_1, \cdots, k_n)} \mathcal{T}_{k_1, \cdots, k_n} \times \mathcal{P}_{k_1, \cdots, k_n}^V \times [1, \cdots, k_1] \times \cdots \times [1, \cdots, k_n],$$

we obtain as an immediate consequence that

$$(1 + k\alpha) \sum_{\Upsilon \in \mathcal{T}_k^n} \mathcal{W}_\Upsilon =$$

$$\sum_{(k_1, \cdots, k_n)} \sum_{(\Upsilon_1, \cdots, \Upsilon_n) \in \mathcal{T}_{k_1, \cdots, k_n}} \frac{\Upsilon_n}{n!} k_1 \mathcal{W}_{\Upsilon_1} \cdots k_n \mathcal{W}_{\Upsilon_n}.$$

Defining $\mathcal{W}^{(k)} \equiv \sum_{\Upsilon \in \mathcal{T}_k} \mathcal{W}_\Upsilon = \sum_n \sum_{\Upsilon \in \mathcal{T}_k^n} \mathcal{W}_\Upsilon$, it is plain that $\sum_k \mathcal{W}^{(k)} z^k$ satisfies equation (7), which has only one formal power series solution vanishing at $z = 0$. QED.

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. M. Bauer, D. Bernard, J. Stat. Phys. **111**(3), 703 (2003)
3. M. Bauer, D. Bernard, *Maximal entropy random networks with given degree distribution*, `ArXiv:cond-mat/0206150`
4. D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman, H. Strogatz, Phys. Rev. E **64**, 041902 (2001)
5. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
6. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. E **64**, 066110 (2001)
7. P. Erdös, A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960)
8. N. Guelzim, S. Bottani, P. Bourgine, F. Képès, Nature Genetics, in press
9. J. Kim, P.L. Krapivsky, B. Kahng, S. Redner, Phys. Rev. E **66**, 055101 (2002)
10. Lee *et al.*, Science **298**, 799 (2002), see also `http://web.wi.mit.edu/young/regulator_network`
11. M. Molloy, B. Reed, Random Struct. Algorithms **6**, 161 (1995); and Comb. Proba. Comput. **7**, 295 (1998)
12. M.E. Newman, S.H. Strogatz, D. Watts, Phys. Rev. E **64**, 026118 (2001)
13. P. Whittle, J. Stat. Phys. **56**, 499 (1989)