Bose-Einstein condensation in random directed networks

O. Sotolongo-Costa[1,3] and G. J. Rodgers[2,3]

[1]*Department of Theoretical Physics, Havana University, 10400 Havana, Cuba*
[2]*Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom*
[3]*Henri Poincaré Chair of Complex Systems, Havana University, 10400 Havana, Cuba*

We consider the phenomenon of Bose-Einstein condensation in a random growing directed network. The network grows by the addition of vertices and edges. At each time step the network gains a vertex with probability p and an edge with probability $1-p$. The new vertex has a fitness (a,b) a,b>0, with probability $f(a,b)$. A vertex with fitness (a,b), with in-degree i and out-degree j, gains a new incoming edge with rate $a(i+1)$ and an outgoing edge with rate $b(j+1)$. The Bose-Einstein condensation occurs as a function of fitness distribution $f(a,b)$.

I. INTRODUCTION

Recently there has been much interest in random growing networks [1–3], both from the point of view of theoretical modeling as well as the empirical study of real networks. There is considerable evidence that while traditional Erdös-Renyi random graphs have a Poisson degree distribution [4], real graphs and random growing graphs have a power-law degree distribution.

Of particular interest are directed networks, which can be used to model systems in which directed flow takes place. Such networks include the world wide web [5], the phone-call graph [6], and the citation graph [7].

As a complement to computer simulations and exact solutions of simplified systems, thermodynamic formulations have been used to study a number of complex cooperative systems including granular media, econophysics, breaking phenomena, and many others. In a recent paper [8], Bianconi and Barabási (BB) mapped the different behavior of random growing networks with fitness onto the thermodynamically distinct phases of a free Bose gas. The fitness model predicts that, in the large network limit, the fittest node will have the most links. This is called the *fit-get-rich* (FGR) phase. Unlike the scale-free model (SFM), in which the degree distribution of the network is power law, the FGR behavior (phase in the thermodynamic language) has nodes with a very large degree which dominate the network. Another phase revealed in Ref. [8] is the Bose-Einstein condensation (BEC), when the fittest node acquires a finite fraction of the total degree. In contrast to BEC, in the FGR phase the richest node is not an absolute winner, since its share of the links (i.e., the ratio of its degree and the total degree of the network) decays to zero for large system sizes, whereas in BEC the winner maintains its share irrespective of system size.

In this case, the fact that this ratio is constant corresponds to the extensivity property of a Bose gas, when the gas keeps a finite fraction of its particles in the ground state. In this paper we examine the phenomenon of Bose-Einstein condensation, which was considered for an undirected graph in Refs. [8,9] on the random directed growing network introduced in Ref. [10].

To do this we work with the model introduced in Ref.

[10], introducing a dependence between the vertex fitness and an energy. As each vertex has two fitnesses (for in-degree and out-degree) then it is necessary to assign two different energy levels to represent each vertex. This is detailed in the first section, together with the description of edges in terms of particles assigned to two energy levels.

The necessary coexistence of two subsystems forming the network and the formulation of an equilibrium condition in a formal way is discussed in detail in Secs. III and IV.

The occurrence of Bose-Einstein condensation, and its interpretation in a physical framework makes it possible to describe the directed network in terms of canonical concepts of statistical physics, such as a phase diagram and first and second order phase transitions. These are introduced in the following sections.

In the conclusions we emphasize the usefulness of this formulation in providing a clear and concise interpretation of different phenomena in networks. So, the possibility of coexistence of phases in which the in-degree distribution exhibits a clear winner whereas the out-degree shows scale-free behavior, or the simultaneous existence of two different winners (bipolarity) and the conditions for its existence are predicted in thermodynamic language.

II. THE MODEL

In Ref. [8], a correspondence between fitness η and energy ϵ given by

$$\eta = e^{-\beta\varepsilon} \tag{1}$$

was introduced. In our model the vertices have, in general, different fitnesses a and b for the addition of an in- or out-degree. As in Ref. [10], here the model consists of the addition of bare vertices (i.e., without edges, but with fitness a for in-degree and b for out-degree) to the network with probability p, and the creation of directed edges between vertices with probability $q = 1 - p$.

A kinetic equation describing the process of directed networks must include the kinetics of both in-degree and out-degree simultaneously since they coexist and influence each

other. Then, the kinetic equation for the mean number of vertices with in-degree i and out-degree j is, as in Ref. [10],

$$\frac{\partial N_{ij}(a,b)}{\partial t} = \frac{qa}{M_1}[iN_{i-1,j}-(i+1)N_{ij}]$$

$$+ \frac{qb}{M_2}[jN_{i,j-1}-(j+1)N_{ij}]+p\delta_{i0}\delta_{j0}f(a,b),$$

$$(2)$$

where the fitnesses a,b are chosen from the fitness distribution $f(a,b)$. The normalization factors M_1 and M_2 are given by

$$M_1 = \sum_{ijab} (i+1)aN_{ij}(a,b) \tag{3}$$

and

$$M_2 = \sum_{ijab} (j+1)bN_{ij}(a,b). \tag{4}$$

The first term in the first square brackets of Eq. (2) represents the increase in N_{ij} when vertices with in-degree $i-1$ and out-degree j gain an incoming edge and the second term represents the corresponding loss. The second square brackets contain the analogous terms for outgoing edges and the last term ensures the continuous addition of new vertices with fitnesses a,b.

The translation of this problem into an energetic formulation is straightforward. We associate the addition of a vertex with fitnesses a,b to the network with the creation of two energy levels; one representing the fitness of the incoming edge and the other for the outgoing edge. This means that the edge can be mapped into two separate isolated subsystems. Then the creation of a directed edge corresponds to the creation of two particles, one in each subsystem, simultaneously. The particles are created in the energy levels corresponding to incoming fitness of the receiving vertex and the outgoing fitness of the emitter vertex.

Then, using the energy variables, rather than the fitness, we have

$$a = e^{-\beta_1\varepsilon_1}, \tag{5}$$

$$b = e^{-\beta_2\varepsilon_2}, \tag{6}$$

and

$$\frac{\partial N_{ij}(\varepsilon_1,\varepsilon_2)}{\partial t} = \frac{qe^{-\beta_1\varepsilon_1}}{M_1}[iN_{i-1,j}-(i+1)N_{ij}]$$

$$+ \frac{qe^{-\beta_2\varepsilon_2}}{M_2}[jN_{i,j-1}-(j+1)N_{ij}]$$

$$+ p\delta_{i0}\delta_{j0}f(\varepsilon_1,\varepsilon_2). \tag{7}$$

And the normalizations are rewritten as

$$M_1 = \sum_{ijab} (i+1)e^{-\beta_1\varepsilon_1}N_{ij}(\varepsilon_1,\varepsilon_2) \tag{8}$$

and

$$M_2 = \sum_{ijab} (j+1)e^{-\beta_2\varepsilon_2}N_{ij}(\varepsilon_1,\varepsilon_2). \tag{9}$$

As we distinguish between incoming and outgoing edges, and their respective fitnesses, when going to the energy representation the edges and their fitnesses must also be distinguished as belonging to different isolated subsystems of the same system. Let us denote the subsystems as 1 and 2. So, the creation of a vertex in the network implies the simultaneous creation of an energy level in each sub-system, i.e., ε_1 and ε_2.

The creation of an edge joining two vertices implies the simultaneous creation of one particle in each subsystem in the energy levels corresponding to the fitnesses of the vertices gaining an in- and out-degree.

It should be noted that we are doing a simultaneous description of two isolated subsystems which compose the whole system (the network). These subsystems do not exchange energy or particles, but are correlated in the sense that the creation of an energy level (a particle) in one of the subsystems implies the simultaneous phenomenon in the other.

By definition of the model, the total number of particles increases linearly with time, so

$$\sum_{i,j,\varepsilon_1,\varepsilon_2} N_{ij}(\varepsilon_1,\varepsilon_2) = pt. \tag{10}$$

Let us define the reduced moments m_1 and m_2 by

$$M_1(t) = m_1 t, \tag{11}$$

$$M_2(t) = m_2 t, \tag{12}$$

and introduce n_{ij} as

$$N_{ij}(t) = n_{ij}t. \tag{13}$$

As shall be seen these magnitudes will be useful to calculate the characteristics of the network. Then, following the procedure in Ref. [10], we can show that

$$m_1 = \frac{q}{m_1} \sum_{i,j,\varepsilon_1,\varepsilon_2} e^{-2\beta_1\varepsilon_1}(i+1)n_{ij}$$

$$+ p\int e^{-\beta_1\varepsilon_1}f(\varepsilon_1,\varepsilon_2)d\varepsilon_1 d\varepsilon_2 \tag{14}$$

and

$$m_2 = \frac{q}{m_2} \sum_{i,j,\varepsilon_1,\varepsilon_2} e^{-2\beta_2\varepsilon_2}(j+1)n_{ij}$$

$$+ p\int e^{-\beta_2\varepsilon_2}f(\varepsilon_1,\varepsilon_2)d\varepsilon_1 d\varepsilon_2, \tag{15}$$

which determines m_1 and m_2 as functions of p, β_1, and β_2.

A. The coexistence of two systems

In this representation, as we already pointed out, the addition of a vertex means the creation of two energy levels, each corresponding to its fitness. When an edge joining two vertices is created, we add one particle in each corresponding level.

The distribution of in- and out-degree then corresponds to the distribution of particles in systems 1 and 2,

$$g_i = \sum_j n_{ij} = p \frac{\Gamma(i+1)\Gamma(1+e^{\beta(\varepsilon_1-\mu_1)})}{\Gamma(i+2+e^{\beta(\varepsilon_1-\mu_1)})} e^{\beta(\varepsilon_1-\mu_1)} f(\varepsilon_1,\varepsilon_2)$$

(16)

is the number of vertices with in-degree i, and

$$h_j = \sum_i n_{ij} = p \frac{\Gamma(j+1)\Gamma(1+e^{\beta(\varepsilon_2-\mu_2)})}{\Gamma(j+2+e^{\beta(\varepsilon_2-\mu_2)})} e^{\beta(\varepsilon_2-\mu_2)} f(\varepsilon_1,\varepsilon_2)$$

(17)

is the number of vertices with out-degree j. For large degree their asymptotic expressions are

$$g_i \sim i^{-(1+e^{\beta(\varepsilon_1-\mu_1)})}$$

(18)

and

$$h_j \sim j^{-(1+e^{\beta(\varepsilon_2-\mu_2)})},$$

(19)

where we have introduced

$$\mu_k = \frac{m_k}{q}$$

(20)

for $k=1$ and $k=2$ as *chemical potentials*. Unlike BB, the chemical potential is introduced here as an exact expression, not as an asymptotic one.

B. "Equilibrium" condition

The equations for m_1 and m_2 can be transformed by introducing the generating function

$$g(x,y,\varepsilon_1,\varepsilon_2) = \sum_{i,j} x^i y^j n_{ij}(\varepsilon_1,\varepsilon_2),$$

(21)

where

$$g(1,1,\varepsilon_1,\varepsilon_2) = p f(\varepsilon_1,\varepsilon_2)$$

(22)

with

$$\frac{\partial g}{\partial x}(1,1,\varepsilon_1,\varepsilon_2) = \frac{p e^{-\beta_1 \varepsilon_1}}{e^{-\beta_1 \mu_1}[1-e^{-\beta_1(\varepsilon_1-\mu_1)}]}$$

(23)

and

$$\frac{\partial g}{\partial y}(1,1,\varepsilon_1,\varepsilon_2) = \frac{p e^{-\beta_2 \varepsilon_2}}{e^{-\beta_2 \mu_2}[1-e^{-\beta_2(\varepsilon_2-\mu_2)}]}.$$

(24)

Then the equations for m_k ($k=1,2$) become

$$\frac{p}{q} \int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_k(\varepsilon_k-\mu_k)}-1} = 1.$$

(25)

This means that

$$\int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_1(\varepsilon_1-\mu_1)}-1} = \int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_2(\varepsilon_2-\mu_2)}-1}.$$

(26)

Unlike BB, this is a different condition generated by the model. Let us denote it as a *generalized equilibrium condition*. A *strong equilibrium* condition implies that

$$\varepsilon_1 = \varepsilon_2$$

(27)

and

$$\mu_1 = \mu_2.$$

(28)

This condition will be discussed later. In this formalism the occupation probability is, then,

$$n_k(\varepsilon_k,\mu_k) = \frac{p}{q} \frac{1}{e^{\beta_k(\varepsilon_k-\mu_k)}-1}$$

(29)

for $k=1,2$, representing the probability for a given edge to belong to a given vertex (probability of a particle to belong to an energy level). If, now, the fitness distribution is separable then

$$f(\varepsilon_1,\varepsilon_2) = f_1(\varepsilon_1) f_2(\varepsilon_2)$$

(30)

and

$$\frac{p}{q} \int \frac{d\varepsilon_k f(\varepsilon_k)}{e^{\beta_k(\varepsilon_k-\mu_k)}-1} = 1$$

(31)

separately for each system $k=1,2$. In each case, μ_k is the solution of this equation, which depends on the density of states $f_k(\varepsilon)$.

C. Bose-Einstein condensation

To illustrate, let us start with a particular case, when $f(\varepsilon) = C\varepsilon^\theta$ with $0 < \varepsilon < \varepsilon_{\max}$. The normalization $C = (\theta+1)/\varepsilon_{\max}^{\theta+1}$ as in Ref. [8].

Then the integral in Eq. (25) becomes

$$I(\beta,\mu) = \frac{p(\theta+1)}{q \varepsilon_{\max}^{\theta+1}} \int \frac{d\varepsilon \epsilon \epsilon^\theta}{e^{\beta(\varepsilon-\mu)}-1} = 1.$$

(32)

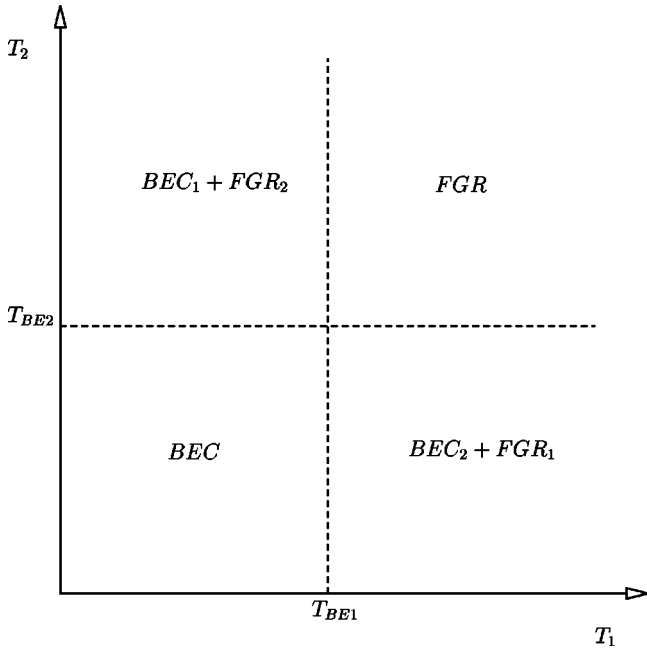But this integral reaches its maximum value when $\mu = 0$. Then, if for a given value of β we have

FIG. 1. Phase diagram.

FIG. 2. First and second order phase transitions.

$$\int \frac{d\varepsilon f(\varepsilon)}{e^{\beta\varepsilon}-1}<1, \tag{33}$$

then we have a BEC. If we calculate the temperature for BEC in this particular case, we obtain

$$T_{BE}=\frac{1}{\varepsilon_{\max}}\left\{\frac{p}{q}\Gamma(\theta+2)\zeta(\theta+1)\right\}^{-(1/\theta+1)}. \tag{34}$$

Unlike BB, the factor (p/q) appears here. If $p=q$ we recover BB result. Even in this simple example, T_{BE} may be different for the subsystems 1 and 2, if the value of θ is different. In a more general form, we can state the condition for Bose-Einstein condensation. Then, as

$$I(\beta_k,\mu_k)=\frac{p}{q}\int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_k(\varepsilon_k-\mu_k)}-1}=1 \tag{35}$$

for $k=1,2$, if $I(\beta_k,\mu_k)<1$ then BEC occurs. The conditions for BEC depend on $f(\varepsilon_1,\varepsilon_2)$.

D. Phase diagram

Below T_{BE} there is BEC. Above it the phase is fit-get-richer as in Ref. [8], but here both phases must be considered for both systems. Then, a phase diagram can be drawn with the temperatures $T_1=(1/\beta_1)$ and $T_2=(1/\beta_2)$ characterizing each subsystem (Fig. 1).

Imposing a given dependence of fitness with energy so that

$$\int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_1(\varepsilon_1-\mu_1)}-1} \neq \int \frac{d\varepsilon_1 d\varepsilon_2 f(\varepsilon_1,\varepsilon_2)}{e^{\beta_2(\varepsilon_2-\mu_2)}-1} \tag{36}$$

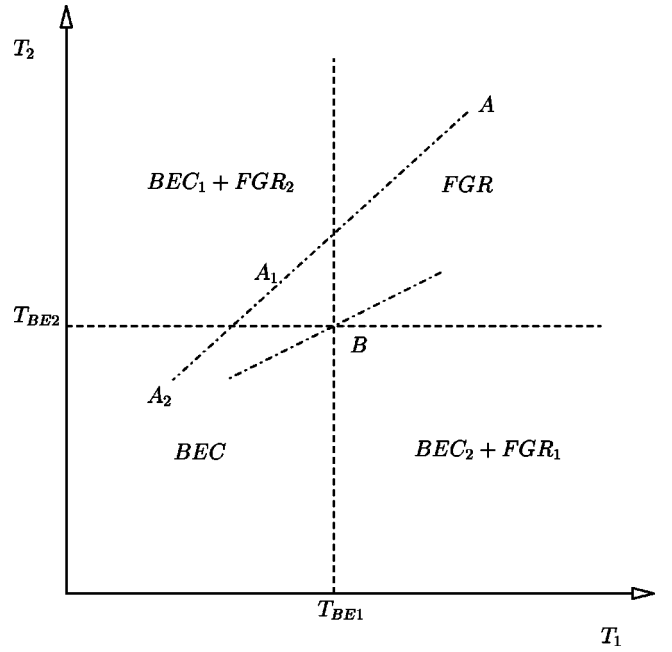means that the equilibrium is broken [i.e., in a computer simulation we would impose temperatures that do not satisfy Eq. (26)]. Then there can be a winner in one of the phases. One of the vertices accumulates, say, most of the in-degrees, coming from the other vertices. If both phases are in BEC there is a vertex with most of the incoming and another with most of the outgoing edges. Then both vertices will be highly correlated.

E. Out of equilibrium first order phase transition

Although more complicated situations are possible, let us illustrate the case $T_{BE1}=T_{BE2}$. In Fig. 2 point A represents the case when both subsystems are in the FGR phase. If, for example, we make the simulations with the parameters given by A_1 then there will be a FGR and a BEC phase coexisting, but not *in equilibrium* since one of them satisfies Eq. (35) but not the other. There is a coexistence of phases out of equilibrium. We might call this an *out of equilibrium first order phase transition*. The motion along $AA_1A_2\cdots$ leads to the region in which BEC occurs for both phases. All we know in the BEC region is that none of the subsystems satisfies Eq. (35).

The concept of equilibrium in this model is purely formal since the subsystems are not in contact, but it is a comfortable tool to make a map between directed networks and Bose-Einstein statistics.

F. The scale-free model

If one of the systems is such that $f(\varepsilon)=\delta(\varepsilon)$, then Eq. (31) gives

$$\frac{p}{q}\frac{1}{e^{-\beta\mu}-1}=1, \tag{37}$$

then

$$g_i \sim i^{-(1+e^{-\beta\mu})} \sim i^{-(2+p/q)} \tag{38}$$

with a corresponding expression for h_j. If $p=q$ the scale-free model is recovered with a degree distribution $P_k \sim k^{-3}$. Thus the model with directed edges gives the possibility of a scale-free phase for both in the in-degree and the out-degree.

G. Second order phase transition

If in Fig. 2 we move in such a way as to cross into the BEC region through the point B then BEC occurs simultaneously for both phases. No coexistence of phases will occur, the system goes from FGR to BEC or vice-versa instantly. This is a *second order phase transition* since the system as a whole changes its phase. Then the point B is the *critical point* for the transition.

BEC is itself a second order phase transition since it affects the system as a whole, so that when subsystem 1 or 2 experiences BEC, the subsystem itself suffers a second order phase transition, but our system contains both subsystems 1 and 2 so that when they are in different phases we consider our whole system as composed of two phases.

III. DISCUSSION

We have studied Bose-Einstein condensation in a random growing network model. In this model, with a formulation considering incoming and outgoing edges, a formal analogy can be adopted to describe the change in the behavior of the directed edges in terms of Bose statistics.

In this formulation, as in that of BB, the temperature plays the role of a dummy variable. Real systems are characterized by the functional form of the distribution function of fitness and there is no β to consider. It only emerges when we translate from the language of fitness to that of energies. But the formulation of networks in terms of quantum statistics and the introduction of temperature is an elegant and simple description of the behavior of networks under different conditions. In this sense, the temperature plays its role in the simulations of the networks, determining the strength of the dependence between the fitness and its associated energy, and this dependence is reflected in the behavior of the network by determining its position in the phase diagram.

Though phase diagrams and phase transitions are not exclusive to thermodynamic formulations (see, e.g., Dorogovtsev and Mendes [11]), such a formulation is a very powerful tool to describe different phases of random directed networks, where the presence of directed edges make the description in terms of fitness difficult. Considering this network as a thermodynamic system, makes it simple to understand, provided we appropriately interpret the thermodynamic parameters.

ACKNOWLEDGMENTS

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] P.L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[3] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).

[4] B. Bollobas, *Random Graphs* (Academic Press, London, 1985).

[5] A. Broder, R. Kumar, F. Maghul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).

[6] W. Aiello, F. Chung, and L. Lu, in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, Portland, OR, 2000* (ACM Publishing, New York, 2000), pp. 171–180.

[7] S. Redner, Europhys. Lett. **4**, 131 (1998).

[8] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).

[9] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).

[10] G. Ergun and G.J. Rodgers, Physica A **303**, 261 (2002).

[11] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **63**, 025101 (2001).