

## Continuum formalism for modeling growing networks with deletion of nodes

Joan Saldaña\*

*Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona, 17071-Girona, Spain*

(Received 25 August 2006; published 8 February 2007)

We present a continuum formalism for modeling growing random networks under addition and deletion of nodes based on a differential mass balance equation. As examples of its applicability, we obtain new results on the degree distribution for growing networks with a uniform attachment and deletion of nodes, and complete some recent results on growing networks with preferential attachment and uniform removal.

DOI: [10.1103/PhysRevE.75.027102](https://doi.org/10.1103/PhysRevE.75.027102)

PACS number(s): 89.75.Hc, 87.23.-n, 89.75.Fb

The subject of growing networks has received a great wave of interest since the publication, in 1999, of a model with preferential attachment during the addition of new nodes [1]. This is the main mechanism that explains the power-law distribution of degrees appearing in the description of many complex networks. Until that moment and although this paper was not the first one proposing this sort of linking mechanism [2–4], the main interest was focused on random models for networks having a fixed set of nodes. None of these models considering a fixed number of nodes, however, were able to reproduce (by adding and/or rewiring links) such a power-law distribution, an important feature of the architecture of complex networks (but see Ref. [5].), in contrast to the suitability of such models to explain another important characteristic of such networks, the so-called “small-world” property.

This new way of studying complex networks is no longer static. Instead, it is intrinsically dynamic and, accordingly, different theoretical approaches have been carried out to describe the dynamics of such growing networks [6]. In all these approaches it is assumed a network growth based on a regular (in time) introduction of new nodes. The first method, already proposed in Ref. [1], assumes that the (expected) degree  $k_i$  of the node  $i$  is a continuum variable with a rate of change proportional to the probability with which a new node attaches to it during the network growth. Then, once the evolution in time of the expected degree of the node  $i$ ,  $k_i(t)$ , is known, the network degree distribution  $p(k, t)$  is obtained from the hypothesis of a regular addition of nodes, which leads to a uniform distribution of introduction times  $\{t_i\}$ , and the fact that, by definition,  $p(k, t) = \partial P[k_i(t; t_i) < k] / \partial k$ . The second analytical method, introduced in Ref. [7], is based on the so-called “master-equation approach,” a gain-loss equation for the probability  $p(k, t_i, t)$  that, at time  $t$ , a node  $i$  introduced at time  $t_i$  has a degree  $k$ . The third theoretical approach, introduced in Ref. [8], is the so-called “rate-equation approach,” which deals with the rate of change of the average number  $n_k(t)$  of nodes in the network with degree  $k$  at time  $t$ . The last two approaches are completely equivalent and lead to the same asymptotic results on  $p(k, t)$  as the one based on the analysis of  $dk_i(t)/dt$  [6].

Surprisingly, after Ref. [9] and until the recent publications [10,11] and the very recent [12], growing network models only deal with the pure addition of nodes, although papers on duplication models in biology also consider deletion of links during the growth of the network [13–15]. However, realistic models of networks usually encounter both addition and removal of nodes. In fact, only few kinds of networks such as, for instance, citation networks or networks of collaborations can be described realistically without removal of nodes. Other classical examples of complex networks like the Internet or the WWW have a varying composition of nodes (computers or documents, respectively) due to the appearance of new computers or documents but, at the same time, to the disappearance of previously existing ones (computer breakdowns). In Ref. [16] a uniformly random removal of nodes is considered in relation with error and attack tolerance of scale-free networks. In Refs. [17,18] such an approach has been extended to analyze the network robustness under a random deletion of nodes depending on their connectivity.

The formalism we present in this paper provides a framework for obtaining the degree distribution of growing networks with deletion of nodes. This framework uses the continuum approximation of the rate equation for  $n_k(t)$  which yields, after passing to the continuum limits of  $k$  and  $t$ , to a first-order partial-differential equation for  $n(k, t)$ , the density of nodes with degree  $k$  at time  $t$ . This formulation allows for a general treatment of nonlinear models that cannot be done by using the generating function formalism for  $n_k(t)$  [9]. An example of this sort of model is given when the removal probability of a selected node at time  $t$  depends in a non-trivial manner on the degree distribution  $n_k(t)$  existing at this time (see the end of the paper for a particular choice of such a dependency). On the other hand, a similar continuum approach has been done for the master equation for  $p(k, t_i, t)$  in Ref. [19]. However, the fact that this is an equation for the evolution of the degree of a single node while we are interested in a general formulation for the evolution of  $n(k, t)$  avoids any overlapping of results.

The idea behind this continuum formalism lies in the notion of a continuously *structured population*, which has been widely used in cell biology and ecology since the 1970s [20]. According to this notion, a growing random network is nothing but a collection of nodes structured by a continuous variable  $k$  that changes during their life span following a certain “growth” rate  $dk/dt = g(k, t)$ . While in biology age and size

\*Electronic address: [jsaldana@ima.udg.es](mailto:jsaldana@ima.udg.es)

are the most important examples of structuring variables, in growing networks the node degree plays this role. In fact, although other node features like age or fitness have also been considered elsewhere [5,8,21,22], such features usually act as dependent variables of the node degree, the true structuring variable in complex networks (but see Ref. [23]). Hence, the description of such a network is given by a density of nodes  $n(k,t)$  with respect to node degree  $k$ , and its dynamics is governed by a mass balance equation for  $n(k,t)$  of the form

$$\frac{\partial}{\partial t}n(k,t) + \frac{\partial}{\partial k}[g(k,t)n(k,t)] = -\alpha(k,t), \quad (1)$$

with  $\alpha(k,t)$  the deletion rate of nodes with degree  $k$  at time  $t$ . Moreover, if attached nodes are assumed to have all the same initial degree  $k_0$ , we need to give a boundary condition at  $k=k_0$  that reflects the inflow of new nodes into network, namely,

$$g(k_0,t)n(k_0,t) = f(t), \quad t > 0.$$

In particular, when the network growth takes place under an addition of one node of degree  $k_0$  at each time step, as in most of the models of growing networks that have been published so far, then  $f(t) \equiv 1 \quad \forall t > 0$ .

If  $N(t)$  is the number of nodes in the network at time  $t$ , and  $\lim_{k \rightarrow \infty} g(k,t)n(k,t) = 0$  for all  $t > 0$  (i.e., the flow of nodes with very large degrees vanishes), then integrating Eq. (1) with respect to  $k$  from 0 to  $\infty$  yields

$$\frac{d}{dt}N(t) = f(t) - \int_0^{\infty} \alpha(k,t)dk.$$

In particular, if a growing network model assumes that, on average, a constant number of nodes  $r_0$  is removed from the network per unit of time, then  $\alpha(k,t)$  has to satisfy the normalization

$$\int_0^{\infty} \alpha(k,t)dk = r_0, \quad \forall t > 0. \quad (2)$$

Hence,  $N(t) = \int_0^t f(s)ds - r_0t + N_0$  with  $N_0$  the initial network size.

On the other hand, although most of the models assume a fixed degree for the newcomer nodes, it is also possible to consider a continuous distribution  $\beta(k,t)$  for the initial degree at attachment of new nodes within the same transportlike framework. In this case, the boundary condition at  $k=k_0 \geq 0$  is simply given by  $g(k_0,t)n(k_0,t) = 0 \quad \forall t > 0$  while the mass balance equation becomes

$$\frac{\partial}{\partial t}n(k,t) + \frac{\partial}{\partial k}[g(k,t)n(k,t)] = \beta(k,t) - \alpha(k,t). \quad (3)$$

If, as before, the total number of new nodes per unit of time is given by  $f(t)$ , then it follows that

$$\int_{k_0}^{\infty} \beta(k,t)dk = f(t), \quad \forall t > 0.$$

It is worth noting that, in this modeling approach, it is

implicitly assumed that any node with degree  $k$  has a similar neighborhood surrounding it. This is a sort of mean-field approximation to the spatial distribution of nodes, also present in the discrete approximations, which is suitable regardless of the kind of mixing of degrees (assortative or disassortative) in the network. The point is that this assumption allows for the same degree growth rate of those nodes having the same degree. In particular, when a fixed initial degree  $k_0$  is considered for the independently added nodes, the degree growth rate is

$$\frac{dk(t)}{dt} = k_0 f(t) \Pi(k,t) - \sum_{j \in \mathcal{N}_k} r(j,t) \nu(j,t), \quad (4)$$

where  $\mathcal{N}_k$  denotes the set of  $k$  neighbors of a node of degree  $k$ ,  $\Pi(k,t)$  is the probability that a given link of a newly added node attaches to an extant node of degree  $k$ ,  $r(j,t)$  is the probability of the neighbor node  $j$  being removed given that this node has been already selected, and  $\nu(j,t)$  is the probability of selecting the neighbor  $j$ . Note that, each time a node is removed, all the links incident on it are also removed, which causes a decrease in the degree of all the neighbors of that node. This is the reason for the sum in the second term in the right-hand side (RHS) of Eq. (4), which accounts for the expected number of neighbors of a node of degree  $k$  that are removed per time unit.

In general, the removal probability of a node  $j$  in Eq. (4), namely,  $r(j,t)\nu(j,t)$ , is a function of the node degree  $k_j$  that can be either increasing or decreasing when deletion of nodes is not uniformly random. This can happen when such a removal is due to intentional external attacks to the network, with the most efficient attempts being those where the highest degree nodes are removed with the highest or the lowest probability, respectively (see Ref. [18] for a discussion).

Finally, in order to have a completely specified model, an expression for  $\alpha(k,t)$  in Eq. (1) is needed. If  $d(k,t)$  denotes the probability of selecting a node of degree  $k$ , and the conditional probability  $r$  of removing a selected node depends on the node degree, then

$$\alpha(k,t) = r(k,t)d(k,t).$$

For instance, under a uniformly random selection of one node and if  $r$  is the same for any chosen node and for any time, then

$$\alpha(k,t) = r \frac{n(k,t)}{N(t)},$$

which satisfies Eq. (2) as long as  $r=r_0$ . In this case, moreover,  $\nu(j,t) = \nu(t) = 1/N(t)$ .

This formalism can be applied in a straightforward manner to obtain the degree distribution for the simplest models of growing networks. From now on, we will assume a uniform removal of one node per time unit with probability  $r_0$ , that is,  $\nu(t) = 1/N(t)$  and  $r=r_0 < 1$ , and a regular addition of a new node with  $k_0$  links per time unit [ $f(t) \equiv 1$ ]. Under these hypotheses and for large values of  $t$ ,  $N(t) \approx (1-r)t$  and the total number of links  $L(t) = (k_0 - r\bar{k})t + L_0 \approx k_0 \frac{1-r}{1+r}t$ , where

$\bar{k}=2L(t)/N(t)$  is the network average degree. These approximations will allow for simpler expressions of the solutions to the examples below.

When the network growth takes place with nonpreferential attachment of the new nodes, the model equation (1) reads

$$\frac{\partial}{\partial t}n(k,t) + \frac{\partial}{\partial k}\left(\frac{k_0 - rk}{N(t)}n(k,t)\right) = -r\frac{n(k,t)}{N(t)}. \quad (5)$$

Integrating Eq. (5) along the characteristic curves, which are the solutions of

$$\frac{d}{dt}k(t;t_0) = \frac{k_0 - rk(t)}{N(t)}, \quad k(t_0;t_0) = k_0, \quad (6)$$

with the boundary condition  $g(k_0,t)n(k_0,t)=1 \forall t>0$  and using that  $N(t) \approx (1-r)t$ , it follows that the solution to Eq. (5) for large  $t$  is given by

$$n(k,t) = \frac{t}{k_0}(1-r)^{-(1-r)t/r} \left(1 - \frac{rk}{k_0}\right)^{(1-r)/r}, \quad (7)$$

for  $k_0 \leq k < k_0/r$ , and  $n(k,t)=0$  otherwise. Interestingly, from Eq. (5) it follows that the density of nodes along characteristics  $n[k(t),t]$  remains constant since

$$\frac{d}{dt}n[k(t),t] = \frac{\partial}{\partial t}n(k,t) + \frac{k_0 - rk}{N(t)} \frac{\partial}{\partial k}n(k,t) = 0.$$

This is so because both the addition and the deletion of nodes take place uniformly all over the network. Moreover, from Eq. (6) it follows that the expected degree of nodes is restricted to take strictly positive values in a *finite interval*, the length of which tends to infinity as  $r \rightarrow 0$ , and tends to 0 as  $r \rightarrow 1^-$ . Now, from Eq. (7) and since the degree distribution  $p(k,t) = n(k,t)/N(t) \approx n(k,t)/(1-r)t$  for  $r \in [0,1)$  and large enough  $t$ , one obtains that, for  $k_0 \leq k < k_0/r$ ,

$$p(k,t) = k_0^{-1}(1-r)^{-1/r} \left(1 - \frac{rk}{k_0}\right)^{(1-r)/r}, \quad (8)$$

and  $p(k,t)=0$  otherwise. From this expression, it follows that  $p(k,t) \rightarrow \delta(k-k_0)$  as  $r \rightarrow 1^-$ , i.e., in nongrowing networks the *expected* degree of each node will be  $k_0$  for large  $t$ . When comparing this result with the exact solution to the discrete version of this model obtained in Ref. [12], it follows that in that case the asymptotic degree distribution is concentrated around  $k_0$  and has tails that decay rapidly (in fact, faster than any exponential). On the other hand,  $p(k,t) \rightarrow k_0^{-1}e^{-k/k_0}$  as  $r \rightarrow 0$ , which is exactly the same expression obtained by using the first of the previously described analytical methods for studying growing networks [6,19]. Therefore, Eq. (8) is the natural generalization of the degree distribution of the so-called exponential networks [19] to networks with a uniform deletion of nodes.

When the network growth takes place with a linear preferential attachment (LPA) of the new nodes as the one in Ref. [1], for large  $t$  and  $r \in [0,1)$  the model equation (1) now reads

$$\frac{\partial}{\partial t}n(k,t) + \frac{\partial}{\partial k}\left(\frac{k}{2t}n(k,t)\right) = -r\frac{n(k,t)}{(1-r)t}. \quad (9)$$

Curiously, in this case the characteristic curves do not depend on the value of  $r$  since now Eq. (4) is given by  $dk(t;t_0)/dt = k/2t$  with  $k(t_0;t_0) = k_0$ . This also implies that the expected degree of a node takes values from  $k_0$  to  $\infty$ . Proceeding along the same lines as in the previous example and using that  $p(k,t) \approx n(k,t)/(1-r)t$  with  $n(k,t)$  a solution to Eq. (9) satisfying  $g(k_0,t)n(k_0,t)=1 \forall t>0$ , it follows that, for  $k \geq k_0$  and large  $t$ ,

$$p(k,t) = \frac{2}{k_0(1-r)} \left(\frac{k}{k_0}\right)^{(r-3)/(1-r)}.$$

Therefore, the scale-free character of a growing network with LPA is not destroyed by a nonselective removal of nodes although the exponent, which is the same that the one obtained in Ref. [12] for the tail of the degree distribution, is always lower than  $-3$  for  $r > 0$  and unbounded from below as  $r \rightarrow 1^-$  (see also Ref. [9]). Moreover, when  $r \rightarrow 0$ , it is recovered the degree distribution of a growing network with LPA obtained in Ref. [1], namely,  $p(k) = 2k_0^2k^{-3}$ , for  $k \geq k_0$  and  $t$  large enough. Note that, when  $r \rightarrow 1^-$ , the network size is constant [ $N(t) \equiv N_0$ ] but not the number of links, which is approximately given by  $L(t) = N_0(k_0t + L_0)/(2t + N_0)$ . Therefore, in this limiting case, Eq. (4) becomes

$$\frac{d}{dt}k(t;t_0) = \frac{k}{2t + \bar{k}_0} \left(1 - \frac{\bar{k}_0}{k_0}\right), \quad (10)$$

with  $k(t_0;t_0) = k_0$  and  $\bar{k}_0 = 2L_0/N_0$  the initial average degree of the network. Hence,  $dk(t)/dt$  will be positive (negative) for  $t < \infty$  as long as  $\bar{k}_0 < (>)k_0$ , inequality that only depends on the initial configuration of the network. In the first case, the range of expected degrees for the added nodes is  $[k_0, \infty)$  while, in the second one, it is  $(0, k_0]$ . In both cases, upon introducing Eq. (10) into the model equation it follows that, for the added nodes in the network,

$$p(k,t) = \frac{2t + \bar{k}_0}{N_0|k_0 - \bar{k}_0|} \left(\frac{k_0}{k}\right)^{\gamma+1} \exp\left\{\frac{2t + \bar{k}_0}{2N_0} \left[\left(\frac{k_0}{k}\right)^\gamma - 1\right]\right\}, \quad (11)$$

where  $\gamma = 2k_0/(k_0 - \bar{k}_0)$ , which tends to a  $\delta(k-k_0)$  as  $t \rightarrow \infty$  since  $(k_0/k)^\gamma - 1 < 0$  for  $k \neq k_0$ . When comparing this result with the exact solution to the discrete version of this model for nongrowing networks obtained in Ref. [12], it follows that in that case the asymptotic degree distribution is concentrated around  $k_0$  and has a stretched-exponential shape. Instead, the exponential term in Eq. (11) vanishes as  $t \rightarrow \infty$ .

The last example deals with a uniform distributed initial degree at attachment between  $k_0$  and  $k_1 > k_0$ . In this case, the expected initial degree of a new node is  $\bar{k}_a = (k_0 + k_1)/2$  and, under a nonpreferential addition of one node each time step, Eq. (4) becomes for  $r \in [0,1)$ ,

$$\frac{d}{dt}k(t;t_0) = \frac{\bar{k}_a - rk(t)}{(1-r)t}, \quad k(t_0;t_0) = k_0, \quad (12)$$

and the model equation (3) yields

$$\frac{\partial}{\partial t}n(k,t) + \frac{\bar{k}_a - rk}{(1-r)t} \frac{\partial}{\partial k}n(k,t) = \frac{H(k_1 - k)}{k_1 - k_0}, \quad (13)$$

where  $H(x)$  is the Heaviside function, with the boundary condition  $n(k_0, t) = 0$  for  $\forall t > 0$ . Note that, now, the solution is no longer constant along characteristics for  $k_0 < k < k_1$  due to the uniform addition of new nodes within this range of degrees. Integrating Eq. (13) along characteristics under the assumption  $r < (k_0 + k_1)/2k_1$  to guarantee that  $\lim_{t \rightarrow \infty} k(t; t_0) > k_1$  [see Eq. (12)], and using that  $p(k, t) \approx n(k, t)/(1-r)t$  for large  $t$  as before, it follows

$$p(k, t) = \frac{1/(1-r)}{k_1 - k_0} \left\{ \left[ 1 - \left( \frac{\bar{k}_a - rk_0}{\bar{k}_a - rk} \right)^{(r-1)/r} \right] H(k_1 - k) + \left[ \left( \frac{\bar{k}_a - rk_1}{\bar{k}_a - rk} \right)^{(r-1)/r} - \left( \frac{\bar{k}_a - rk_0}{\bar{k}_a - rk} \right)^{(r-1)/r} \right] \times H(k - k_1) \right\}, \quad (14)$$

with  $k_0 \leq k \leq \bar{k}_a/r$ , and  $p(k, t) = 0$  otherwise. Note that for  $r \rightarrow 0$ , Eq. (14) amounts to

$$p(k, t) = \frac{1}{k_1 - k_0} \left\{ (1 - e^{(k_0 - k)/\bar{k}_a}) H(k_1 - k) + (e^{k_1/\bar{k}_a} - e^{k_0/\bar{k}_a}) e^{-k/\bar{k}_a} H(k - k_1) \right\},$$

and that, from this expression, it is easy to see that we re-

cover the exponential degree distribution (i.e.,  $p(k) = k_0^{-1} e^{-k/k_0}$ ,  $k \geq k_0$ ) when  $k_1 \rightarrow k_0$ , which constitutes an example of the consistency of this approach.

In this Brief Report we have presented a continuum formulation for the evolution of the number of nodes with degree  $k$  at time  $t$  in growing networks with the deletion of nodes. Although it has been recognized elsewhere that analytical methods that do not use a continuum assumption appear more suitable to obtain exact results in network models [6], the simplicity and elegance of the formulation and resolution of models based on the continuum approach make them a good alternative to the former as long as  $r < 1$ . However, in the limiting case  $r = 1$ , this approach offers a poorer description of the evolution of (nongrowing) networks (cf. Ref. [12]) since it has been specially derived for systems whose evolution in time can be described as a transport process in a space of traits (age or size in population biology or node degree in complex networks). On the other hand, the theory of nonlinear mass balance equations allows for a good chance for dealing with challenging network models. A nice example could be the assumption of a conditional removal probability  $r$  depending on a weighted integral of the node density  $\int_{k_0}^{\infty} \omega(k) n(k, t) dk$  in such a way that the higher the value of this integral, the higher the value of  $r$ . In the case of an increasing (decreasing)  $\omega(k)$ , this assumption represents that, instead of being independent of the network configuration, the probability of a node breakdown increases with the presence of highly (poorly) connected nodes in the network.

The author thanks Albert Avinyó, Narcís Clara, Josep Lluís Garcia-Domingo, David Juher, and Marta Pellicer for discussions on complex networks and on some contents of this Brief Report. This work has been partially supported by Grant No. MTM2005-07660-C02-02 of the Spanish government.

- 
- [1] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).  
 [2] G. Yule, *Philos. Trans. R. Soc. London, Ser. B* **213**, 21 (1925).  
 [3] H. A. Simon, *Biometrika* **42**, 425 (1955).  
 [4] D. J. de S. Price, *Science* **149**, 510 (1965).  
 [5] G. Caldarelli, A. Capocci, P. DeLosRios, and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).  
 [6] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).  
 [7] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).  
 [8] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).  
 [9] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **63**, 056125 (2001).  
 [10] F. Chung and L. Lu, *Internet Math.* **1**, 409 (2004).  
 [11] C. Cooper, A. Frieze, and J. Vera, *Internet Math.* **1**, 463 (2004).  
 [12] C. Moore, G. Ghoshal, and M. E. J. Newman, *Phys. Rev. E* **74**, 036121 (2006).  
 [13] F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, *J. Comput. Biol.* **10**, 677 (2003).  
 [14] R. Pastor-Satorras, E. Smith, and R. V. Solé, *J. Theor. Biol.* **222**, 199 (2003).  
 [15] G. Bebek, P. Berenbrink, C. Cooper, T. Friedetzky, J. H. Nadeau, and S. C. Sahinalp, *Theor. Comp. Sci.* **369**, 234 (2006).  
 [16] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).  
 [17] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).  
 [18] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, *Phys. Rev. Lett.* **94**, 188701 (2005).  
 [19] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, UK, 2003).  
 [20] J. A. J. Metz and O. Diekmann, *The Dynamics of Physiologically Structured Populations* (Springer-Verlag, Berlin, 1986).  
 [21] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **68**, 036112 (2003).  
 [22] G. Bianconi and A.-L. Barabási, *Europhys. Lett.* **54**, 436 (2001).  
 [23] V. D. P. Servedio, G. Caldarelli, and P. Buttà, *Phys. Rev. E* **70**, 056126 (2004).