

# Emergence of network structure in models of collective evolution and evolutionary dynamics

BY HENRIK JELDTOFT JENSEN<sup>1,2,\*</sup>

<sup>1</sup>*Institute for Mathematical Sciences, Imperial College London,  
53 Prince's Gate, South Kensington campus, London SW7 2PG, UK*

<sup>2</sup>*Department of Mathematics, Imperial College London,  
South Kensington campus, London SW7 2AZ, UK*

We consider an evolving network of a fixed number of nodes. The allocation of edges is a dynamical stochastic process inspired by biological reproduction dynamics, namely by deleting and duplicating existing nodes and their edges. The properties of the degree distribution in the stationary state is analysed by use of the Fokker–Planck equation. For a broad range of parameters, exponential degree distributions are observed. The mechanism responsible for this behaviour is illuminated by use of a simple mean field equation and reproduced by the Fokker–Planck equation. The latter is treated exactly, except for an approximate treatment of the degree–degree correlations. In the limit of 0 mutations, the degree distribution becomes a power law with exponent 1.

**Keywords:** networks; dynamics; evolution; degree distribution

## 1. Introduction: networks and evolutionary dynamics

Whenever a phenomenon can be thought of in terms of components and relations between components, the mathematical language of graph theory or networks may be helpful to the description, analysis and the understanding of the relevant problem of interest. A large amount of work is currently being carried out with the aim to understand the structure and statistical properties of networks in the hope that certain aspects of the general mathematical characterization of network structure may be related to common functional properties, e.g. vulnerability to the breakdown of part of the network (Biggs 1994; Albert & Barabási 2002; Newman 2003).

Our aim in the present paper is to discuss an example of a persistently evolving network of fixed size. The dynamics is driven by a Moran process (Moran 1962) inspired by evolutionary dynamics. A time step consists of two events: (i) a randomly selected node is removed with all its attached edges and (ii) a randomly selected node is selected for duplication. When a node is removed, all its edges disappear obviously as well. The duplication event involves the creation of new edges. The daughter node inherits edges with tunable probabilities. Mutations are represented in two ways. Firstly, the daughter

\*h.jensen@imperial.ac.uk

One contribution of 6 to a Special Feature ‘Stochastic networks: from theory to practice’.

inherits edges to neighbours of the parent with a probability that can be smaller than 1. Secondly, edges to nodes not connected to the parent are added to the daughter with a probability that may be larger than 0. Both these processes tend to make the daughter differ from the parent. Finally, the parent–daughter relationship suggests that an edge should be established between daughter and parent with a certain probability.

Other models in the literature have considered aspects of the network dynamics described above. For persistently growing networks, the process consisting solely of the second (duplication) step has been considered (e.g. Vázquez 2003; Ipolatov *et al.* 2005; Krapivsky & Redner 2005). One finds typical power-law degree distributions. A model considering a stochastic combination of rewiring, addition of new links and creation of new nodes was studied by Albert & Barabási (2000). They found power-law degree distributions with exponents above 2. When the probability for rewiring of edges is above a certain limit, the degree distribution becomes exponential. A model consisting of adding and removing edges to a fixed set of nodes was studied by Epstein & Wang (2002). The model generates power-law degree distributions. Power-law degree distributions with exponents above 2 were also found in other models of fixed node number in which preferential attachment is an explicit part of the dynamics (see Cheng & Tang 2004; Sarshar & Roychowdhury 2004; Salathe *et al.* 2005). Our model, consisting of a fixed number of nodes, produces typically exponential degree distributions, except in the limit of perfect inheritance where a degree distribution  $p(k) \propto 1/k$  is obtained, i.e. a power-law distribution with an exponent 1.

The paper is organized as follows. In the next section, we describe the dynamical algorithm of the model and, to develop some intuition, discuss the degree distribution by a simple mean field argument. Next, we derive the Fokker–Planck equations for the degree distribution and discuss approximations involved in these equations. In the summary and discussion section, we relate the simple node-and-edge model to emerging network structures in the individual-based tangled nature model (Anderson & Jensen 2005; Laird & Jensen 2006*b*, 2007) of evolutionary ecology and also discuss more broadly the relevance of the simple node-and-edge dynamics and the results derived.

## 2. Simple node model

Let us consider the following simple node-and-edge model of a network of  $N$  nodes and associated edges. The dynamics conserve the number of nodes. A time step consists of choosing a node at random and removing it, together with all its connected edges. Next, another node, a parent, is randomly selected from the remaining  $N-1$  nodes and is duplicated in the form of a daughter. All nodes connected to the parent are now given connections to the daughter with probability  $p_e$ . All nodes unconnected to the parent are given connections to the daughter with probability  $p_n$ . An edge between the daughter and the parent is placed with probability  $p_p$  (for a similar model with  $p_n=0$ , see Farid & Christensen 2006). These probabilities represent the degree of similarity or correlations between the daughter and the parent. The daughter will be a complete copy of the parent if  $p_e=1$  and  $p_n=0$ . It seems natural to allow for a possible ‘interaction’ between the parent and the offspring, which is represented

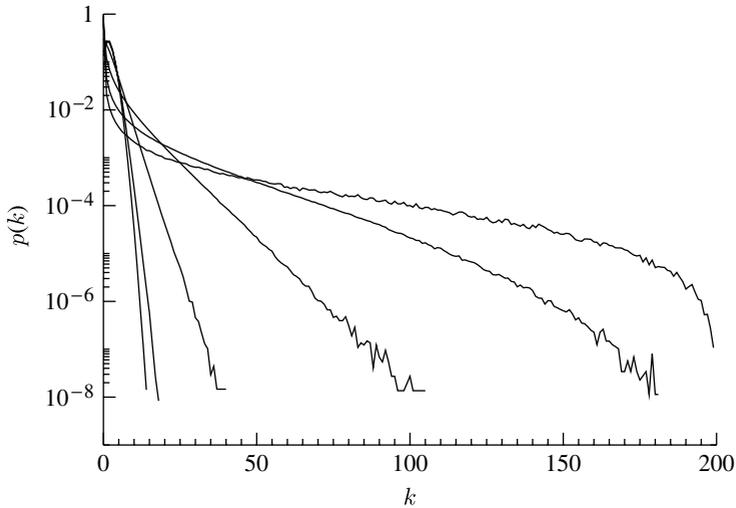


Figure 1. Simulated degree distributions of the node model for  $N = 200$  and  $p_p = 0.01$  using the imperfect duplication process. From short to long tail we have  $p_e = 0.01, 0.25, 0.75, 0.95, 0.99$  and  $0.999$  and  $p_n$  was chosen to be given by equation (2.2) in order to keep the connectance fixed (Laird & Jensen 2007).

by the possibility of establishing an edge between the parent and the offspring with probability  $p_p$ . It is straightforward to check in the mean field (Laird & Jensen 2006a) that the described edge and node dynamics converge towards a steady-state network with a time-averaged connectance

$$\langle C \rangle \equiv \frac{\text{no. of edges}}{\text{maximum no. of edges}} = \frac{p_n(N-2) + p_p}{N-1 - (p_e - p_n)(N-2)}. \tag{2.1}$$

The dynamics is simple to simulate. The results are independent of initial configuration. To make the transient very short, one may start the simulation from a binomial network of  $N$  nodes where edges between any two nodes are established with a probability equal to the mean field connectance given in equation (2.1). After a short transient, a steady state is established. The time-averaged degree distribution behaves exponentially for all values of the control parameters  $p_e$ ,  $p_n$  and  $p_p$  (see figure 1 and Laird & Jensen 2006a), except in the limit  $p_e \rightarrow 1$  and vanishing  $p_n$  and  $p_p$ , where the distribution falls off like one over the degree. Inspired by the relation between the node-and-edge model and the tangled nature model (see §4 below for details), we choose  $p_p$  equal to the connectance in equation (2.1); i.e. we link  $p_e$ ,  $p_n$  and  $p_p$  together by solving the equation  $p_p = \langle C \rangle$  and obtain

$$p_n = \frac{p_p(1 - p_e)}{1 - p_p}. \tag{2.2}$$

We will, in a moment, write down the complete Fokker–Planck equation for the degree distribution of a network evolving according to the process described above. The full equation is, however, rather involved and can only be solved by numerical iteration. It is therefore illuminating to make the following simplistic and heuristic considerations. Let  $n_k(t)$  denote the number of nodes of degree  $k$  after  $t$  time steps. Let us focus solely on the following aspects of the dynamics.

- (i) *Removal.* A node of degree  $k$  is selected for annihilation that occurs with probability  $n_k(t)/N$ . Nodes sharing edges with the removed node decrease their degree by 1. The probability that a node of degree  $k$  ends up as a degree  $k-1$  node through this process is  $k_r n_k(t)/N \mapsto \langle k \rangle n_k(t)/N$ , where the degree  $k_r$  of the removed node is replaced by the average degree.
- (ii) *Duplication.* The process of attaching edges to the new daughter node will increase the degree of existing nodes with probabilities that depend on whether these share edges with the parent node or not. A node receives an edge because it is selected to become a neighbour of the daughter (of a degree  $k_p$  parent) with probability

$$[p_e k_p + p_n(N - 1 - k_p)] \frac{n_k(t)}{N} \mapsto [p_e \langle k \rangle + p_n(N - 1 - \langle k \rangle)] \frac{n_k(t)}{N}. \tag{2.3}$$

As we are seeking a qualitative self-consistent mean field equation, we substituted the average degree  $\langle k \rangle$  in the last expression. A parent node of degree  $k$  receives an edge to the daughter, which occurs with probability  $p_p n_k(t)/N$ . It is, in general, rather complicated to estimate the probability with which the new daughter node ends up with a specific degree (see §3 below for details). However, in the limit  $p_e \rightarrow 1$ , for fixed  $p_p$  so  $p_n \rightarrow 0$  according to equation (2.2), the probability that the daughter is allocated  $k$  edges can be estimated as

$$p_p \frac{n_{k-1}}{N} + (1 - p_p) \frac{n_k}{N} + \mathcal{O}(p_n). \tag{2.4}$$

The first term corresponds to the daughter connecting to the parent and inheriting  $k-1$  edges from the parent. The second term corresponds to no edge between the parent and the daughter and  $k$  edges inherited. Here, and in the following few equations, we denote by  $\mathcal{O}(p_n)$  those terms of order  $p_n$  arising from the allocation of edges between the daughter and the nodes not connected to the parent node. We combine these events to obtain the mean field equation for the evolution of  $n_k(t)$

$$\begin{aligned} n_k(t + 1) = n_k(t) + \frac{1}{N} &[-n_k + \langle k \rangle (n_{k+1} - n_k) + (p_p + p_e \langle k \rangle \\ &+ p_n(N - 1 - \langle k \rangle))(n_{k-1} - n_k) + p_p n_{k-1} \\ &+ (1 - p_p) n_k] + \mathcal{O}(p_n). \end{aligned} \tag{2.5}$$

In the stationary limit  $n_k(t + 1) = n_k(t)$ , we obtain the following solution:

$$n_{k+1} = \alpha n_k - \beta n_{k-1}, \tag{2.6}$$

where

$$\beta = \frac{2p_p + (p_e - p_n)\langle k \rangle + p_n(N - 1)}{\langle k \rangle} + \mathcal{O}(p_n) \tag{2.7}$$

and  $\alpha = 1 + \beta$ . Using the normalization and the self-consistent equations

$$N = \sum_{k=0}^{N-1} n_k \quad \text{and} \quad \langle k \rangle = \frac{1}{N} \sum_{k=0}^{N-1} k n_k, \tag{2.8}$$

we obtain the exponential solution  $n_k = n_0 \exp(-k/k_0)$  with  $n_0 = N(1 - \exp[-1/k_0])$  and

$$k_0 = \frac{-1}{\ln \left[ 1 - \frac{1+p_n-p_e}{2p_p+p_n(N-1)+1} + \mathcal{O}(p_n) \right]} \simeq \frac{-1}{\ln \left( \frac{p_e+2p_p}{1+2p_p} \right)}, \tag{2.9}$$

where the approximation refers to the limit  $p_n \rightarrow 0$  for fixed  $N$ . The divergence of the exponential cut-off in the limit  $p_e \rightarrow 1$  and  $p_p \rightarrow 0$  is in qualitative agreement with the change in the exponential part of the degree distribution obtained in simulations of the network (figure 1). However, this simplistic mean field discussion is only of heuristic value. We now present the full Fokker–Planck-like equation for the process.

### 3. Fokker–Planck equation

Some care has to be taken when we develop the Fokker–Planck equation for the ensemble-averaged time-dependent number of nodes of degree  $k$ ,  $n_k(t)$ , constrained by the condition  $\sum_k n_k(t) = N$ . Firstly, it is worth mentioning that the equations are concerned with the ensemble-averaged quantity  $n_k(t)$  and accordingly neglect ‘microscopic’ fluctuations in the number of nodes of degree. Secondly, to make a closed set of equations, one needs to perform some kind of truncation of a hierarchy of equations, which couples the degree distribution to the degree–degree correlation functions, which, in turn, is coupled to triple correlation functions, etc. This is usually the case. We will first write down the equations formally, including the needed degree–degree correlation functions, and then make clear the nature of the heuristic approximation we have used to estimate this correlation function.

We structure the analysis in the following way. Removal (R): the effect of removing, from a population of  $N$ , a node and its edges described by a rate term  $\Gamma_R(N, k, t)$ . Duplication (D): the effect of introducing, into a population of  $N-1$ , a new daughter node and attaching edges described by a rate term  $\Gamma_{Du}(N-1, k, t)$ . Our equation has accordingly the form

$$n_k(t+1) = n_k(t) + \Gamma_R(N, k, t) + \Gamma_{Du}(N-1, k, t). \tag{3.1}$$

Removal of a node affects the network in two ways:  $\Gamma_R^r$ , the node being removed from the network and  $\Gamma_R^a$ , the effect on the nodes being adjacent, i.e. sharing edges, to the node being removed. Therefore,

$$\Gamma_R(k) = -\Gamma_R^r(k) + \Gamma_R^a(k+1) - \Gamma_R^a(k). \tag{3.2}$$

The effect of the duplication process is conveniently broken up into three sub-effects:  $\Gamma_{Du}^p$ , the effect on the parent;  $\Gamma_{Du}^d$ , the effect on the daughter; and  $\Gamma_{Du}^a$ , the effect on the adjacent nodes, i.e. those that will receive an extra edge as a result of the duplication. Hence,

$$\Gamma_{Du}(k) = \Gamma_{Du}^p(k-1) - \Gamma_{Du}^p(k) + \Gamma_{Du}^d(k) + \Gamma_{Du}^a(k-1) - \Gamma_{Du}^a(k). \tag{3.3}$$

We have suppressed the time step,  $t$ , and network size,  $N$ , for notational ease.

Next, we derive detailed expressions for each of the terms above. The direct effect on  $n_k$  of removing a node of degree  $k$  is to decrement  $n_k$ . The probability of selecting a node of degree  $k$  is  $n_k/N$  and therefore,

$$\Gamma_R^r(k) = \frac{n_k}{N}. \tag{3.4}$$

After the removal, the degree of the nodes connected to the removed node, i.e. the adjacent nodes, will decrease by 1. For this we need the *Edge* probability,

$$p_{Ed}(k_1, k_2, q) = \text{Prob}\{\text{node of degree } k_1 \text{ is connected to } q \text{ nodes of degree } k_2\}. \tag{3.5}$$

In general we do not have a closed analytic expression for  $p_{Ed}(k_1, k_2, q)$ , but below we give approximate forms neglecting, or treating non-rigorously, degree-degree correlations. Here, we note

$$\Gamma_R^a(k) = \sum_{k_r=1}^{N-1} \frac{n_{k_r}}{N} \sum_{q=1}^{k_r} q p_{Ed}(k_r, k, q). \tag{3.6}$$

The first sum is over the degree of the removed node and the second sum is over the number,  $q$ , of nodes of degree  $k = 0, 1, \dots, N - 1$  the removed node is connected to.

A node of degree  $k$  is selected for duplication with probability  $n_k/(N - 1)$ . The daughter of this node receives an edge to the parent with probability  $p_p$ . Thus, the parent increases its degree by 1 with probability

$$\Gamma_{Du}^p(k) = p_p \frac{n_k}{N - 1}. \tag{3.7}$$

The new daughter node can add to  $n_k$  by an amount determined by the probability of finishing with  $k$  edges,

$$\Gamma_{Du}^d(k) = p_p A(k - 1) + (1 - p_p) A(k). \tag{3.8}$$

To keep track of the bookkeeping, we have introduced a new probability

$$A(k) = \text{Prob}\{\text{daughter receives } k \text{ edges to nodes different from the parent}\}, \tag{3.9}$$

which is given by

$$A(k) = \sum_{k_p=0}^{N-2} \sum_{k_1=0}^{\min[k_p, k]} \sum_{k_2=0}^{\min[N-2-k_p, k]} \frac{n_{k_p}}{N - 1} \delta(k_1 + k_2 - k) \Omega(k_1, k_2, k_p). \tag{3.10}$$

The r.h.s. adds up the probabilities associated with the process where the daughter inherits  $k_1$  edges to nodes already connected to the parent. Each of these edges is inherited by the daughter with probability  $p_e$ . The daughter may receive an additional  $k_2 = k - k_1$  edges to nodes not connected to the parent. Each of these edges are attached to the daughter with probability  $p_n$ . The factor  $\Omega(k_1, k_2, k_p)$  denotes the probability of  $k = k_1 + k_2$  edges allocated to the daughter,  $k_1$  of the edges are inherited, i.e. these edges connect to some of

the  $k_p$  nodes connected to the parent. In addition, the daughter receives  $k_2$  edges, which connect to nodes not connected to the parent. The probability for this event is

$$\Omega(k_1, k_2, k_p) = \binom{k_p}{k_1} p_e^{k_1} (1 - p_e)^{k_p - k_1} \binom{N - 2 - k_p}{k_2} p_n^{k_2} (1 - p_n)^{N - 2 - k_p - k_2}. \tag{3.11}$$

Next, we consider the effect of the duplication on the adjacent nodes, and we need to distinguish between the nodes sharing an edge with the parent (Ed) and nodes not connected to the parent (nEd). Let us first consider the Ed nodes. We introduced previously  $p_{Ed}(k_p, k, q_E)$  as the probability that a node, here the parent, of degree  $k_p$  is connected to  $q_E$  nodes of degree  $k$ . The duplication process will, with probability  $p_e$ , attach a new edge from the daughter to each of these nodes and thereby increase their degree from  $k$  to  $k + 1$ . Let us now turn to the nEd nodes. There are  $N - 2 - k_p$  nodes that share no edge with the parent. With probability  $p_n p_{nEd}(N - 2 - k_p, k, q_{nE})$ , a total of  $q_{nE}$  of these nodes are of degree  $k$  and will receive a new edge to the daughter. Here,  $p_{nEd}(N - 2 - k_p, k, q_{nE})$  is equivalent to  $p_{Ed}(N, k, q)$  introduced in equation (3.5), though  $p_{nEd}(k_p, k, q_{nE})$  is concerned with the  $N - 2 - k_p$  nodes that a node of degree  $k_p$  (in a set of  $N - 1$  nodes) is *not* connected to. Among these  $N - 2 - k_p$  nodes,  $q_{nE}$  have degree  $k$  with probability  $p_{nEd}(N - 2 - k_p, k, q_{nE})$ . Therefore, we have

$$\begin{aligned} \Gamma_{Du}^a(k) &= \sum_{q=0}^{N-2} \sum_{k_p=0}^{N-2} \sum_{k_1=0}^{k_p} \sum_{k_2=0}^{N-2-k_p} \sum_{q_1=0}^{k_1} \sum_{q_2=0}^{k_2} \delta(q_1 + q_2 - q) q \frac{n_{k_p}}{N - 1} p_{Ed}(k_p, k, k_1) \\ &\quad \times \binom{k_1}{q_1} p_e^{q_1} (1 - p_e)^{k_1 - q_1} p_{nEd}(k_p, k, k_2) \binom{k_2}{q_2} p_n^{q_2} (1 - p_n)^{k_2 - q_2}. \end{aligned} \tag{3.12}$$

Degree-degree correlations induced by the evolutionary dynamics makes it difficult to write an explicit form for  $p_{Ed}(k_1, k_2, q)$  and  $p_{nEd}(k_1, k_2, q)$ . Numerical simulations show that the network is disassortative (Newman 2003), i.e. nodes with a high degree tend to attach to nodes with a low degree, and that the Pearson correlation coefficient decreases rapidly with size of the network and increased connectance. This numerical finding suggests that it makes sense to analytically treat the degree correlations approximately. One can choose to neglect the correlations altogether and try to estimate  $p_{Ed}$  and  $p_{nEd}$  by purely binomial arguments in the following way. First, we deal with  $p_{Ed}(k_1, k_2, q)$ . The  $k_1$  edges emerging from the degree  $k_1$  node connects (in this approximation) to nodes of degree  $k_2$  with probability  $(n_{k_2} - \delta_{k_1, k_2}) / (N - 1)$  (remember there are  $N - 1$  nodes when the duplication takes place), hence, we use

$$P_{Ed}(k_1, k_2, q) = \binom{k_1}{q} \left( \frac{n_{k_2} - \delta_{k_1, k_2}}{N - 1} \right)^q \left( 1 - \frac{n_{k_2} - \delta_{k_1, k_2}}{N - 1} \right)^{k_1 - q}. \tag{3.13}$$

When we treat  $p_{nEd}(k_1, k_2, q)$  in the same approximation, we obtain  $p_{nEd}(k_1, k_2, q) = p_{Ed}(N - 2 - k_1, k_2, q)$  since we now pick  $q$  nodes among the  $N - 2 - k_1$  nodes not connected to the degree  $k_1$  node under consideration. It appears to be better to treat the correlations by a somewhat different

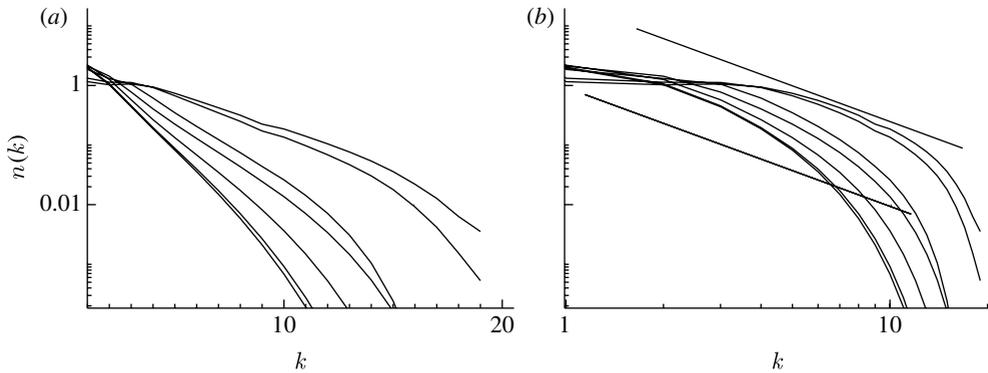


Figure 2. The degree distribution obtained by iteration of the Fokker–Planck equation (3.1). (a) The exponential form for a broad range of parameter values in a linear–log plot. (b) The approach towards a  $1/k$  dependence in the limit of  $p_e \rightarrow 1$  in a log–log plot. The two straight lines have slope  $-1$ . The parameters are  $N = 20$ ,  $p_e = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95$  and  $p_p = 0.01$ .  $p_n$  was chosen to be given by equation (2.2).

argument that focuses on the edge dynamics. This approach leads to better numerical convergence towards the results obtained by direct simulation (figures 1 and 2). We use the following urn argument. We place  $M = \sum_k n_k$  edges in an urn. The edges are of two types. Type  $\mathcal{A}$  edges correspond to the  $|\mathcal{A}| = k_2(n_{k_2} - \delta_{k_1, k_2})$  edges connecting nodes of degree  $k_2$ . In addition, we have  $|\mathcal{B}| = M - |\mathcal{A}|$  type  $\mathcal{B}$  edges connecting nodes of degree different from  $k_2$ . The probability that among  $k_1$  randomly picked edges  $q$  are of type  $\mathcal{A}$  and  $k_1 - q$  are of type  $\mathcal{B}$  is given by

$$p_{\text{Ed}}(k_1, k_2, q) = \binom{k_1}{q} \left(\frac{k_2 n_{k_2}}{M}\right)^q \left(1 - \frac{k_2 n_{k_2}}{M}\right)^{k_1 - q}. \tag{3.14}$$

Again, we assume  $p_{\text{nEd}}(k_1, k_2, q) = p_{\text{Ed}}(N - 2 - k_1, k_2, q)$ . In general, it is not simple to find analytic solutions to this somewhat involved set of equations. The result of iterating the Fokker–Planck equation (3.1) using these estimates is shown in figure 2 for diversity  $N = 20$ , which makes the numerical iteration manageable. We note good qualitative agreement with the behaviour of simulation results presented in figure 1.

Let us finally mention that direct simulations (Laird & Jensen 2006a) of the simple model described in §2 show that in the limit  $p_e \rightarrow 1$ , for fixed  $p_p \ll 1$ , so  $p_n \rightarrow 0$  according to equation (2.2), the degree distribution  $n_k$  behaves like  $1/k$ . The Fokker–Planck equation (3.1) confirms this result. In this limit, the Fokker–Planck equation reduces to

$$\begin{aligned} n_k(t + 1) = & n_k(t) + n_k \left( \frac{1}{N - 1} - \frac{1}{N} \right) + \frac{2p_p}{N - 1} (n_{k-1} - n_k) \\ & + \sum_{k_1=1}^{N-2} \sum_{q=1}^{k_1} q n_{k_1} \left[ \frac{1}{N} p_{\text{Ed}}(k_1, k + 1, q) - \frac{1}{N - 1} p_{\text{Ed}}(k_1, k, q) \right. \\ & \left. - \frac{1}{N} p_{\text{Ed}}(k_1, k, q) + \frac{1}{N - 1} p_{\text{Ed}}(k_1, k - 1, q) \right]. \end{aligned} \tag{3.15}$$

Including only the leading terms from  $k_1 = 1$  and  $q = 1$ , one obtains

$$n_k(t + 1) = n_k + \frac{n_k}{N(N - 1)} + \frac{2p_p}{N - 1} [n_{k-1} - n_k] + \frac{n_1}{M} \left[ \frac{1}{N} \{ (k + 1)n_{k+1} - kn_k \} - \frac{1}{N - 1} \{ kn_k - (k - 1)n_{k-1} \} \right]. \quad (3.16)$$

In the limit  $N \rightarrow \infty$  and  $p_p \ll 1$ , this equation has the stationary solution  $n_k \propto 1/k$ . For detailed numerical study of the  $1/k$  behaviour, see Laird & Jensen (2006a).

#### 4. Summary and discussion

Let us now briefly address the relevance of the simplistic network model discussed above. The inspiration to the model came from a study of emergent networks in the individual-based tangled nature (Christensen *et al.* 2002; Laird & Jensen 2006b, 2007). The basics of the tangled nature model (Christensen *et al.* 2002; Laird & Jensen 2006b, 2007) are as follows. Individuals  $\{\alpha, \beta, \dots\}$  are described by type vectors  $\mathbf{T}^\alpha = (T_1^\alpha, T_2^\alpha, \dots, T_L^\alpha)$ . The number of individuals of type  $\mathbf{T}$  at time  $t$  is denoted by  $n(\mathbf{T}, t)$ . Different types influence each other through an interaction matrix ( $\mathbf{J}$ -matrix) that accounts for all possible interactions between any possible set of types. Once the matrix  $\mathbf{J}$  is defined it never changes. The dynamics consists of the configuration of occupied types changing around in the fixed space given by the positions  $\mathbf{T}$  and the coupling matrix  $\mathbf{J}$ . Selection leads to only a small fraction of types being occupied, and their interactions will be described by a small subset of the elements of this complete matrix. Species are defined as emergent structures in the type space in the following way. At time  $t$  the local maxima,  $\mathbf{T}_{\max}$ , of the occupancy  $n(\mathbf{T}, t)$  are identified. All occupied types within a distance from a given  $\mathbf{T}_{\max}$  smaller than the correlation length of the matrix  $\mathbf{J}$  are considered to be belonging to the species defined by  $\mathbf{T}_{\max}$ .

The structure of the interaction network between extant species is found to depend on the statistical properties of  $\mathbf{J}$ . A proportion,  $\theta$ , of the elements of the  $\mathbf{J}$ -matrix,  $\mathbf{J}(\mathbf{T}^\alpha, \mathbf{T}^\beta)$  is assigned non-zero (and non-symmetric) values; all other elements are zero. The interactions assigned in the type space can either be uncorrelated (Hall *et al.* 2002; Anderson & Jensen 2005) or correlated (Laird & Jensen 2006b, 2007). If no correlations are present in the type space, the evolved networks of interactions between extant species exhibit a binomial degree distribution as does the underlying network of non-zero  $\mathbf{J}$ -matrix elements (Anderson & Jensen 2005). The correlated case is more interesting and is relevant to this paper. Correlations are made to decay exponentially with the separation in type space. This implies that offspring will see a set of interactions that are very similar to the interactions of the parent, even when mutations make the offspring differ slightly from the parent. When correlations are present in  $\mathbf{J}$ , the evolutionary dynamics is able to generate a network of interactions, between extant species, described by an exponential degree distribution. This is very different from the binomial distributions exhibited by a network constructed by randomly selecting positions in the same type space. This is interesting since

Stumpf & Wiuf (2005) have studied the properties of sub-networks obtained by random sampling nodes in a larger network. They showed that the binomial (or Poisson) networks are invariant under decimation. If the large network has a binomial degree distribution, a randomly sampled subset of the network will also exhibit a binomial degree distribution. The results derived by Stumpf and Wiuf are concerned with the statistical properties of sub-networks obtained by *random* sampling. What we have found is that evolutionary dynamics is able to generate sub-networks typically characterized by the exponential degree distributions, even when the full network has a binomial degree distribution.

The simple node-and-edge model discussed in §2 can be seen as the explanation of how the evolutionary dynamics is able to produce sub-networks with a degree distribution of a functional form totally different from the one describing the entire network, from which they are sampled. A qualitative link between the tangled nature model and the simple node-and-edge model can readily be established. To understand this phenomenology of the tangled nature model, we now neglect the fluctuations present at the level of individual-based dynamics and assume a more coarse-grained viewpoint in which we consider species as either occupied or not. That is, we turn the coarse-grained measure  $n(\mathbf{T}, t)$  into a binary equal to 1 when  $n(\mathbf{T}, t) > 0$  and 0 when  $n(\mathbf{T}, t) = 0$ . We consider the dynamics at the level of species, which implies that creation events correspond to one species splitting into two species (a speciation event) and annihilation events correspond to a species going extinct. We elevate the dynamics of the individual-based tangled nature model to the level of species and describe this high-level dynamics by the node-and-edge model. The more correlated the interaction matrix  $\mathbf{J}$  of the tangled nature model is, the more similar will the offspring species be to the parent, implying that the edge probabilities  $p_e$  should be large and  $p_n$  small, respectively. If the connectance of the interaction matrix  $\mathbf{J}$  is large it is probable that the offspring will end up with a link to the parent species, i.e. the probability  $p_p$  should be high. An exact link between the set of probabilities ( $p_e$ ,  $p_n$  and  $p_p$ ) of the node-and-edge model and the parameters defining the tangled nature model is not possible, nor is it needed, since the obtained results are robust for a broad range of control parameters.

The exponential degree distributions found in the tangled nature model and node-and-edge model may also be of relevance to naturally occurring food webs (see Dunne *et al.* 2002) and of interest to protein–protein interaction networks (Mering *et al.* 2002; Ipolatov *et al.* 2005).

I am grateful for very fruitful collaboration with Dr Simon Laird.

## References

- Albert, R. & Barabási, A.-L. 2000 Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* **85**, 5234–5237. (doi:10.1103/PhysRevLett.85.5234)
- Albert, R. & Barabási, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–57. (doi:10.1103/RevModPhys.74.47)
- Anderson, P. & Jensen, H. J. 2005 Network properties, species abundance and evolution in a model of evolutionary ecology. *J. Theor. Biol.* **232**, 551–558. (doi:10.1016/j.jtbi.2004.03.029)
- Biggs, N. L. 1994 *Algebraic graph theory*. Cambridge, UK: Cambridge University Press.
- Cheng, K. & Tang, Y. 2004 Growing networks based on the mechanism of addition and deletion. *Chin. Phys. Lett.* **21**, 1858–1860. (doi:10.1088/0256-307X/21/9/052)

- Christensen, K., Collobiano, S. A., di Hall, M. & Jensen, H. J. 2002 Tangled nature: a model of evolutionary ecology. *J. Theor. Biol.* **216**, 73–84. (doi:10.1006/jtbi.2002.2530)
- Dunne, J. A., Williams, R. J. & Martinez, N. D. 2002 Food web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12 917–12 922. (doi:10.1073/pnas.192407699)
- Epstein, D. & Wang, J. 2002 A steady state model for graph power law. In *Second Int. Workshop on Web Dynamics*, 5234.
- Farid, N. & Christensen, K. 2006 Evolving networks through deletion and duplication. *New J. Phys.* **8**, 212. (doi:10.1088/1367-2630/8/9/212)
- Hall, M., Christensen, K., di Collobiano, S. A. & Jensen, H. J. 2002 Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E* **66**, 1–10. (doi:10.1103/PhysRevE.66.011904)
- Ipolatov, I., Krapivsky, P. & Yuryev, A. 2005 Duplication-divergence model of protein interaction network. *Phys. Rev. E* **71**, 061 911. (doi:10.1103/PhysRevE.71.061911)
- Krapivsky, P. & Redner, S. 2005 Network growth by coping. *Phys. Rev. E* **71**, 036 118. (doi:10.1103/PhysRevE.71.036118)
- Laird, S. & Jensen, H. J. 2006*a* A steady state network model with a  $1/k$  scale-free degree distribution. *Europhys. Lett.* **76**, 710–716. (doi:10.1209/epl/i2006-10319-x)
- Laird, S. & Jensen, H. J. 2006*b* The tangled nature model with inheritance and constraint: evolutionary ecology restricted by a conserved resource. *Ecol. Complexity* **3**, 253–262. (doi:10.1016/j.ecocom.2006.06.001)
- Laird, S. & Jensen, H. J. 2007 Correlation, selection and the evolution of species networks. *Ecol. Modell.* **209**, 149–156. (doi:10.1016/j.ecolmodel.2007.06.016)
- Mering, C., von Krause, R., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. 2002 Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403. (doi:10.1038/nature750)
- Moran, P. 1962 *The statistical processes of evolutionary theory*. Oxford, UK: Clarendon Press.
- Newman, M. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)
- Newman, M. E. J. 2003 Mixing patterns in networks. *Phys. Rev. E* **67**, 026 126. (doi:10.1103/PhysRevE.67.026126)
- Salathe, M., May, R. & Bonhoeffer, S. 2005 The evolution of network topology by selective removal. *J. R. Soc. Interface* **2**, 533–536. (doi:10.1098/rsif.2005.0072)
- Sarshar, N. & Roychowdhury, V. 2004 Scale-free and stable structures in complex ad hoc networks. *Phys. Rev. E* **69**, 026 101. (doi:10.1103/PhysRevE.69.026101)
- Stumpf, M. & Wiuf, C. 2005 Sampling properties of random graphs: the degree distribution. *Phys. Rev. E* **72**, 036 118. (doi:10.1103/PhysRevE.72.036118)
- Vázquez, A. 2003 Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* **67**, 056 104. (doi:10.1103/PhysRevE.67.056104)