Emergence of Network Structure in Models of Collective Evolution and Evolutionary Dynamics.

Henrik Jeldtoft Jensen[*†‡]

January 25, 2008

Abstract

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 zero mutations the degree distribution becomes a power law with exponent one.

Keywords: Networks, Dynamics, Evolution, Degree distribution

1 Introduction: Networks and evolutionary dynamics

Whenever a phenomena 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 done with the aim to understand the structure and statistical properties of networks in the hope that certain aspects of the general mathematical characterisation of network structure may be related to common functional

---

[*]Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, South Kensington campus, SW7 2PG, UK.

[†]Department of Mathematics, Imperial College London, South Kensington campus, London SW7 2AZ, UK.

[‡]Email and URL (h.jensen@imperial.ac.uk; www.ma.ic.ac.uk/~hjjens)

properties, e.g. vulnerability to breakdown of part of the network (Biggs, 1994; Albert & Barabasi, 2002; M. 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. 1) a randomly selected node is removed with all its attached edges and 2) 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 inherits edges to neighbours of the parent with a probability that can be smaller than one. Secondly, edges to nodes not connected to the parent are added to the daughter with a probability that may be larger than zero. Both these processes tends 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. in (Vázquez, 2003; Ipolatov, Krapivsky, & Yuryev, 2005; Krapivsky & Redner, 2005). One finds typically power law degree distributions. A model considering a stochastic combination of rewiring, addition of new links and creation of new nodes were studied by Albert and Barabási in (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 and Wang(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 e.g. (Sarshar & Roychowdhury, 2004; Cheng & Tang, 2004; Salathe, May, & Bonhoeffer, 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. power law distribution with an exponent one.

The paper is organised 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 discussion and summary 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, 2006b, 2007) of evolutionary ecology and we 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 conserves the number of nodes. A timestep consists of choosing a node at random and remove 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 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 daughter and 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 possible "interaction" between parent and offspring, which is represented by the possibility of establishing an edge between parent and offspring with probability p_p. It it is straight forward to check in mean field(Laird & Jensen, 2006a) that the described edge and node dynamics converges towards a steady state network with a time averaged connectance

$$\langle C \rangle \equiv \frac{\text{number of edges}}{\text{maximum number of edges}} = \frac{P_n(N-2) + P_p}{N - 1 - (P_e - P_n)(N-2)}. \tag{1}$$

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 Eq. (1). After a short transient a steady state is established. The time averaged degree distribution behave exponentially for all values of the control parameters P_e, P_n and P_p (see Fig. 1 below and (Laird & Jensen, 2006a)) except in the limit $P_e \rightarrow 1$ and vanishing P_n and P_p where the distribution falls of like one over the degree. Inspired by the relation between the node-and-edge model and the Tangled Nature model (see Sec. 4 below for details) we choose P_p equal to the connectance in Eq. (1). I.e. we link P_e, P_n and P_p together by solving the equation $P_p = \langle C \rangle$ and obtain

$$P_n = \frac{P_p(1 - P_e)}{1 - P_p} \tag{2}$$

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 timesteps. Let us focus solely on the following aspects of the dynamics:

1) *Removal*: a node of degree k is selected for annihilation, this occurs with probability $n_k(t)/N$. Nodes sharing edges with the removed node decreases their degree by one. 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.

2) *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

$$(P_e k_p + P_n(N - 1 - k_p))]\frac{n_k(t)}{N}] \mapsto [P_e\langle k \rangle + P_n(N - 1 - \langle k \rangle)]\frac{n_k(t)}{N}. \tag{3}$$

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, this 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 Sec. 3 below for details). However in the limit $P_e \to 1$ and $P_n \to 0$ the probability that the daughter is allocated k edges can be estimated as

$$P_p \frac{n_{k-1}}{N} + (1 - P_p)\frac{n_k}{N} + \mathcal{O}(P_n). \tag{4}$$

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 parent and 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 nodes not connected to the parent node. We combine these events to obtain the mean field equation for the evolution of $n_k(t)$

$$
\begin{aligned}
n_k(t+1) &= n_k(t) + \frac{1}{N}[-n_k + \langle k \rangle(n_{k+1} - n_k) \\
&+ (P_p + P_e\langle k \rangle + P_n(N - 1 - \langle k \rangle))(n_{k-1} - n_k) + P_p n_{k-1} + (1 - P_p)n_k] \\
&+ \mathcal{O}(P_n).
\end{aligned} \tag{5}
$$

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

$$n_{k+1} = \alpha n_k - \beta n_{k-1}, \tag{6}$$

where

$$\beta = \frac{2P_p + (P_e - P_n)\langle k \rangle + P_n(N - 1)}{\langle k \rangle} + \mathcal{O}(P_n) \tag{7}$$

and $\alpha = 1 + \beta$. Using the normalisation and the self-consistent equation

$$N = \sum_{k=0}^{N-1} n_k \quad \text{and} \quad \langle k \rangle = \frac{1}{N}\sum_{k=0}^{N-1} k n_k, \tag{8}$$

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

$$k_0 = -1/\ln[1 - \frac{1 + P_n - P_e}{2P_p + P_n(N - 1) + 1} + \mathcal{O}(P_n)] \simeq -1/\ln(\frac{P_e + 2P_p}{1 + 2P_p}), \tag{9}$$

where the approximation refers to $N \gg 1$ and the limit $P_n \to 0$. The divergence of the exponential cutoff in he limit $P_e \to 1$ and $P_p \to 0$ is in qualitative agreement with the change in the exponential part of the degree distribution obtained in simulations of the network, see Fig. 1. However, this simplistic mean field discussion is only of heuristic value. We now present the full Fokker-Planck like equation for the process.
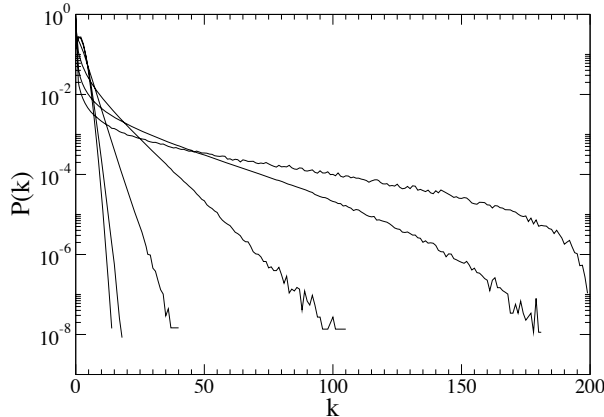
Figure 1: Simulated degree distributions of the node model for $N = 200$ and $P_p = 0.01$ using the imperfect duplication process. From short to long tail we have $P_e = 0.01$, 0.25, 0.75, 0.95, 0.99, 0.999 and P_n was chosen to be given by Eq. (2) in order to keep the connectance fixed (Laird & Jensen, 2007).

3 Fokker-Planck equation

Some care has to be observed 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 to mention 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 function, 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 function 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.

$$n_k(t + 1) = n_k(t) + \Gamma_R(N, k, t) + \Gamma_{Du}(N - 1, k, t). \tag{10}$$

Removal of a node affects the network in two ways: (Γ_R^r) the node removed from the network and (Γ_R^a) the effect on the nodes adjacent, i.e. sharing edges, with the node being removed. Therefore

$$\Gamma_R(k) = -\Gamma_R^r(k) + \Gamma_R^a(k+1) - \Gamma_R^a(k). \tag{11}$$

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

5

Hence
$$\Gamma_{Du}(k) = \Gamma^p_{Du}(k-1) - \Gamma^p_{Du}(k) + \Gamma^d_{Du}(k) + \Gamma^a_{Du}(k-1) - \Gamma^a_{Du}(k). \qquad (12)$$

We have suppressed the timestep, 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,

$$\Gamma^r_R(k) = \frac{n_k}{N}. \qquad (13)$$

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

$$P_{Ed}(k_1, k_2, q) = \mathrm{Prob}\{\text{node of degree } k_1 \text{ is connected to } q \text{ nodes of degree } k_2\}. \qquad (14)$$

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

$$\Gamma^a_R(k) = \sum_{k_r=1}^{N-1} \frac{n_{k_r}}{N} \sum_{q=1}^{k_r} q P_{Ed}(k_r, k, q). \qquad (15)$$

The first sum is over the degree of the removed node, the second sum is over the number, q, of nodes of degree $k = 0, 1, ..., N-1$ 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 one with probability

$$\Gamma^p_{Du}(k) = P_p \frac{n_k}{N-1}. \qquad (16)$$

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

$$\Gamma^d_{Du}(k) = P_p \Lambda(k-1) + (1 - P_p)\Lambda(k). \qquad (17)$$

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

$$\Lambda(k) = \mathrm{Prob}\{\text{daughter receives } k \text{ edges to nodes different from the parent}\}, \qquad (18)$$

which is given by

$$\Lambda(k) = \sum_{k_p=0}^{N-2} \sum_{k_1=0}^{\min[k_p,k]} \sum_{k_2=0}^{\min[N-2-k_p,k]} \frac{n_{k_p}}{N-1} \delta(k_1+k_2-k)\Omega(k_1, k_2, k_p), \qquad (19)$$

The right hand side 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 that

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

$$\Omega(k_1, k_2, k_p) = \begin{pmatrix} k_p \\ k_1 \end{pmatrix} P_e^{k_1}(1-P_e)^{k_p-k_1} \begin{pmatrix} N-2-k_p \\ k_2 \end{pmatrix} P_n^{k_2}(1-P_n)^{N-2-k_p-k_2}. \qquad (20)$$

Next we consider the effect of the duplication on the adjacent nodes and we need to distinguish between nodes sharing an edge with the parent (Ed) and nodes not connect to the parent (nEd). Let us first consider the Ed nodes. We introduced above $P_{Ed}(k_p, k, q_E)$ as the probability that a mode, 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 which 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 edged to the daughter. Here $P_{nEd}(N - 2 - k_p, k, q_{nE})$ is equivalent to $P_{Ed}(N, k, q)$ introduced in Eq. (14), 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,

$$\begin{aligned} \Gamma_{Du}^a(k) &= \sum_{q=0}^{N-2} \sum_{k_p=0}^{N-2} \sum_{\kappa_1=0}^{k_p} \sum_{\kappa_2=0}^{N-2-k_p} \sum_{q_1=0}^{\kappa_1} \sum_{q_2=0}^{\kappa_2} \delta(q_1+q_2-q) \\ &\quad q\frac{n_{k_p}}{N-1} P_{Ed}(k_p, k, \kappa_1) \begin{pmatrix} \kappa_1 \\ q_1 \end{pmatrix} P_e^{q_1}(1-P_e)^{\kappa_1-q_1} \\ &\quad P_{nEd}(k_p, k, \kappa_2) \begin{pmatrix} \kappa_2 \\ q_2 \end{pmatrix} P_n^{q_2}(1-P_n)^{\kappa_2-q_2}. \end{aligned}$$

(21)

Degree-degree correlations induced by the evolutionary dynamics makes is difficult to write an explicit form for $P_{Ed}(k_1, k_2, q)$ and $P_{nEd}(k_1, k_2, q)$. The numerical simulation show that the network is disassortative(M. E. J. Newman, 2003), i.e. nodes with high degree tend to attach to nodes with low degree, and that the Pearson correlation coefficient(M. E. J. Newman, 2003) decreases rapidly with size of the network and with 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

$$P_{Ed}(k_1, k_2, q) = \begin{pmatrix} k_1 \\ q \end{pmatrix} \left(\frac{n_{k_2} - \delta_{k_1,k_2}}{N-1}\right)^q \left(1 - \frac{n_{k_2} - \delta_{k_1,k_2}}{N-1}\right)^{k_1-q}. \qquad (22)$$

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 argument, which focus on the edge dynamics. This approach leads to better numerically convergence towards the results obtained by direct simulation (See Fig. 1 and Fig. 2). We use the 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 to nodes of degree k_2. In addition we have $|\mathcal{B}| = M - |A|$ 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

$$P_{Ed}(k_1, k_2, q) = \begin{pmatrix} k_1 \\ q \end{pmatrix} \left(\frac{k_2 n_{k_2}}{M}\right)^q \left(1 - \frac{k_2 n_{k_2}}{M}\right)^{k_1-q}. \tag{23}$$

Again we assume $P_{nEd}(k_1, k_2, q) = P_{Ed}(N - 2 - k_1, k_2, q)$. In general it is not simple to find analytic solutions to this somewhat involved set of equations. The result of iterating the Fokker-Planck Eq. (10) using these estimates is shown in Fig. 2 for diversity $N = 20$, which makes the numerical iteration manageable. We notice good qualitative agreement with the behaviour of simulation results presented Fig. 1.
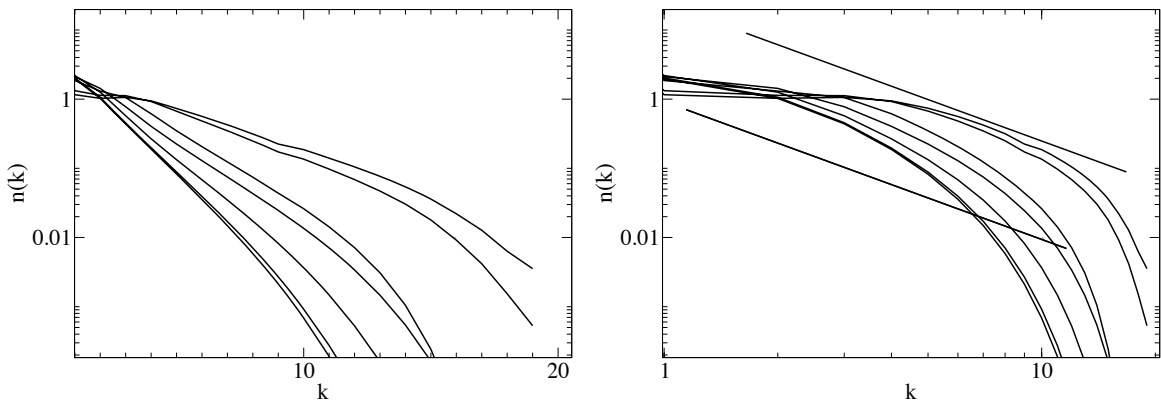

Figure 2: The degree distribution obtained by iteration of the Fokker-Planck equation (10). The exponential form is visible for a broad range of parameter values in the linear-log plot to the left. The approach towards a $1/k$ dependence in the limit of $P_e \to 1$ can be seen in the log-log plot to the right. 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 Eq. (2).

Let us finally mention that direct simulations (Laird & Jensen, 2006a) of the simple model described in Sec. 2 show that in the limit $P_e \to 1$, $P_n \to 0$ and $P_p \ll 1$ the degree distribution n_k behaves like $1/k$. The Fokker-Planck equation Eq. (10) confirms this result. In the limit $P_e = 1$ and $P_n \to 0$ (i.e. perfect replication) the Fokker-Planck

equation reduces to

$$
\begin{aligned}
n_k(t+1) &= n_k(t) + n_k(\frac{1}{N-1} - \frac{1}{N}) + \frac{2P_p}{N-1}(n_{k-1} - n_k) \\
&+ \sum_{k_1=1}^{N-2} \sum_{q=1}^{k_1} q n_{k1}[\frac{1}{N} P_{Ed}(k_1, k+1, q) - \frac{1}{N-1} P_{Ed}(k_1, k, q) \\
&- \frac{1}{N} P_{Ed}(k_1, k, q) + \frac{1}{N-1} P_{Ed}(k_1, k-1, q)].
\end{aligned} \tag{24}
$$

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

$$
\begin{aligned}
n_k(t+1) &= n_k + \frac{n_k}{N(N-1)} + \frac{2P_p}{N-1}[n_{k-1} - n_k] \\
&+ \frac{n_1}{M}[\frac{1}{N}\{(k+1)n_{k+1} - kn_k\} - \frac{1}{N-1}\{kn_k - (k-1)n_{k-1}\}].
\end{aligned} \tag{25}
$$

In the limit $N \gg 1$ 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, Collobiano, Hall, & Jensen, 2002; Laird & Jensen, 2006b, 2007). The basic of the Tangled Nature model (Christensen et al., 2002; Laird & Jensen, 2006b, 2007) is as follows. Individuals, $\{\alpha, \beta, ...\}$ are described by type vectors $\mathbf{T}^\alpha = (T_1^\alpha, T_2^\alpha, ..., T_L^\alpha)$. 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 belong 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, θ, of the elements of the $\mathbf{J}$-matrix, $\mathbf{J}(\mathbf{T}^\alpha, \mathbf{T}^\beta)$ are assigned non-zero (and non symmetric) values all other elements are zero. The interactions assigned in the type space can either be uncorrelated (Hall, Christensen, Collobiano, & Jensen, 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 the one of relevance to the this paper. Correlations are made to

decay exponentially with separation in type space. This implies that off-spring will see a set of interactions which are very similar to the interactions of the parent even when mutations make the off-spring 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 distributions. 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 and co-workers have studied the properties of sub-networks obtained by random sampling nodes in a larger network (Stumpf & Wiuf, 2005). They showed that only 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 et al. are concerned with the statistical properties of sub-nets obtained by *random* sampling. What we have found is that evolutionary dynamics is able to generate sub-networks typically characterised by exponential degree distributions, even when the full network has a binomial degree distribution.

The simple node-and-edge model discussed in Sec. 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 view point in which we consider species as either occupied or not. I.e. we turn the coarse grained measure $n(\mathbf{T}, t)$ into a binary equal to 1 when $n(\mathbf{T}, t) > 0$ and zero 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 likely that the off-spring will end up with a link to the parent species, i.e. the probability P_p should by high. An exact link between the set of probabilities (P_e, P_n, 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 in the node-and-edge model may also be of relevance to naturally occurring food webs, see e.g. (Dunne, Williams, & Martinez, 2002) and may also be of interest to protein-protein interaction networks(Ipolatov et al., 2005; Mering et al., 2002).

5 Acknowledgement

References

Albert, R., & Barabási, A.-L. (2000). Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.*, *85*, 5234.

Albert, R., & Barabasi, A.-L. (2002). Statistical mechnics of complex networks. *Rev. Mod. Phys.*, *74*, 47-57.

Anderson, P., & Jensen, H. J. (2005). Network properties, species abundance and evolution in a model of evolutionary ecology. *J. Theor. Biol.*, *232*, 551-558.

Biggs, N. L. (1994). *Algebraic graph theory.* Cambridge: Cambridge University Press.

Cheng, K., & Tang, Y. (2004). Growing networks based on the mechanism of addition and deletion. *Chin. Phys. Lett.*, *21*, 1858 – 1860.

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.

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*(20), 12917–12922.

Epstein, D., & Wang, J. (2002). A steady state model for graph power law. *2nd International Workshop on Web Dynamics*, 5234.

Farid, N., & Christensen, K. (2006). Evolvong networks through deletion and duplication. *New J. Phys.*, *8*, 212.

Hall, M., Christensen, K., Collobiano, S. A. di, & Jensen, H. J. (2002). Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E*, *66*(011904), 1–10.

Ipolatov, I., Krapivsky, P., & Yuryev, A. (2005). Duplication-divergence model of protein interaction network. *Phys. Rev. E*, *71*, 061911.

Krapivsky, P., & Redner, S. (2005). Network growth by coping. *Phys. Rev. E*, *71*, 036118.

Laird, S., & Jensen, H. J. (2006a). A steady state network model with a 1/k scale-free degree distribution. *Europhys. Lett.*, *76*, 710-716.

Laird, S., & Jensen, H. J. (2006b). The tangled nature model with inheritance and constraint: Evolutionary ecology restricted by a conserved resource. *Ecol. Complexity*, *3*, 253-262.

Laird, S., & Jensen, H. J. (2007). Correlation, selection and the evolution of species networks. *Ecol. Modelling, In Press*.

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.

Moran, P. (1962). *The statistical processes of evolutionary theory.* Oxford: Claredon Press.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, *45*, 167-256.

Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, *67*, 026126.

Salathe, M., May, R., & Bonhoeffer, S. (2005). The evolution of network topology by selective removal. *Soc. Interface*, *2*, 533 – 536.

Sarshar, N., & Roychowdhury, V. (2004). Scale-free and stable structures in complex ad hoc networks. *Phys. Rev. E*, *69*, 026101.

Stumpf, M., & Wiuf, C. (2005). Sampling properties of random graphs: The degree distribution. *Phys. Rev. E, 72*, 036118.

Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E, 67*, 056104.