

A DYNAMIC NETWORK IN A DYNAMIC POPULATION: ASYMPTOTIC PROPERTIES

TOM BRITTON,* *Stockholm University*

MATHIAS LINDHOLM,** *Uppsala University*

TATYANA TUROVA,*** *Lund University*

Abstract

We derive asymptotic properties for a stochastic dynamic network model in a stochastic dynamic population. In the model, nodes give birth to new nodes until they die, each node being equipped with a social index given at birth. During the life of a node it creates edges to other nodes, nodes with high social index at higher rate, and edges disappear randomly in time. For this model, we derive a criterion for when a giant connected component exists after the process has evolved for a long period of time, assuming that the node population grows to infinity. We also obtain an explicit expression for the degree correlation ρ (of neighbouring nodes) which shows that ρ is always positive irrespective of parameter values in one of the two treated submodels, and may be either positive or negative in the other model, depending on the parameters.

Keywords: Degree correlation; dynamic network; phase transition; random graph; stationary distribution

2010 Mathematics Subject Classification: Primary 92D30
Secondary 60J80

1. Introduction

The models of dynamical graphs are defined by the rules of attachment and deletion of vertices and edges chosen to fit particular processes in nature. Intensive study in this area began at the end of the twentieth century when a number of models were formulated primarily in the physics literature (see, e.g. Barabási *et al.* (1999), Callaway *et al.* (2001), and Malyshev (1998)). The important paper by Bollobás *et al.* (2007) provided a unified approach based on branching processes to many known models of random networks (see the reference list in Bollobás *et al.* (2007)).

In the present paper we continue the study of a Markovian model describing a random time-dynamic network in a random time-dynamic population. The model was originally defined in Britton and Lindholm (2010), extending an earlier model of Turova (2003), which in turn was derived from a general model of Malyshev (1998). The population process is a Markovian linear birth-and-death process in which individuals are assigned random independent and identically distributed (i.i.d.) social indices. Given the population process, a Markovian network is defined where edges between individuals appear and disappear in such a way that ‘social’ individuals

Received 12 April 2011; revision received 14 June 2011.

* Postal address: Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden.

Email address: tomb@math.su.se

** Current address: AFA Insurance, SE-106 27 Stockholm, Sweden.

Email address: mathias.lindholm@afaforsakring.se

*** Postal address: Mathematical Center, Lund University, Lund S-221 00, Sweden.

Email address: tatyana@maths.lth.se

tend to have more neighbours. Note also that it was shown in Turova (2002) that the model without social index and also without deletion of nodes includes as a subcase yet another model studied in Callaway *et al.* (2001).

We study the asymptotic properties of the network: the properties that the network (and population) will have after having evolved for a long time, assuming that the size of the population has grown large. We can check that a snapshot of the limiting network of Britton and Lindholm (2010) falls into the general class of inhomogeneous random graphs introduced in Bollobás *et al.* (2007). In particular, we can consult Britton and Lindholm (2010) as well as Bollobás *et al.* (2007) for the age, the type, and the degree distribution in this model.

In the present paper we shall make essential use of the theory of Bollobás *et al.* (2007) to study the phase transition in the model of Britton and Lindholm (2010). More precisely, we will determine the critical values which separate the set of parameters under which the model with a high probability has a giant connected component (i.e. of order of the entire network), and the area of parameters which do not produce a giant component. This extends earlier results of Turova (2006) for the model without social index.

Besides global properties of the network, e.g. phase transition, we shall study a characteristic which reveals more about the local structure, namely the degree correlation, or the mixing coefficient, as introduced in Callaway *et al.* (2001). Inspiring numerical results on the mixing coefficient in different empirical networks have already been presented and analyzed in Newman (2002). Recently, Bollobás *et al.* (2011) derived formulae for the mixing coefficient in terms of small subgraph counts for a rather general graph model, which includes, in particular, the inhomogeneous graphs.

According to empirical results provided in Newman (2002), real-life networks appear to be of two classes: assortative, when the mixing coefficient is positive, as in many social networks, or disassortative, when the mixing coefficient is negative, which is a feature of technological or biological networks. Still, most of the models tend to have positive degree correlation/mixing coefficient. However, Newman (2002) constructed an example of a network model where the mixing coefficient changes sign depending on the parameters of the model. More recently, Bollobás *et al.* (2011) gave another approach for constructing a network with a negative mixing coefficient. Therefore, it is a challenge to look for an example of a somewhat 'naturally grown' random network which possesses the property of enabling both positive and negative mixing coefficients depending on the parameters. As pointed out in Newman (2002), the property of assortativity affects qualitatively the sharpness of the phase transition with respect to the size of the giant component: the phase transition is sharper for the disassortative networks.

Here we provide an example of a dynamic (growing) social network with a degree correlation (mixing coefficient) that may be either negative or positive depending on the model parameters. We derive an explicit formula for the degree correlation, showing the dependence of the sign of the mixing coefficient on the parameters. Note that the formula for the mixing coefficient for the model of Callaway *et al.* (2001) derived in Newman (2002) is a subcase of the formula we prove here.

Although our model is a subcase of the general model studied in Bollobás *et al.* (2011), and, hence, the general formula from their paper should be applicable for our model as well, we find a somewhat more direct way to obtain a formula for the degree correlation in our case. We use the invariant measures for the random walk associated with the graph. It is worth noting that our method is not restricted to the particular model we study here, but would work as well for a class of inhomogeneous random graphs. However, this will be the subject of a separate study.

2. The Markovian random network in a Markovian dynamic population

Below we define the Markovian random network in a Markovian dynamic population, originally defined in Britton and Lindholm (2010). There are two versions of the model, referred to as the uniform (U) version and the preferential (P) version. In what follows a network denotes a finite set of nodes (the population) together with undirected edges connecting pairs of nodes. Nodes that are directly connected by an edge are called neighbours (in sociological applications nodes correspond to individuals, edges to some type of friendship, and neighbours are referred to as friends). The model is dynamic in the double sense that nodes are born and may die, and the same applies to edges.

2.1. The model and its two versions

We first define the node population dynamics. Let $Y(t)$ denote the number of nodes alive at time t , and assume that $Y(0) = 1$. While alive, each node gives birth to new nodes at the constant rate λ and each node lives for an exponentially distributed time having mean $1/\mu$ (denoted $\text{Exp}(\mu)$), so each node dies at the rate μ . We assume that $\lambda > \mu$, implying that the expected number of births during a life span is larger than 1. This means that the node population is modelled by a Markovian supercritical branching process. Additional to this, each node i is at birth given a random ‘social index’ S_i having distribution F_S on \mathbb{R}^+ , i.i.d. for different nodes. We will throughout assume that S has finite mean i.e. $E[S] := \mu_S < \infty$.

There are two versions of the model for births and deaths of edges, the U- and P-versions, both being Markovian given the node population and their social indices. In both versions, nodes are isolated at birth, i.e. have no neighbours. During the life of a node i , having social index $S_i = s$, it creates new neighbour edges at rate αs . The difference between the two versions lies in how neighbour nodes are selected. In the U-version the neighbour node is chosen uniformly at random among all living nodes. In the P-version the neighbour node is instead chosen at random among all living nodes with probabilities proportional to their social index. Finally, in both versions each edge is removed, independently of everything else, at the rate β . If a node dies, all edges connected to the node in question are removed. We emphasize that a node gets new neighbours in two ways: it ‘creates’ new neighbour edges itself but may also be selected as the neighbour of another node that has created a neighbour edge.

2.2. Comments on the model

The model has four parameters: the birth and death rates of nodes, λ and μ , respectively, the death rate β of edges, and α , which is related to the birth rate of new edges. Beside these four parameters, we have the distribution F_S for the social indices $\{S_i\}$ (in what follows we let f_s denote the corresponding density function).

The P-version of the model is inspired by the preferential attachment model (see Barabási and Albert (1999)); however, in the P-version the probability of receiving edges is determined at birth, whereas it is determined by random events during life in the preferential attachment model. In both models there is also an age factor, in the sense that older nodes tend to have more edges.

Some submodels are worth mentioning, such as the case where there is no node heterogeneity and nodes have the same social index $S \equiv s$ (for example, set to 1 without loss of generality). Loosely speaking, allowing node heterogeneity makes it possible to achieve degree distributions having heavy tails: if F_S is heavy tailed, there will be some nodes with very high social indices that hence have a large number of neighbours (cf. the next section). The case where $\mu = 0$ and $S \equiv 1$ has been studied in Turova (2002), (2006), who derived more results for this submodel.

2.3. Known results of the model

The model allows for multiple edges and self-loops, which usually does not make sense in applications. However, it was shown in Britton and Lindholm (2010) that the *proportion* of such edges is asymptotically negligible as long as $E[S] < \infty$ for the U-version, and as long as $E[S^2] < \infty$ for the P-version. As a consequence, the network will then have identical properties if loops and multiple edges are ignored or not allowed.

In the rest of the paper we will consider only the case where the node population tends to ∞ , i.e. we condition on the event $B := \{\omega; Y(t) \rightarrow \infty\}$, which has positive probability since it was assumed that $\lambda > \mu$. The first two results follow from standard branching process theory.

Asymptotic population size. As $t \rightarrow \infty$,

$$Y(t)e^{-t(\lambda-\mu)} \xrightarrow{D} \text{Exp}(1).$$

Stable age distribution. As $t \rightarrow \infty$, the age A of a randomly selected living node will satisfy $A \sim \text{Exp}(\lambda)$, having mean $1/\lambda$; see, e.g. Section 3.4 of Haccou *et al.* (2005).

Social index distribution. Since the social indices are defined to be i.i.d. with distribution F_S , and they affect only the occurrence of edges, the social index of a randomly selected living node will have the same social index distribution F_S .

Asymptotic degree distribution. As $t \rightarrow \infty$, the degree (i.e. the number of neighbours) D of a randomly selected living individual has a mixed Poisson distribution in both versions of the model (see Britton and Lindholm (2010)). For the U-version,

$$D^U \sim \text{MixPo}\left(\frac{\alpha(S + \mu_S)(1 - e^{-(\beta+\mu)A})}{\beta + \mu}\right), \tag{2.1}$$

where $A \sim \text{Exp}(\lambda)$ and $S \sim F_S$ are independent. This simply means that, conditional on $A = a$ and $S = s$, the degree is Poisson distributed with mean parameter equal to $\alpha(s + \mu_S) \times (1 - e^{-(\beta+\mu)a})/(\beta + \mu)$. The mean and variance of this distribution are given by

$$\begin{aligned} E[D^U] &= \frac{2\alpha}{\lambda + \beta + \mu} \mu_S, \\ \text{var}[D^U] &= \frac{2\alpha}{\lambda + \beta + \mu} \mu_S + \frac{4\lambda\alpha^2}{(\lambda + \beta + \mu)^2(\lambda + 2(\beta + \mu))} \mu_S^2 \\ &\quad + \frac{2\alpha^2}{(\lambda + \beta + \mu)(\lambda + 2(\beta + \mu))} \text{var}[S]. \end{aligned}$$

For the P-version of the model, the degree distribution satisfies

$$D^P \sim \text{MixPo}\left(\frac{2\alpha S(1 - e^{-(\beta+\mu)A})}{\beta + \mu}\right), \tag{2.2}$$

where $A \sim \text{Exp}(\lambda)$ and $S \sim F_S$ are independent. The mean and variance are given by

$$\begin{aligned} E[D^P] &= \frac{2\alpha}{\lambda + \beta + \mu} \mu_S, \\ \text{var}[D^P] &= \frac{2\alpha}{\lambda + \beta + \mu} \mu_S + \frac{4\lambda\alpha^2}{(\lambda + \beta + \mu)^2(\lambda + 2(\beta + \mu))} \mu_S^2 \\ &\quad + \frac{8\alpha^2}{(\lambda + \beta + \mu)(\lambda + 2(\beta + \mu))} \text{var}[S], \end{aligned}$$

i.e. the same mean degree but a larger variance. For both versions, it is seen that the variance of the degree distribution increases with the variance of the social index distribution, so having a heavy-tailed social index distribution F_S implies that the degree distribution will also be heavy tailed.

3. Results

In what follows we derive some further properties of the model, still assuming that $t \rightarrow \infty$ and conditioning on the fact that the node population grows beyond all limits. We do this for the U-version and only present the results for the P-version, which are obtained in a similar way.

3.1. Type distribution of neighbour nodes

Assume that t is large, and consider a randomly selected individual alive at this time of age a and social index s . We derive the distribution of the number of (a', s') -neighbours that this (a, s) -node has. The number of neighbours of any type (a', s') will be Poisson distributed, so we first compute the mean of the distribution and then divide by the total mean, thus obtaining the type distribution.

How many neighbours of type (a', s') does the (a, s) -individual have? Any edge must have been created at some time $\tau \in (t - (a \wedge a'), t)$ (' \wedge ' denotes the minimum: both individuals must have been born). At such τ the (a, s) -individual creates edges with density $\alpha s \lambda e^{-\lambda(\tau - (t - a'))} f_S(s')$: αs is the rate at which the individual creates edges and $\lambda e^{-\lambda(\tau - (t - a'))} f_S(s')$ is the density of the fraction of nodes having age $\tau - (t - a')$ (at τ) and type s' . Similarly, the (a, s) -individual receives edges with density $\alpha s' \lambda e^{-\lambda(\tau - (t - a'))} f_S(s')$, as there are $Y(\tau) \lambda e^{-\lambda(\tau - (t - a'))} f_S(s')$ having the 'correct' age and type, each of which creates edges at rate $\alpha s'$ and, with probability $1/Y(\tau)$, each of these created edges reaches the (a, s) -individual. Any such edge created at τ remains at t if and only if both the (a', s') -individual and the edge survive until t (we have already conditioned on the fact that the (a, s) -individual is alive at time t). The probability of this equals $e^{-(\beta + \mu)(t - \tau)}$. Hence, the total expected number of edges with (a', s') -individuals equals

$$\begin{aligned}
 m^U(a', s' \mid a, s) da' ds' &= da' ds' \int_{t - (a \wedge a')}^t \alpha(s + s') \lambda e^{-\lambda(\tau - (t - a'))} f_S(s') e^{-(\beta + \mu)(t - \tau)} d\tau \\
 &= \alpha(s + s') f_S(s') \lambda e^{-\lambda a'} \frac{e^{(\lambda - \beta - \mu)(a \wedge a')} - 1}{\lambda - \beta - \mu} da' ds'. \tag{3.1}
 \end{aligned}$$

It is easy to check from this that the integral over all a' and s' of the expected total number of neighbours of any type of the (a, s) -node equals

$$m^U(\cdot, \cdot \mid a, s) = \frac{\alpha(s + \mu_S)}{\beta + \mu} \left(1 - e^{-(\beta + \mu)a} \right). \tag{3.2}$$

This also confirms that the unconditional degree distribution has the mixed Poisson distribution specified in (2.1) (while, conditionally on the age a and type s , it has Poisson distribution with mean given by (3.2)).

Consequently, for the U-version of the model, the type distribution of the neighbours of an individual of type (a, s) , which equals the expected number of (a', s') -types divided by the

total expected number of types, becomes

$$f^U(a', s' | a, s) = \frac{m^U(a', s' | a, s)}{m^U(\cdot, \cdot | a, s)} = \frac{(s + s') f_S(s')}{s + \mu_S} \frac{(\beta + \mu) \lambda e^{-\lambda a'} (e^{(\lambda - \beta - \mu)(a \wedge a')} - 1)}{(\lambda - \beta - \mu)(1 - e^{-(\beta + \mu)a}}.$$

We hence have conditional independence, i.e. $f^U(a', s' | a, s) = f^U(a' | a) f^U(s' | s)$, where

$$f^U(a' | a) = \frac{(\beta + \mu) \lambda e^{-\lambda a'} (e^{(\lambda - \beta - \mu)(a \wedge a')} - 1)}{(\lambda - \beta - \mu)(1 - e^{-(\beta + \mu)a}} \tag{3.3}$$

$$f^U(s' | s) = \frac{(s + s') f_S(s')}{s + \mu_S} \tag{3.4}$$

We draw attention to the abuse of notation here, and in what follows, in that $f^U(a' | a)$ and $f^U(s' | s)$ really denote different density functions, as indicated by the difference in arguments.

Using similar arguments, for the P-version of the model, we obtain $f^P(a', s' | a, s) = f^P(a' | a) f^P(s' | s)$, where

$$f^P(a' | a) = \frac{(\beta + \mu) \lambda e^{-\lambda a'} (e^{(\lambda - \beta - \mu)(a \wedge a')} - 1)}{(\lambda - \beta - \mu)(1 - e^{-(\beta + \mu)a}} \tag{3.5}$$

$$f^P(s' | s) = \frac{s' f_S(s')}{\mu_S} \tag{3.6}$$

i.e. the same density as for the U-version with respect to age but a different density with respect to social index.

3.2. Phase transition

Here we will apply the results of Bollobás *et al.* (2007) to the above models in the limit as $t \rightarrow \infty$. Therefore, we will first put our model into a general setup of the theory (and notation) of inhomogeneous random graphs of Bollobás *et al.* (2007).

Let us enumerate the individuals at time t by $i = 1, \dots, n := Y(t)$. Let $X_i = (A_i, S_i)$, where A_i and S_i respectively denote the age and social index of individual i . Note that this means that this individual was born at time $t - A_i$ (and remains alive at time t). The random variables X_i are independent for different i . For a randomly selected i , as we argued above, the distribution of the social index follows the original social index distribution and A_i is $\text{Exp}(\gamma)$; denote the corresponding measures by μ_1 and μ_2 :

$$\mu_1(da) = \lambda e^{-\lambda a} da, \quad \mu_2(ds) = f_S(s) ds. \tag{3.7}$$

We shall call X_i the type of individual i . Given the types $X_i = (a_i, s_i)$ and $X_j = (a_j, s_j)$ of individuals i and j , we derive the probability $p_{ij}(n)$ that they are connected. From above we know that the expected number of (a', s') -individuals individual i is connected to equals $m(a', s' | a_i, s_i) da' ds'$, defined by (3.1) for the U-version (and similarly for the P-version). In total, if $Y(t) = n$, the expected total number of individuals of type (a', s') equals $n \lambda e^{-\lambda a'} da' \times f_S(s') ds'$. This reasoning implies that the probability that i is connected to j (of type (a_j, s_j)) equals the ratio of these two expressions:

$$p_{ij}(n) = \frac{1}{n} \kappa_1(a_i, a_j) \kappa_2(s_i, s_j). \tag{3.8}$$

Here

$$\kappa_1(a_i, a_j) = \frac{e^{(\lambda - \beta - \mu)(a_i \wedge a_j)} - 1}{\lambda - \beta - \mu}$$

and

$$\kappa_2(s_i, s_j) = \begin{cases} \kappa_2^U(s_i, s_j) = \alpha(s + s') & \text{in the U-version,} \\ \kappa_2^P(s_i, s_j) = 2\alpha s s' & \text{in the P-version.} \end{cases}$$

Let $\mu = \mu_1 \times \mu_2$, which, by (3.7), is a probability measure on $\mathcal{S} = \mathbb{R}_+^2$, and let

$$\kappa = \kappa(x_i, x_j) = \kappa_1(a_i, a_j)\kappa_2(s_i, s_j).$$

Given the sequence x_1, \dots, x_n , we let $G^{\mathcal{V}}(n, \kappa)$ be the random graph on $\{1, \dots, n\}$ such that any two vertices i and j are connected by an edge, independently of the other vertices, with probability given by (3.8), which we write as

$$p_{ij}(n) = \min\left\{\frac{\kappa(x_i, x_j)}{n}, 1\right\}.$$

Hence, conditionally on $Y(t) = n$, the graph $G^{\mathcal{V}}(n, \kappa)$ describes our model. On the other hand, $G^{\mathcal{V}}(n, \kappa)$ with κ as above satisfies the definition of the inhomogeneous random graph from Bollobás *et al.* (2007), and, thus, we can apply some known results here.

Recall first the fundamental result on phase transitions in the inhomogeneous random graphs. Define

$$T_{\kappa} f(x) = \int_S \kappa(x, y) f(y) \, d\mu(y)$$

and

$$\|T_{\kappa}\| = \sup\{\|T_{\kappa} f\|_2 : f \geq 0, \|f\|_2 \leq 1\}.$$

Then, by Theorem 3.1 of Bollobás *et al.* (2007), the largest connected component of $G^{\mathcal{V}}(n, \kappa)$, denoted by $C_1(G^{\mathcal{V}}(n, \kappa))$, satisfies

$$\frac{1}{n} C_1(G^{\mathcal{V}}(n, \kappa)) \xrightarrow{P} \rho_{\kappa} = \int_S \rho(x) \, d\mu(x) \quad \text{as } n \rightarrow \infty,$$

where $\rho_{\kappa} > 0$ if and only if $\|T_{\kappa}\| > 1$.

Then, for our model as well, conditionally on the number $Y(t)$ of individuals in the population at time t , we have, with a high probability, approximately $\rho_{\kappa} Y(t)$ individuals connected; this holds unconditionally as well. The function $\rho_{\kappa}(x)$ is described more precisely in the following theorem.

Theorem 3.1. (Theorem 6.1 of Bollobás *et al.* (2007).) *Suppose that κ is the kernel on (\mathcal{S}, μ) , that $\kappa \in L^1$, and that*

$$\int_{\mathcal{S}} \kappa(x, y) \, d\mu(y) < \infty$$

for every $x \in S$. Then $\rho_{\kappa}(x)$ is the maximal solution to

$$f(x) = 1 - \exp\{-T_{\kappa} f(x)\}. \tag{3.9}$$

Furthermore, the following statements hold.

- (i) *If $\|T_{\kappa}\| \leq 1$ then $\rho_{\kappa}(x) = 0$ for every x , and (3.9) has only the zero solution.*

(ii) If $1 < \|T_k\| \leq \infty$ then $\rho_\kappa(x) > 0$ on a set of positive measure. If, in addition, κ is irreducible, i.e. if

$A \subseteq S$ and $\kappa = 0$ almost everywhere on $A \times (S \setminus A)$ implies that $\mu(A)$ or $\mu(S \setminus A) = 0$, then $\rho_\kappa(x) > 0$ for almost every x , and $\rho_\kappa(x)$ is the only nonzero solution of (3.9).

First we shall derive conditions for $\rho_\kappa(x) > 0$. By the cited results, this amounts to computing $\|T_k\|$. Unfortunately, there is no general formula to simply compute the norm $\|T_k\|$. Therefore, we shall use a criterion for the condition when $\|T_k\| > 1$, which is what we need here.

Proposition 3.1. (Proposition 17.2 of Bollobás *et al.* (2007).) For $k \geq 1$, let

$$\alpha(k) := \frac{1}{2} \int_{\mathcal{S}^{k+1}} \kappa(x_0, x_1)\kappa(x_1, x_2) \cdots \kappa(x_{k-1}, x_k) \, d\mu(x_0) \cdots d\mu(x_k).$$

Then $\|T_k\| > 1$ if and only if $\alpha(k) \rightarrow \infty$ as $k \rightarrow \infty$.

It follows from definition (3.7) that, for our model,

$$\alpha(k) = \frac{1}{2}\alpha_1(k)\alpha_2(k), \tag{3.10}$$

where, for $i = 1, 2$,

$$\alpha_i(k) = \int_{\mathbb{R}_+^{k+1}} \kappa_i(u_0, u_1)\kappa_i(u_1, u_2) \cdots \kappa_i(u_{k-1}, u_k) \, d\mu_i(u_0) \cdots d\mu_i(u_k). \tag{3.11}$$

First we shall compute $\alpha_2(k)$. We start with the U-version in the next proposition.

Proposition 3.2. Consider the matrix

$$A = \begin{pmatrix} \mathbb{E}[S] & 1 \\ \mathbb{E}[S^2] & \mathbb{E}[S] \end{pmatrix}.$$

Then, for all $k \geq 3$,

$$\alpha_2^U(k) = \alpha^k \left\{ A^{k-1} \begin{pmatrix} 2 \mathbb{E}[S] \\ (\mathbb{E}[S])^2 + \mathbb{E}[S^2] \end{pmatrix} \right\}_1, \tag{3.12}$$

where, for a vector B , we denote by $\{B\}_i$ component i .

Proof. Let S_0, S_1, \dots be i.i.d. copies of the random variable S . Then, by definition (3.11),

$$\alpha_2^U(k) = \alpha^k \mathbb{E} \left[\prod_{i=1}^k (S_{i-1} + S_i) \right] =: \alpha^k f(k), \tag{3.13}$$

where

$$f(k) = \mathbb{E} \left[\prod_{i=1}^k (S_{i-1} + S_i) \right].$$

Also, define

$$g(k) = \mathbb{E} \left[\left(\prod_{i=1}^k (S_{i-1} + S_i) \right) S_k \right].$$

Then we recursively derive

$$f(k) = g(k - 1) + f(k - 1)E[S]$$

and

$$g(k) = g(k - 1)E[S] + f(k - 1)E[S^2]$$

for $k > 1$, with

$$f(1) = 2E[S], \quad g(1) = (E[S])^2 + E[S^2]. \tag{3.14}$$

Hence, for all $k > 1$,

$$\begin{pmatrix} f(k) \\ g(k) \end{pmatrix} = A \begin{pmatrix} f(k - 1) \\ g(k - 1) \end{pmatrix} = A^{k-1} \begin{pmatrix} f(1) \\ g(1) \end{pmatrix},$$

and (3.12) follows from (3.13) upon using the above formula and (3.14).

Corollary 3.1. *We have*

$$\lim_{k \rightarrow \infty} (\alpha_2^U(k))^{1/k} = \alpha(E[S] + \sqrt{E[S^2]}). \tag{3.15}$$

In the case of the P-version, when $\alpha_2 = \alpha_2^P$, it is straightforward to derive from (3.11) that

$$\alpha_2^P(k) = (2\alpha)^k (E[S])^2 (E[S^2])^{k-2},$$

which yields

$$\lim_{k \rightarrow \infty} (\alpha_2^P(k))^{1/k} = 2\alpha E[S^2]. \tag{3.16}$$

Now define

$$c^{cr}(\lambda, \mu + \beta) := \sup \left\{ x > 0 : \sum_{k=2}^{\infty} x^k \alpha_1(k) < \infty \right\}.$$

Then, owing to definitions (3.10) and (3.11), together with asymptotics (3.15) and (3.16), we have the following criteria, which hold for both the P-version and U-version.

Corollary 3.2. *If*

$$\lim_{k \rightarrow \infty} (\alpha_2(k))^{1/k} > c^{cr}(\lambda, \mu + \beta)$$

then $\alpha(k) \rightarrow \infty$; *if*

$$\lim_{k \rightarrow \infty} (\alpha_2(k))^{1/k} < c^{cr}(\lambda, \mu + \beta)$$

then $\alpha(k) < \infty$.

The value $c^{cr}(\lambda, \mu + \beta)$ is known from Turova (2003). Let us record this result here. Write first

$$\begin{aligned} \alpha_1(k) &= \int_{\mathbb{R}_+^{k+1}} \kappa_1(a_0, a_1) \cdots \kappa_1(a_{k-1}, a_k) \lambda e^{-\lambda a_0} da_0 \cdots \lambda e^{-\lambda a_k} da_k \\ &= E \left[\kappa_1 \left(\frac{X_0}{\lambda}, \frac{X_1}{\lambda} \right) \cdots \kappa_1 \left(\frac{X_{k-1}}{\lambda}, \frac{X_k}{\lambda} \right) \right], \end{aligned}$$

where X_0, \dots, X_k are i.i.d. random variables with a common $\text{Exp}(1)$ distribution. Then, by the results of Turova (2007b) (see Equation (1.7) therein, or consult Bollobás *et al.* (2007)), the value $c^{\text{cr}}(\lambda, \mu + \beta)$ is the smallest positive root of

$$H(x) = 1 + \sum_{n=1}^{\infty} (-1)^n \left(\frac{x}{\mu + \beta}\right)^n \frac{1}{n!} \left[\prod_{l=1}^n \frac{1}{1 + (l-1)(\mu + \beta)/\lambda} \right]. \tag{3.17}$$

Finally, combining Corollary 3.2 and Proposition 3.1 with asymptotics (3.15) and (3.16), we can derive, apart from the critical case, the necessary and sufficient conditions for the existence of the giant component.

Corollary 3.3. *Define*

$$\kappa^{\text{U}} = \kappa_1 \kappa_2^{\text{U}}, \quad \kappa^{\text{P}} = \kappa_1 \kappa_2^{\text{P}},$$

and let

$$R^{\text{U}}(\alpha, \beta, \lambda, \gamma, S) := \frac{\alpha(\mathbb{E}[S] + \sqrt{\mathbb{E}[S^2]})}{c^{\text{cr}}(\lambda, \mu + \beta)} \quad \text{and} \quad R^{\text{P}}(\alpha, \beta, \lambda, \gamma, S) := \frac{2\alpha \mathbb{E}[S^2]}{c^{\text{cr}}(\lambda, \mu + \beta)}. \tag{3.18}$$

Then there exists a giant component in the U-model if $R^{\text{U}}(\alpha, \beta, \lambda, \gamma, S) > 1$, and there is no giant component if $R^{\text{U}}(\alpha, \beta, \lambda, \gamma, S) < 1$.

Similarly, there exists a giant component in the P-model if $R^{\text{P}}(\alpha, \beta, \lambda, \gamma, S) > 1$, and there is no giant component if $R^{\text{P}}(\alpha, \beta, \lambda, \gamma, S) < 1$.

Note also that although we define the function $c^{\text{cr}}(\lambda, u)$ for $u > 0$, the model with $\mu + \beta = 0$ is treated similarly; in fact, when $S \equiv 1$ also, this model becomes the model studied in Callaway *et al.* (2001). We define $c^{\text{cr}}(\lambda, 0)$ as the critical parameter, above which there is a giant component. It is known (see Callaway *et al.* (2001)) that $c^{\text{cr}}(\lambda, 0) = \frac{1}{4}$, and, moreover, the function $c^{\text{cr}}(\lambda, u)$ is continuous at $u = 0$ (see Bollobás *et al.* (2007) and Turova (2007)).

In general, there is no closed form for the function c^{cr} , but we can mention some qualitative properties of this function. First, simply from the model it follows that $c^{\text{cr}}(\lambda, \mu + \beta)$ is decreasing in λ and increasing in $\mu + \beta$, i.e. both in μ and β . For any $u > 0$, denote by $c(u)$ the smallest positive root of the rescaled function (3.17)

$$H_u(x) = 1 + \sum_{n=1}^{\infty} (-1)^n x^n \frac{1}{n!} \left[\prod_{l=1}^n \frac{1}{1 + (l-1)u} \right].$$

Then

$$c^{\text{cr}}(\lambda, \mu + \beta) = (\mu + \beta)c\left(\frac{\mu + \beta}{\lambda}\right).$$

In conclusion, both $R^{\text{U}}(\alpha, \beta, \lambda, \gamma, S)$ and $R^{\text{P}}(\alpha, \beta, \lambda, \gamma, S)$ are increasing in λ and α (the birth rate of individuals and edges, respectively) as well as the first and second moments of S , whereas they are both decreasing in μ and β (the death rate of individuals and edges, respectively).

3.3. Degree correlation

We now derive an expression for the degree correlation ρ (also known as the ‘mixing coefficient’) of the network when it has grown and reached its stationary phase. That is, after a long time t we take a snapshot of the network and compute the degree correlation ρ , which is

defined as the correlation of the degrees of the two adjacent nodes of a randomly selected edge in the network. We derive ρ in two different ways. In the first method we pick a random *node* in the snapshot network and then perform a random walk on this fixed network. When this random walk has reached stationarity, we consider the node in one step as the ‘first’ node and then pick the ‘second’ node randomly among the neighbouring nodes, and let the edge between these two nodes be our randomly selected edge. In the second method we pick the ‘first’ node randomly in a size-biased way in the snapshot network (such that the node is adjacent to a randomly selected *edge*) and we pick the ‘second’ node randomly among the neighbours of this individual (the size-biased distribution of a nonnegative random variable X with probability density function (PDF) $f_X(x)$ and finite mean μ_X has PDF $xf_X(x)/\mu_X$ —below we denote such a random variable by \bar{X}). As will be seen, the two methods of picking neighbouring nodes give the same result. The degree correlation is then obtained by computing the degree correlation of the two neighbouring nodes.

3.3.1. *The type distribution obtained using the random walk method.* We start with the first method, in which we pick a node at random among all living nodes. From the results of Section 2.3, this means that the node has an exponential age distribution and an independent social index distribution given by the original distribution. More precisely, the type distribution of a randomly selected node equals $\lambda e^{-\lambda a} da f_S(s) ds$. The neighbours of this node, conditional on its type, all have distribution $f(a', s' | a, s)$, defined in Section 3.1, where its form depends on whether we consider the U-version or the P-version of the model. Iterating this procedure gives the type distribution of nodes after additional steps in the random walk. Eventually, this random walk reaches stationarity and the distribution $f_\infty(a', s')$ of the type of node is then given by the solution to the functional equation

$$f_\infty(a', s') = \int \int f_\infty(a, s) f(a', s' | s, a) ds da.$$

Since $f(a', s' | s, a) = f(a' | a) f(s' | s)$ in both versions of the model, it follows that the age and social index distributions are independent in the stationary distribution as well: $f_\infty(a', s') = f_\infty(a') f_\infty(s')$. The corresponding densities are therefore the solutions to

$$f_\infty(a') = \int f_\infty(a) f(a' | a) da, \tag{3.19}$$

$$f_\infty(s') = \int f_\infty(s) f(s' | s) ds. \tag{3.20}$$

The functions $f(a' | a)$ and $f(s' | s)$ were defined in Section 3.1, with $f(a' | a)$ being the same for the two versions and $f(s' | s)$ being different for the two versions. The (unique) solution $f_\infty(a')$ to (3.19) is given by

$$f_\infty(a') = \lambda \left(1 + \frac{\lambda}{\beta + \mu} \right) e^{-\lambda a'} (1 - e^{-(\beta + \mu) a'}). \tag{3.21}$$

The solution to (3.20) depends on which version of the model we consider, i.e. whether we use $f^P(s' | s)$ or $f^U(s' | s)$. The (unique) solutions $f_\infty^P(s')$ and $f_\infty^U(s')$ are given by

$$f_\infty^P(s') = \frac{s' f_S(s')}{\mu_S}, \tag{3.22}$$

and

$$f_{\infty}^U(s') = \frac{s' + \mu_S}{2\mu_S} f_S(s'). \tag{3.23}$$

Hence, the P-version has the size-biased version of the social index distribution, whereas the U-version has a mixture of the original and the size-biased social index distribution.

The stationary distribution defined by (3.21) and (3.22) for the P-version, and (3.21) and (3.23) for the U-version, is the distribution of the ‘first’ node of a randomly selected edge. Given the type (a', s') of this individual, a randomly selected neighbour of this individual has the previously defined type distribution $f(a, s | a', s') = f(a | a')f(s | s')$ (where $f(s | s')$, but not $f(a | a')$, depends on the model version we consider).

3.3.2. *The type distribution obtained using the size-biased method.* We now derive the type distribution of the nodes of a randomly selected edge using the second method described above, which will be seen to give the same solution. We do this for the P-version of the model.

We know from before (see (2.2)) that the degree D of a randomly selected node in the snapshot network is mixed Poisson with (random) parameter

$$\frac{2\alpha S(1 - e^{-(\beta+\mu)A})}{\beta + \mu}, \tag{3.24}$$

where S has the original social index distribution f_S and A is independent $\text{Exp}(\lambda)$. We now instead pick our ‘first’ node as a node of a randomly selected edge. As a consequence, this node will have the size-biased version \tilde{D} of this random variable (because a node of degree k has probability proportional to k of being selected). The size-biased distribution of mixed Poisson is also mixed Poisson, but with the size-biased version of the random parameter. The size-biased distribution of the random parameter (3.24) equals

$$\frac{2\alpha}{\beta + \mu} \tilde{S} (1 - \widetilde{e^{-(\beta+\mu)A}}).$$

By this tilde notation we mean that the type (A^*, S^*) of the ‘first’ node will have social index distribution \tilde{S} (the size-biased distribution of the original density, having density $sf_S(s)/\mu_S$, which agrees with (3.22)). The age distribution A^* will be independent and have distribution defined by $1 - e^{-(\beta+\mu)A^*}$, being distributed like $(1 - \widetilde{e^{-(\beta+\mu)A}})$, where $A \sim \text{Exp}(\lambda)$. But, since $A \sim \text{Exp}(\lambda)$, it follows that

$$1 - e^{-(\beta+\mu)A} \sim \text{Beta}\left(1, \frac{\lambda}{\beta + \mu}\right).$$

From this, it follows that the size-biased distribution

$$(1 - \widetilde{e^{-(\beta+\mu)A}}) \sim \text{Beta}\left(2, \frac{\lambda}{\beta + \mu}\right).$$

So $1 - e^{-(\beta+\mu)A^*}$ has this distribution, implying that $e^{-(\beta+\mu)A^*} \sim \text{Beta}(\lambda/(\beta + \mu), 2)$. From this, it follows after some straightforward calculations that A^* has exactly the density given in (3.21). As in the previous subsection, given the type (a', s') of the ‘first’ node, a randomly selected neighbour of this node has the type distribution $f(a, s | a', s') = f(a | a')f(s | s')$.

3.3.3. *The degree correlation ρ .* In the two previous subsections we derived the type distribution of the two nodes of a randomly selected edge. The type (A^*, S^*) of the ‘first’ node has density $f_\infty(a')f_\infty(s')$, where $f_\infty(a')$ is defined in (3.21) for both versions of the model, and $f_\infty(s')$ is given by (3.22) for the P-version (and not explicit for the U-version). Given the type (a', s') of the first node, the second node has type distribution $f(a, s | a', s') = f(a | a')f(s | s')$, where $f(a | a')$ and $f(s | s')$ are as defined in (3.3)–(3.6). We now derive the degree correlation between two such randomly selected nodes.

In order to compute the degree correlation of the nodes adjacent to the randomly selected edge, we condition on the types. Given the types, we know from before that the degree distribution is Poisson: an individual of type (a, s) has a Poisson number of neighbours and the mean $g(a, s) = g_1(a)g_2(s)$ equals

$$g_1(a) = \frac{1 - e^{-(\beta+\mu)a}}{\beta + \mu} \quad \text{and} \quad g_2(s) = \begin{cases} g_2^U(s) = \alpha(s + \mu_S) & \text{for the U-version,} \\ g_2^P(s) = 2\alpha s & \text{for the P-version.} \end{cases} \quad (3.25)$$

We now compute the degree correlation using these results. Let D_1 denote the degree of the ‘first’ node, and let D_2 denote its selected neighbour. In order to compute the degree correlation $\rho(D_1, D_2)$, we need to compute the following expectations: $E[D_1] = E[D_2]$, $E[D_1^2] = E[D_2^2]$, and $E[D_1 D_2]$ (the first as well as the second moment of D_1 and D_2 are the same since we have reached stationarity). We do this excluding the edge connecting the two nodes (subtracting the value 1 everywhere does not change the degree correlation). The conditional degree distributions remain Poisson with the same means because if $X \sim \text{Po}(\beta)$ then the conditional distribution of $X - 1$, conditional on $X \geq 1$, is also $\text{Po}(\beta)$. We hence obtain

$$\begin{aligned} E[D_1] &= E[D_2] = \int g_1(a)f_\infty(a) da \int g_2(s)f_\infty(s) ds, \\ E[D_1^2] &= E[D_2^2] = \int \int (g_1(a)g_2(s) + g_1^2(a)g_2^2(s))f_\infty(a)f_\infty(s) da ds, \\ E[D_1 D_2] &= \int \int g_1(a)g_1(a')f_\infty(a)f(a' | a) da da' \int \int g_2(s)g_2(s')f_\infty(s)f(s' | s) ds ds'. \end{aligned}$$

Given these expressions, the degree correlation is given by

$$\rho(D_1, D_2) = \frac{E(D_1 D_2) - E(D_1) E(D_2)}{E(D_1^2) - (E(D_1))^2} = \frac{C(D_1, D_2)}{\text{var}[D_1]}. \quad (3.26)$$

Perhaps the most important issue is to determine if ρ , or, equivalently, the covariance C , is positive or negative. For this, it is sufficient to compute the numerator of (3.26). Using $f_\infty(a, s)$ (defined by (3.21), (3.22), and (3.23)) and (3.25) in the expressions above for the P-version and U-version, separately, standard but tedious calculations reveal that

$$\begin{aligned} E[D_1^P] &= E[D_2^P] = \frac{2}{\lambda + 2(\beta + \mu)} \frac{2\alpha E[S^2]}{E[S]}, \\ E[D_1^P D_2^P] &= \frac{5\lambda + 6(\beta + \mu)}{(\lambda + \beta + \mu)(\lambda + 2(\beta + \mu))(\lambda + 3(\beta + \mu))} \left(\frac{2\alpha E[S^2]}{E[S]} \right)^2, \end{aligned}$$

as well as

$$E[D_1^U] = E[D_2^U] = \frac{2}{\lambda + 2(\beta + \mu)} \frac{\alpha(E[S^2] + 3(E[S])^2)}{2E[S]},$$

$$E[D_1^U D_2^U] = \frac{5\lambda + 6(\beta + \mu)}{(\lambda + \beta + \mu)(\lambda + 2(\beta + \mu))(\lambda + 3(\beta + \mu))} 2\alpha^2(E[S^2] + (E[S])^2).$$

This gives, for the P-version,

$$C(D_1^P, D_2^P) = \frac{\lambda^2}{(\lambda + \beta + \mu)(\lambda + 2(\beta + \mu))^2(\lambda + 3(\beta + \mu))} (2\alpha E(S^2))^2.$$

Since all parameters are positive, we conclude that, for the P-version of the model, the covariance, and, hence, also the degree correlation ρ , is always positive. This is true irrespective of the model parameters and choices of social index distribution.

For the U-version, the picture is different. Introduce $\gamma = \beta + \mu$, and compute

$$C(D_1^U, D_2^U) = \alpha^2 \frac{6\gamma + 5\lambda}{(\lambda + \gamma)(\lambda + 2\gamma)(\lambda + 3\gamma)} 2(E[S^2] + (E[S])^2) \tag{3.27}$$

$$- \alpha^2 \frac{1}{(\lambda + 2\gamma)^2} \left(\frac{E[S^2] + 3(E[S])^2}{E[S]} \right)^2$$

$$= \frac{\alpha^2}{(\lambda + 2\gamma)^2} (E S)^2$$

$$\times \left(\frac{(6\gamma + 5\lambda)(\lambda + 2\gamma)}{(\lambda + \gamma)(\lambda + 3\gamma)} 2 \left(\frac{E[S^2]}{(E[S])^2} + 1 \right) - \left(\frac{E[S^2]}{(E[S])^2} + 3 \right)^2 \right).$$

Let us also define

$$a = \frac{\lambda^2}{(\lambda + \gamma)(\lambda + 3\gamma)}.$$

Then we can rewrite (3.27) as

$$C(D_1^U, D_2^U) = \frac{\alpha^2}{(\lambda + 2\gamma)^2} (E S)^2 \left(2(a + 4) \left(\frac{E[S^2]}{(E[S])^2} + 1 \right) - \left(\frac{E[S^2]}{(E[S])^2} + 3 \right)^2 \right)$$

$$= \frac{\alpha^2}{(\lambda + 2\gamma)^2} (E S)^2 \left(- \left(\frac{E[S^2]}{(E[S])^2} \right)^2 + 2 \frac{E[S^2]}{(E[S])^2} (a + 1) - (1 - 2a) \right).$$

It follows from this, together with the facts that $E[S^2]/(E[S])^2 \geq 1$ and $0 < a < 1$, that if

$$\frac{E[S^2]}{(E[S])^2} < 1 + a + \sqrt{a^2 + 4a}$$

then $C(D_1^U, D_2^U) > 0$, and, hence, also that $\rho^U > 0$. If

$$\frac{E[S^2]}{(E[S])^2} = 1 + a + \sqrt{a^2 + 4a}$$

then $\rho^U = 0$, otherwise, $\rho^U < 0$.

We conclude that, for a fixed value of $E[S]$ and the other model parameters, increasing the variance of S from 0 (constant S) to a very large variance allows us to pass through the

assortative regime, to neutral (no assortativity), and then to disassortative. We can also observe that increasing the variance (i.e. $E[S^2]$) also increases R^U and R^P , respectively (see (3.18)), which thus might cross the critical value of 1, leading to percolation. This second effect, that disassortativity facilitates percolation, was also reported in Newman (2002) for a different model.

4. Discussion

In the present paper we studied properties of a stochastic dynamic network in which nodes are born and die randomly in time, and while alive, pairs of nodes are connected and disconnected by edges randomly in time. The rate of connecting to other nodes depended on the social index, given at birth, of the node. For this model, we derived limiting properties of a snapshot of the network after a long time, assuming that the node population grew large.

The three main results were: the type distribution of neighbour nodes (Section 3.1), a criterion for when phase transition occurs, above which the network has a giant component (Section 3.2), and an expression for the degree correlation of connected edges (Section 3.3). One of the most interesting features of the model is that the U-version of the model (in which neighbouring nodes are selected uniformly among all living nodes) may have either positive or negative degree correlation depending on the numerical values of the model parameters (birth and death rates of nodes and edges) and the social index distribution.

Together with previous results of the model (see Britton and Lindholm (2010)), the most important limiting local properties of the network are known, as well as whether or not the network has a giant component. Other global properties, such as the diameter of the network, remains to be analysed. An interesting problem would be to study the limiting properties of some process taking place ‘on’ the network. For example, can an epidemic, forest fire, or similar persist on the network forever or will it die out? What happens if a lightning process kills randomly selected nodes and all their neighbours at a constant rate?

The model can of course also be generalized in several ways to make it more applicable to certain situations. Nodes could be of different types with different attachment rates depending on the types of node in question, for example male or female if mimicking a sexual network. Similarly, edges could be of different types, perhaps reflecting the ‘degree’ of acquaintance which, for example, could affect the transmission probability of an epidemic taking place on the network. Still, the current model does capture some important properties of a real-world network in that the nodes as well as the edges are created and cease to exist. Also, depending on the model parameters, the model is quite flexible in producing different degree distributions and degree correlations (the network has no clustering asymptotically).

Acknowledgements

T. B. is grateful to The Bank of Sweden Tercentenary Foundation. Part of this work was done while the authors were visiting Institut Mittag-Leffler to which we are grateful. We also thank Pieter Trapman for fruitful discussions.

References

- BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509–512.
- BARABÁSI, A.-L., ALBERT, R. AND JEONG, H. (1999). Mean-field theory for scale-free random networks. *Physica A*, **272**, 173–187.
- BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31**, 3–122.

- BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2011). Sparse random graphs with clustering. *Random Structures Algorithms*, **38**, 269–323.
- BRITTON, T. AND LINDHOLM, M. (2010). Dynamic random networks in dynamic populations. *J. Statist. Phys.* **139**, 518–535.
- CALLAWAY, D. S. *et al.* (2001). Are randomly grown graphs really random? *Phys. Rev. E* **64**, 041902, 7 pp.
- HACCOU, P., JAGERS, P. AND VATUTIN, V. A. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.
- MALYSHEV, V. A. (1998). Random graphs and grammars on graphs. *Discrete Math. Appl.* **8**, 247–262.
- NEWMAN, M. E. J. (2002). Assortative Mixing in Networks. *Phys. Rev. Lett.* **89**, 208701, 4 pp.
- TUROVA, T. S. (2002). Dynamical random graphs with memory. *Phys. Rev. E* **65**, 066102, 9 pp.
- TUROVA, T. S. (2003). Long paths and cycles in dynamical graphs. *J. Statist. Phys.* **110**, 385–417.
- TUROVA, T. S. (2006). Phase transitions in dynamical random graphs. *J. Statist. Phys.* **123**, 1007–1032.
- TUROVA, T. S. (2007). Continuity of the percolation threshold in randomly grown graphs. *Electron. J. Probab.* **12**, 1036–1047.