Mathematical Statistics
Stockholm University

Dynamic random networks in dynamic populations

Tom Britton
Mathias Lindholm

Research Report 2009:4

Dynamic random networks in dynamic populations

Tom Britton
Mathias Lindholm *

May 2009

Abstract

We consider a random network evolving in continuous time in which new nodes are born and old may die, and where undirected edges between nodes are created randomly and may also disappear. The node population is Markovian and so is the creation and deletion of edges, given the node population. Each node is equipped with a random social index and the intensity at which a node creates new edges is proportional to the social index, and the neighbour is either chosen uniformly or proportional to the its social index in a modification of the model. We derive properties of the network as time and the node population tends to infinity. The degree distribution is shown to have a mixed Poisson distribution which may exhibit a heavy tail if the social index distribution has a heavy tail.

Key words: Random networks, dynamic networks, birth and death process, mixed Poisson distribution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: tomb@math.su.se, lindholm@math.su.se.

Dynamic random networks in dynamic populations

Tom Britton, Stockholm University[*]
Mathias Lindholm, Stockholm University[†]

May 13, 2009

Abstract

We consider a random network evolving in continuous time in which new nodes are born and old may die, and where undirected edges between nodes are created randomly and may also disappear. The node population is Markovian and so is the creation and deletion of edges, given the node population. Each node is equipped with a random social index and the intensity at which a node creates new edges is proportional to the social index, and the neighbour is either chosen uniformly or proportional to the its social index in a modification of the model. We derive properties of the network as time and the node population tends to infinity. The degree distribution is shown to have a mixed Poisson distribution which may exhibit a heavy tail if the social index distribution has a heavy tail.

Keywords: Random networks, dynamic networks, birth and death process, mixed Poisson distribution.

1 Introduction

With the increasing interest of analysing more and more complex networks the need for suitable network models increases as well. For static network models the pioneering Erdős-Rényi random graph model has been generalised successful in many ways, most recently in [3, 4]. As for random graph models having time dynamics built in to them, the perhaps most well-known time dynamic model is the preferential attachment model introduced by Barabási and Albert in 1999 [1], which was elaborated on in [2]. This model is a model for growing networks and it evolves in discrete time. As it was formulated, at each time step a node is born with a number of edges which then are attached to the older nodes with a probability proportional to their degrees. One of the key features of this model is that it generates asymptotically heavy-tailed degree distributions, which is a feature often seen

[*]Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden. *E-mail*: tom.britton@math.su.se

[†]Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden. *E-mail*: lindholm@math.su.se

in empirical networks. Since the model was formulated in 1999 various generalisations have been considered where e.g. edges as well as nodes may die, see e.g. [5, 6, 8].

In the present paper we consider a natural continuous time network model (defined in Section 2), where both nodes and edges are born according to a Markovian mechanism. The node population is modelled by a linear birth and death process assumed to be super-critical. Nodes are born without edges and at the time of birth, each node is assigned a (random) "social index". This social index affects the rate with which the node "creates" new edges. Once an edge is created, the other end of the edge is attached to another node: in our original model the other node is picked uniformly at random, but we also treat the case where the other node is selected with a probability being proportional to the social index of the node. Edges disappear at constant rate.

In Section 3 and 4 we derive asymptotic properties of the network. On the part of the sample space where the node population grows beyond all limits we derive the asymptotic degree distribution and also show that, asymptotically, there will be no clustering or degree correlation. The main result states that the asymptotic degree distribution is mixed Poisson, which may have a heavy tail if the social index distribution does. The probability distribution is not explicit, but we derive explicit expressions for the mean and variance. We give some numerical examples of the asymptotic degree distribution for a number of different social index distributions in Section. In Section 5 the asymptotic findings are also compared with simulated networks, where the model is simulated until the node population has reached a pre-specified number for the first time. For all studied social-index distributions, the empirical degree distribution is close to the asymptotic degree distribution already when simulations are stopped once thousand nodes are alive.

In Section 6 we discuss possible extensions of the model that may give clustering and degree correlation.

2 The Markovian random network in a Markovian dynamic population

Below we define a model which we denote the Markovian random network in a Markovian dynamic population. 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. The model is dynamic in the double sense that nodes are born and may die, and the same applies to edges.

In sociological applications, which is what we have in mind, nodes correspond to individuals, edges to some type of relation (e.g. friendship), and neighbours to friends, but in what follows we use the notation "nodes", "edges" and "neighbours".

2.1 The model

The model consists of two parts: births and deaths of nodes and, given the node population, births and deaths of edges.

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 $\text{Exp}(\mu)$ distributed time with mean $1/\mu$, 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 super critical branching process. Additional to this, each node i is at birth given a random "social index" S_i having distribution F_S on $\mathbb{R}^+$, independent and identically distributed for different nodes. We will throughout assume that S has finite mean $\mathsf{E}[S]$.

The model for births and deaths of edges is Markovian given the node population and the social indices. At birth nodes are isolated, i.e. have no neighbours. During the life of a node having social index s, new edges are created at rate αs, each time the neighbour is chosen uniformly among all living nodes (below we also study a different model for choosing neighbours). 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.

2.2   A modification of the model

As mentioned in the previous paragraph we also study a modification of the original model defined above. This new model is the same as the original except in the way that neighbours are chosen: instead of choosing neighbours uniformly the modified model assumes that neighbours are chosen with probabilities proportional to their social index. This modified model is inspired by the preferential attachment model [1], still being different in that here the probability of receiving edges is determined at birth whereas it is determined by random events during life in the preferential attachment model.

2.3   Comments on the model

The model has four parameters: the birth and death rates of nodes, λ and μ respectively, and the death rate β of edges and α which is related to the birth rate of new edges. Beside these four parameters there is the distribution F_S for the social indices $\{S_i\}$ of nodes.

Note that the model allows for multiple edges and self-loops, but in the Appendix we show that the *proportion* of such edges is asymptotically negligible if $\mathsf{E}[S] < \infty$ for the original model, and if $\mathsf{E}[S^2] < \infty$ for the modified model. As a consequence the network will then have identical properties if loops and multiple edges are ignored or not allowed.

Some submodels are of particular interest. One is 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 possibly 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. The case where $\mu = 0$ and $S \equiv 1$ has been studied by Turova in [13, 14, 15, 16] who derive more results for this submodel.

2.4   Population properties of the model

The population of nodes is a simple Markovian super-critical branching process with individual birth rate λ and death rate μ (we have assumed that $\lambda > \mu$. Properties of this type of model have been studied extensively, see e.g. [10]. For example, it is well-known that the population may either die out or else it will grow at the exponential rate

$$Y(t) \sim e^{t(\lambda-\mu)}.$$

In what follows we will only consider the part of the sample space where the population grows $B := \{\omega; Y(t) \to \infty\}$, the "explosion event". On this part of the sample space the population will have what is called a *stable age distribution*. As t tends to infinity (implying that the population size also tends to infinity) the age distribution of living nodes will converge to the exponential distribution having mean $1/\lambda$, i.e. a node picked uniformly at random at $t \ (\to \infty)$ will have age distribution $\text{Exp}(\lambda)$, see e.g. Section 3.4 in [10].

3   The degree distribution

We are interested in the asymptotic degree distribution, i.e. the number of neighbours of a randomly selected living node, selected at $t \to \infty$ when the node population tends to infinity and the population attains the stable age distribution. We let D be a random variable having this distribution, specified by $\{p_k; k \geq 0\}$, where $p_k = \mathsf{P}(D = k)$.

3.1   The degree distribution conditional on age and social index

Label the selected node i. Two properties of the selected node affects the distribution of the number of neighbours it has: the social index S_i of the node and the age A_i of the node. The social index S_i has distribution F_S and the age A_i has the stable age distribution $\text{Exp}(\lambda)$. Condition on the social index of the node: $S_i = s$. The node gets new neighbours in two ways: either it creates new neighbours itself at rate αs, or it obtains new neighbours from other nodes throwing edges to this particular node. At time t there are $Y(t)$ nodes (labelled $1, \ldots, Y(t)$ say) and hence the over-all rate at which other nodes throw out edges equals $\alpha \sum_{j=1}^{Y(t)} S_j$.

Each time, node i receives such an edge with probability $1/Y(t)$, so the rate at which i receives edges equals

$$\alpha \sum_{j=1}^{Y(t)} S_j \frac{1}{Y(t)}. \tag{1}$$

But, since $Y(t) \to \infty$ as $t \to \infty$ (on the explosion event B), it follows that $Y^{-1}(t) \sum_{j=1}^{Y(t)} S_j \to \mathsf{E}[S]$ as $t \to \infty$ by the law of large numbers. From this it follows that, asymptotically, node i attains edges from other nodes at rate $\alpha \mathsf{E}[S]$ and it creates new edges at rate αs implying that the overall rate equals

$$\alpha(s + \mathsf{E}[S]).$$

4

An edge can be removed for two reasons: either the edge "dies", or one of the neighbouring nodes "dies". During the life of node i, each edge is hence removed if the edge dies or if the neighbour connected by the edge dies. The rate at which a given edge is removed hence equals $\beta + \mu$, and all edges are removed independently at the same rate.

The above reasoning implies that the number of neighbours of a fixed living node having social index s can be modelled by a birth and death process (e.g. [12]) with (constant) birth rate $\lambda_n = \alpha(s + \mathsf{E}[S])$ and (linear) death rate $\mu_n = n(\beta + \mu)$. If $X_s(a)$ denotes the number of neighbours of our node (with social index s) at age a, then, during the life of the node, $X_s(u)$ is hence a birth and death process with the rates specified above and with initial condition $X_s(0) = 0$. We now claim that this implies that $X_s(a)$ is Poisson distributed for the following reason. The number of edges that connect to the node up until a is Poisson distributed with mean $\alpha(s + \mathsf{E}[S])a$ (new edges connect at constant rate $\alpha(s + \mathsf{E}[S])$). Each such edge will remain at time a independently of other edges, so the remaining number of edges will be a thinned Poisson distribution which is also a Poisson distribution. To compute the probability for this event (for an edge that has attached during (o, a) to remain at time a) we condition on the time that the edge attaches, a time which is uniformly distributed between 0 and a. The probability for the edge to remain at a if attached at x is $e^{-(\beta+\mu)(a-x)}$ implying that the unconditional probability equals

$$\mathsf{P}(\text{remain at } a) = \int_0^a e^{-(\beta+\mu)(a-x)} \frac{1}{a} dx = \frac{1 - e^{-(\beta+\mu)a}}{a(\beta + \mu)}.$$

To conclude, we have shown that

$$X_s(a) \sim \mathrm{Po}\left(\frac{\alpha(s + \mathsf{E}[S])\left(1 - e^{-(\beta+\mu)a}\right)}{\beta + \mu}\right). \tag{2}$$

(The result that a birth and death process with constant birth rates and linear death rates starting at 0 has Poisson distributed number of individuals at time a is not new, see e.g. Feller [9, p. 481] who obtained the result through solving a certain partial differential equation.)

3.2   The unconditional degree distribution

In the task of obtaining the unconditional degree distribution, recall that the age of a node picked uniformly at random follows the stable age distribution $A \sim \mathrm{Exp}(\lambda)$. Unconditionally, the number of edges our node is connected to $X_s(A)$, hence has a mixed Poisson distribution where A is exponentially distributed and where the degree distribution given $A = a$ is the one given in (2). Finally, removing also the conditioning on $S = s$ we obtain the community degree distribution $D = X_S(A)$, which hence also has a mixed Poisson distribution:

$$X_S(A) \sim \mathrm{MixPo}\left(\frac{\alpha(S + \mathsf{E}[S])\left(1 - e^{-(\beta+\mu)A}\right)}{\beta + \mu}\right), \tag{3}$$

where $A \sim \mathrm{Exp}(\lambda)$ and $S \sim F_S$ are independent. This simply means that, conditional on $A = a$ and $S = s$, $X_s(a)$ has the Poisson distribution defined in (2). This specifies the

degree distribution $\{p_k\}$ completely. The probabilities $p_k = \mathsf{P}(D = k) = \mathsf{P}(X_S(A) = k)$ can be obtained by integrating the Poisson probability over the age- and social-index-distribution. A description of the (unconditional) degree distribution may also be given using the probability generating function $g_{X_S(A)}(u) = \mathsf{E}\left[u^{X_S(A)}\right]$. Unfortunately, neither of these ways lead to any closed form expressions, in Section 5 we give some numerical illustrations.

It is worth mentioning that the effect of β and μ on the whole distribution of $X_S(A)$, and hence also on its mean and variance, is only through their sum $\beta + \mu$. Further, looking at the conditional degree distribution (2) and remembering that A is exponential with parameter μ, we see that the tail of the distribution of $X_S(A)$ will have the same form as the tail of S. If for example S has a power-law distribution, so will $X_S(A)$.

3.3   Moments of the degree distribution

We now derive explicit expression for the mean and variance of $D = X_S(A)$. To simplify notation we define $\theta = \beta + \mu$ and $\mu_S = \mathsf{E}[S]$. From Equation (2) we know that the conditional distribution of $X_s(a)$ is Poisson. It follows that

$$\mathsf{E}[X_s(a)] = \mathsf{V}[X_s(a)] = \frac{\alpha(s + \mu_S)(1 - e^{-\theta a})}{\theta}.$$

Removing the conditioning on A, which is distributed according to $\mathrm{Exp}(\lambda)$, we get

$$\mathsf{E}[X_s(A)] = \int_0^\infty \frac{\alpha(s + \mathsf{E}[S])(1 - e^{-\theta a})}{\theta} \lambda e^{-\lambda a} da = \frac{\alpha(s + \mu_S)}{\lambda + \theta}.$$

From this we get the unconditional mean:

$$\mathsf{E}[X_S(A)] = \frac{2\alpha}{\lambda + \theta}\mu_S. \tag{4}$$

For the variance we recall that if $Z \sim \mathrm{Po}(\tau)$ it holds that $\mathsf{E}[Z^2] = \tau + \tau^2$. As a consequence, removing the conditioning on the age A, we get

$$\mathsf{E}[X_s^2(A)] = \int_0^\infty \mathsf{E}[X_s^2(a)] \lambda e^{-\lambda a} da$$

$$= \int_0^\infty \left(\frac{\alpha(s + \mu_S)(1 - e^{-(\theta)a})}{\theta} + \left(\frac{\alpha(s + \mu_S)(1 - e^{-\theta a})}{\theta} \right)^2 \right) \lambda e^{-\lambda a} da$$

$$= \frac{\alpha(s + \mu_S)}{\lambda + \theta} + \left(\frac{\alpha(s + \mu_S)}{\theta} \right)^2 \left[1 - 2\frac{\lambda}{\lambda + \theta} + \frac{\lambda}{\lambda + 2\theta} \right].$$

After removing the condition on the social index s and some simplifications we get

$$\mathsf{E}[X_S^2(A)] = \frac{2\alpha\mu_S}{\lambda + \theta} + \frac{2\alpha^2 \mathsf{E}[(S + \mu_S)^2]}{(\lambda + \theta)(\lambda + 2\theta)}.$$

Finally, using that $\mathsf{V}[X_S(A)] = \mathsf{E}[X_S(A)^2] - (\mathsf{E}[X_S(A)])^2$ and simplifying the expression we get the following formula for the variance:

$$\mathsf{V}[X_S(A)] = \frac{2\alpha}{\lambda + \theta}\mu_S + \frac{4\lambda\alpha^2}{(\lambda + \theta)^2(\lambda + 2\theta)}\mu_S^2 + \frac{2\alpha^2}{(\lambda + \theta)(\lambda + 2\theta)}\mathsf{V}[S]. \tag{5}$$

From (4) we see that the mean degree distribution $\mathsf{E}[D] = \mathsf{E}[X_S(A)]$ is proportional to the mean social index μ_S and to α, the birth rate of edges. Further, the mean degree distribution decreases with $\theta = \beta + \mu$, where β is the rate at which edges are removed and μ is the rate at which nodes (and there adjacent edges) are removed. More surprising is perhaps the fact that $\mathsf{E}[D] = \mathsf{E}[X_S(A)]$ *decreases* with λ, the birth rate of nodes. An explanation to this is that a high λ implies a young stable age distribution, which in turn implies that nodes will tend to have low degree.

As for the variance $\mathsf{V}[X_S(A)]$ we see from (5) that it is increasing in the birth rate of edges, α, and decreasing in both the death rate of edges and nodes, β and μ respectively (recall that $\theta = \beta + \mu$). How the variance depends on the birth rate of nodes could go either way, or not be monotone, depending on the relation between $\mu_S = \mathsf{E}[S]$ and $\mathsf{V}[S]$. Finally, $\mathsf{V}[X_S(A)]$ is increasing *both* in $\mu_S = \mathsf{E}[S]$ and $\mathsf{V}[S]$, and the variance is infinite if S has infinite variance.

3.4 The degree distribution for the modified model

Recall that the difference between the modified model and the original is that, in the modified model, a created edge (by some node) is attached to a given node with a probability proportional to the social index of that node. In the same notation as above and conditioning on $\{S_j\}$, this means that an edge created by i will connect to node j with probability

$$\frac{S_j}{\sum_{k=1}^{Y(t)} S_k}. \tag{6}$$

This modification will not change the over-all rate that node i creates new edges; it is still αS_i. However, the rate at node i receives edges is now changed, since the probability of receiving an edge now equals $S_i / \sum_{k=1}^{Y(t)} S_k$ as opposed to $1/Y(t)$ in the original model. The over-all rate with which node i receives edges is hence changed to

$$\alpha \sum_{j=1}^{Y(t)} S_j \frac{S_i}{\sum_{k=1}^{Y(t)} S_k} = \alpha S_i, \tag{7}$$

(cf. Equation 3.1 for the original model), i.e. the same rate as the rate of creating edges. The over-all rate of getting new neighbours is hence $2\alpha S_i$.

Beside this modification the same analysis applies to the modified model. Conditional on the social index being s and the age being a, the number of neighbours node i has at age u ($0 \le u \le a$) is described by a birth and death process with constant birth rate $\lambda_n 2\alpha s$ and linear death rate $\mu_n = n(\beta + \mu) =: n\theta$. The same analysis as for the original model goes through, thus showing that the conditional degree distribution is

$$X_s(a) \sim \mathrm{Po}\left(\frac{2\alpha s\left(1 - \mathrm{e}^{-\theta a}\right)}{\theta}\right). \tag{8}$$

Similarly, the unconditional degree distribution is mixed Poisson:

$$X_S(A) \sim \mathrm{MixPo}\left(\frac{2\alpha S\left(1 - e^{-\theta A}\right)}{\theta}\right), \tag{9}$$

7

where $A \sim \text{Exp}(\lambda)$ and $S \sim F_S$ are independent. From this it can be shown that the mean of the modified model is the same as for the original model (recall that $\theta := \beta + \mu$)

$$\mathsf{E}[X_S(A)] = \frac{2\alpha}{\lambda + \theta}\mu_S, \tag{10}$$

but having a different variance:

$$\mathsf{V}[X_S(A)] = \frac{2\alpha}{\lambda + \theta}\mu_S + \frac{4\lambda\alpha^2}{(\lambda + \theta)^2(\lambda + 2\theta)}\mu_S^2 + \frac{8\alpha^2}{(\lambda + \theta)(\lambda + 2\theta)}\mathsf{V}[S]. \tag{11}$$

If we compare this variance with the variance for the original model (5) we see that the first two terms are identical and the last term is a factor 4 larger here as compared with (5). This hence implies that the degree distribution for the modified model has the same mean but a strictly larger variance (unless S is nonrandom, $S \equiv k$, when the two models are in fact identical). It should not come as a surprise that the modified model has larger variance: by selecting neighbours with probabilities proportional to their social index, the effect will be that nodes with high social index will get many edges and nodes with low social index will get less, thus making the degree distribution more skew.

4 Other network properties

In the previous section it was shown that the asymptotic degree distribution of our model (and the modified model) is quite flexible by choosing an appropriate distribution for the social index. Two other features of interest in many applications of network models (e.g. [11]) are *clustering*, measuring the frequency of triangles, and the *degree correlation*, measuring the correlation of node degrees of connected nodes, a positive degree correlation being termed an assortative network and and negatively correlated networks being called disassortative. We refer to Newman [11] for exact definitions of these quantities, but the essence of them are that the clustering coefficient c equals the probability that two nodes that are connected by a third node, are connected directly, and the degree correlation ρ is the correlation between the degrees of the nodes of a randomly selected edge.

Unfortunately, the present model (or the modified model) cannot produce networks having positive clustering nor any degree correlation asymptotically. That $c = 0$ asymptotically follows from the fact that each edges is connected between edges independently of other edges present in the network, implying that there is no increased chance that edges between nodes connected by a third node will appear. That $\rho = 0$ follows because each node that creates an edge, selects its neighbour edge in the same way independent of its social index and present degree.

In the Section 6 we discuss some possible extensions which might allow for positive clustering ($c > 0$) and degree correlation different from 0 ($\rho \neq 0$).

5 Examples and illustrations

In the present section we compute the degree distribution numerically for some particular parameter choices. This distribution is of asymptotic nature and it is hence of practical

interest to try and quantify how large t must be in order for the asymptotics to be valid. In order to do this we compare simulations of the network model and their associated empirical degree distributions with the numerical evaluations of the theoretical degree distribution. We also compare the original model, see Section 2.1, with the extended model, see Section 2.2.

Throughout this section we will consider the following set of parameter values: $\lambda = 1.55/80$, $\mu = 1/80$, $\alpha = 1$ and $\beta = 1/2$. These parameter values can be given some sociological justification. The choice of $\mu = 1/80$ corresponds to that the average life-length of an individual is 80 years, and the choice of λ gives us a doubling time for the node population which is approximately 100 years. The choice of α is however somewhat more arbitrary but it corresponds to that each individual makes roughly one new friend per year, and the value of β corresponds to that each such friendship on average lasts for two years. Moreover, we have restricted all examples to the following distributions for the social indices:

(i) $S \equiv 1$,

(ii) $S \sim \mathrm{Exp}(1)$,

(iii) $S \sim \Gamma(1/2, 1/2)$,

(iv) $S \sim \mathrm{Pareto}(1/2, 2)$.

In common for all these distributions are that $\mu_S = 1$. Further, they are ordered in increasing variance with the Pareto distribution having infinite variance.

5.1 Numerical evaluation of the degree distribution

The degree distributions for the original model and the modified model are both of mixed Poisson type. We have already seen that the two distributions have the same mean, but the variance is greater for the latter model. In Figure 1 and 2 we have used the above mentioned parameter values and evaluated the theoretical degree distribution numerically for each of the four different choices of distributions of social indices $(i) - (iv)$.

For these particular parameter values we see that the tails of the degree distribution for both models are heaviest when the social indices follows the Pareto distribution. It is however worth noting that in the original model the probability of having degrees of ≈ 30 is still higher when S is gamma distributed. For the modified model this seems to be true for degrees up to ≈ 50, but due to instability of the numerical integration for degrees larger than ≈ 30 these results should be treated with some caution.

5.2 Simulations of dynamic network

In our simulations we let the network develop until the first time point when there are 1000 nodes alive starting from a single node with no edges. This corresponds to that the node population has evolved for approximately 1000 years. This procedure has then been repeated 50 times for each of the distributions $(i) - (iv)$ and with the above discussed
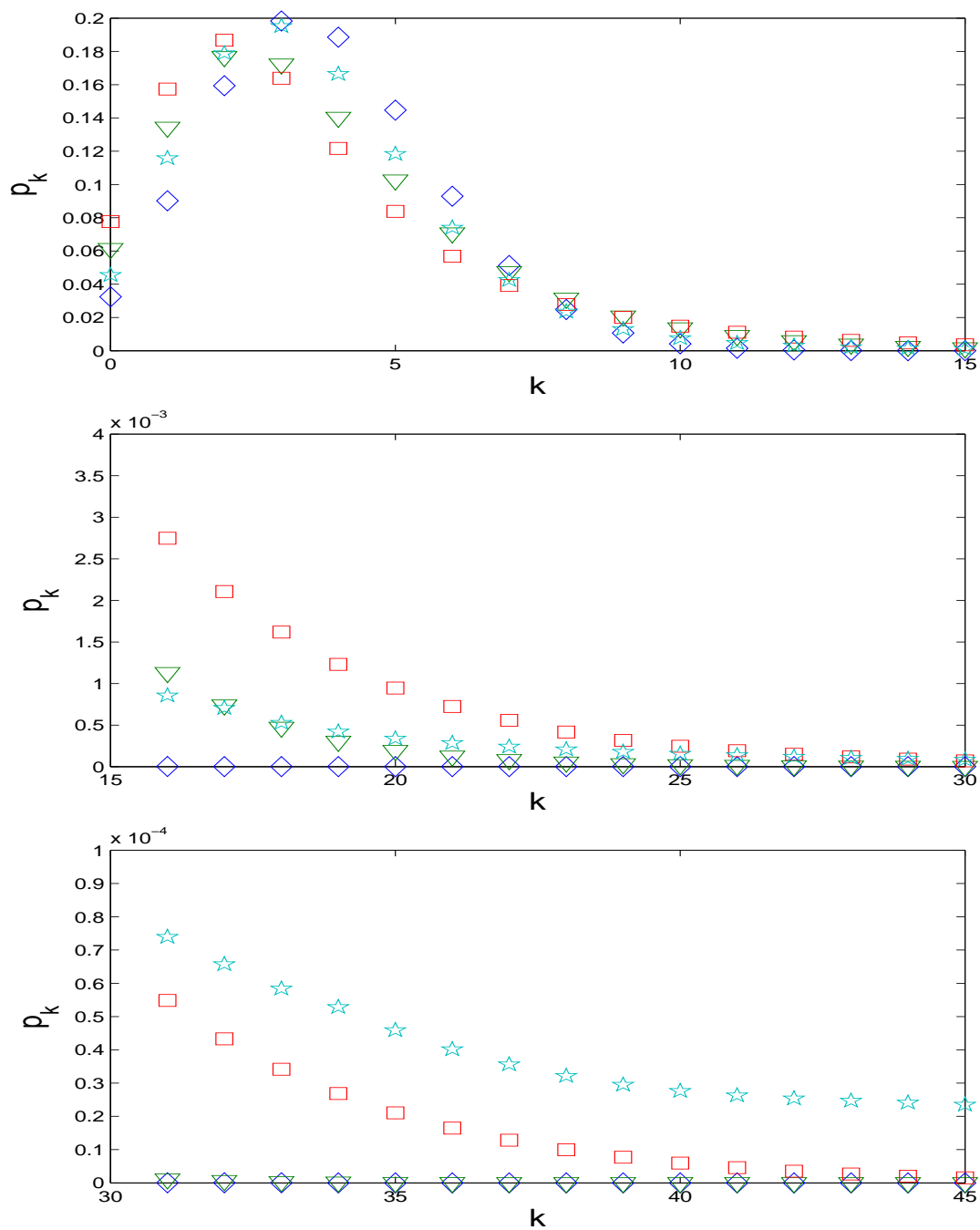
9

Figure 1: The degree distribution $p_k = P(X_S(A) = k)$ for the original model (numerically evaluated) for four different choices of distributions for the social indices: $S \equiv 1$ ("diamonds"), $S \sim \mathrm{Exp}(1)$ ("triangles"), $S \sim \Gamma(1/2, 1/2)$ ("squares") and $S \sim \mathrm{Pareto}(1/2, 2)$ ("stars"). The top figure are for small k, the middle for intermediate k and the bottom for large k (note the different scales on the y-axis).
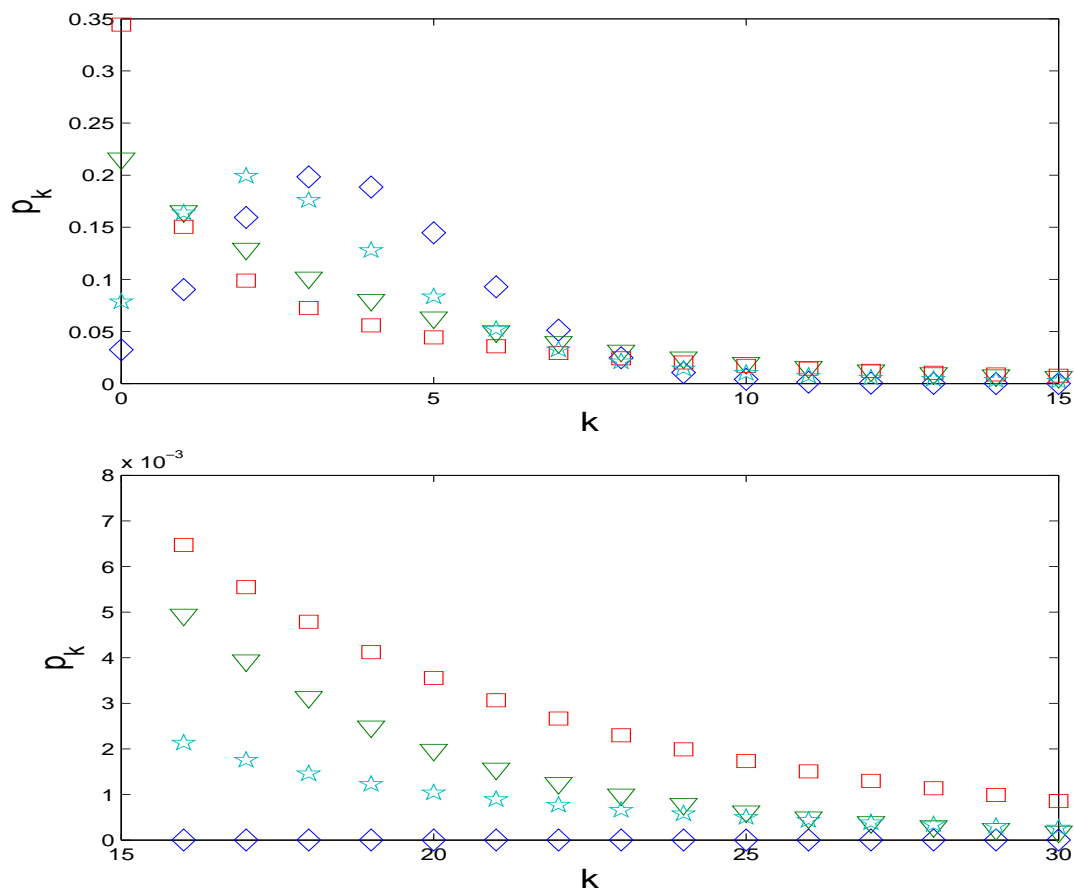
Figure 2: The degree distribution $p_k = P(X_S(A) = k)$ for the modified model (numerically evaluated) for four different choices of distributions for the social indices: $S \equiv 1$ ("diamonds"), $S \sim \text{Exp}(1)$ ("triangles"), $S \sim \Gamma(1/2, 1/2)$ ("squares") and $S \sim \text{Pareto}(1/2, 2)$ ("stars"). The top figure are for small k and the bottom figure are for large k (note the different scales on the y-axis).

Table 1: Empirical mean and variance for the original model

Distribution			$Y(t) = 1000$	$Y(t) = \infty$
$S \equiv 1$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7437	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	4.0471	4.0226
$S \sim \mathrm{Exp}(1)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7596	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	7.6113	7.6231
$S \sim \Gamma(1/2, 2)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7518	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	11.1136	11.2236
$S \sim \mathrm{Pareto}(1/2, 2)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7165	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	12.1321	∞

Table 2: Empirical mean and variance for the modified model

Distribution			$Y(t) = 1000$	$Y(t) = \infty$
$S \equiv 1$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7560	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	4.0364	4.0226
$S \sim \mathrm{Exp}(1)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7637	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	18.4698	18.4246
$S \sim \Gamma(1/2, 2)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.7757	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	34.4582	32.8267
$S \sim \mathrm{Pareto}(1/2, 2)$,	$\mathsf{E}[X_S(A)]$	$\approx$	3.6579	3.7603
	$\mathsf{V}(X_S(A))$	$\approx$	173.2699	∞

parameter values. The resulting empirical means and variances for both models are tabulated in Table 1 and 2. It is worth noting that the empirical means and variances are in good agreement with the theoretical ones except for the situation when $S \sim \mathrm{Pareto}$. In this situation the theoretical variance of S, and hence the theoretical variance of the degree distribution, is infinite. But, as can be seen from Table 1, the difference between the variance when S follows a gamma distribution and when S follows a Pareto distribution is marginal for the original model. This is however *not* the case for the modified model. This behaviour seems to be an artefact of the finiteness of the node population and the weight of the tail of the degree distribution.

In Figure 3 we have plotted the empirical degree distribution for one simulated realization with the theoretical distribution for the case $S \sim \mathrm{Pareto}(1/2, 2)$ (other distributions show similar result), both for the original and the modified model. We can again see that the agreement between the simulations and the numerically evaluated theoretical results are good for both models.

In the Appendix the number of multiple edges and loops is analysed. It is shown that the fraction of such edges will be small if $\mathsf{E}[S] < \infty$ in the original model, and if $\mathsf{E}[S^2] < \infty$ in the modified model. These results are confirmed by simulations where the number of loops and multiple edges is small in all cases except for the modified model when $S \sim \mathrm{Pareto}(1/2, 2)$. To illustrate this we picked one "typical" simulation of both the original model and the modified model for the case that $S \sim \mathrm{Pareto}(1/2, 2)$. The original model had 4 loops and 24 multiple edges out of 1738 edges. The fraction of multiple edges
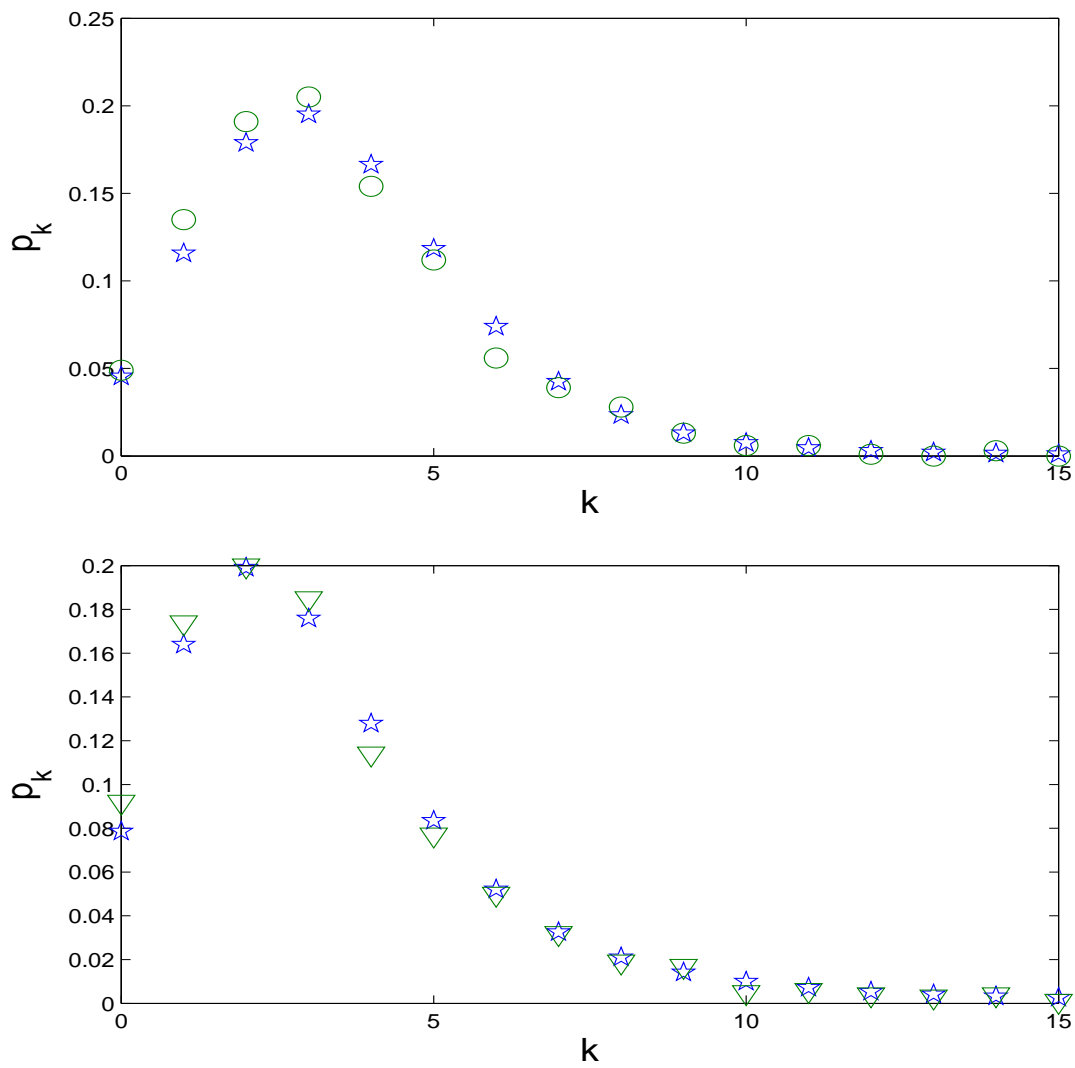
Figure 3: Comparison between the empirical degree distribution ("circles") and the corresponding theoretically evaluated one ("stars") for $S \sim \text{Pareto}(1/2, 2)$; top figure shows the original model and the bottom figure shows the modified model.

13

and loops is hence 1.6%. The modified model has a markedly higher fraction: there were 50 loops and 366 multiple edges out of 1908 edges in total, giving 21.8% of the edges being multiple or loops. This hence agrees well with the asymptotic results since $\mathsf{E}[S] < \infty$ but $\mathsf{E}[S^2] = \infty$ for the Pareto$(1/2, 2)$-distribution. One can also remark that out of the 50 loops 48 of these were multiple loops connected to the same node, a node with degree 306. The other S-distributions were also examined and showed small fractions of loops and multiple edges, as suggested by the theoretical results.

6   Discussion

The present model is a simple continuous time Markovian network model in which nodes give birth to new nodes and old nodes die, and in which edges between existing nodes are created and disappear. A special feature of the model is that, at birth, nodes are given i.i.d. social indices, and by choosing an appropriate distribution for these social indices, the asymptotic degree is rather flexible.

It would be interesting to extend the present model to allow also for positive clustering and/or having degree correlation different from 0. One possibility for obtaining positive clustering would be to let newborn individuals (i.e. nodes) "inherit" edges from the mother as well as having a node from the mother, thus creating triangles at birth. This is a variation of a mechanism described in [7]. A perhaps more elegant way of creating a positive fraction of triangles would be to add an extra rate for creating edges between nodes connected by a third node. However, we believe this second model is harder to analyse. As regards to degree correlation this can be achieved by letting the *probability* for a node to select a given neighbour to depend not only on the social index of the neighbour, but also on the social index of the node itself. If this probability is given by $p(s, s')$ for some kernel function, a kernel with high probability mass near the diagonal should produce positive degree correlation (assortativity) whereas high probability mass on the off-diagonal should result in negative degree correlation (disassortativity). This might be done using similar arguments as in Bollobás et al. [3] and [4]

As mentioned in the Section 2.3 a special case of the original formulation of this model given in Section 2.1 has been treated in Turova [13, 14, 15, 16]. The model treated by Turova corresponds to that μ in our model is set to 0 and that $S \equiv 1$. Among other things, Turova analyse the size of the giant component and derive conditions for when it may emerge. In [16] the methods of [3] are discussed, and a continuation of the present work may hence be to see to what extent the methods of Turova and [3] are applicable to the current situation.

Acknowledgements

Appendix: Multiple edges and loops

As previously mentioned the degree of a node, and hence also the degree distribution, may include multiple edges and/or loops. We now show that the fraction of such edges is negligible under certain conditions on the social index distribution F_S, thus making the degree distribution *excluding* such edges, which is a more reasonable quantity, asymptotically the same. As a consequence, if the model is modified such that multiple edges and loops are not allowed, this model will have the same asymptotic degree distribution.

The original model

We start with the original model in which neighbours are chosen uniformly. Pick t and consider the next edge to appear in the dynamic network. We now show that the probability that this edge will be a multiple edge or a loop converges to 0 as t tends to infinity under a certain moment condition. This implies that the fraction of edges at $2t$ that are loops or multiple will tend to 0 in probability since nearly all edges at $2t$ will have been created after t, thus giving the desired result.

First we condition on $Y(t)$, the number of nodes alive at t, and their social indices $\{S_j\}$. Then we condition on the node i that creates the next edge, the probability of this event being equal to $S_i / \sum_j S_j$. The degree D_i of this node i has distribution $X_{S_i}(A)$ described in the beginning of Section 3.2, a mixed Poisson distribution. This distribution is stochastically smaller than a Poisson distribution with mean parameter $\alpha(S_i + \mathsf{E}[S])/(\beta + \mu)$, since the random component $1 - \exp(-(\beta + \mu)A)$ is always smaller than 1. Given the degree D_i, the probability that the next edge is a multiple edge or a loop is at most $(D_i + 1)/Y(t)$ (with equality if there are no loops or multiple edges from i). If M denotes the event that the next edge is a multiple edge or loop this gives the following:

$$\mathsf{P}(M|\{S_j\}, \{D_j\}, Y(t)) = \sum_{i=1}^{Y(t)} \mathsf{P}(M|\{S_j\}, \{D_j\}, Y(t), i \text{ selected}) \mathsf{P}(i \text{ selected}|\{S_j\}, \{D_j\}, Y(t))$$

$$\leq \sum_{i=1}^{Y(t)} \frac{D_i + 1}{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j}.$$

By taking expectations we then get that

$$\mathsf{P}(M) \leq \mathsf{E}\left[\sum_{i=1}^{Y(t)} \frac{D_i + 1}{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j}\right]$$

$$= \mathsf{E}\left[\mathsf{E}\left[\sum_{i=1}^{Y(t)} \frac{D_i + 1}{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j} \middle| \{S_j\}, Y(t)\right]\right]$$

$$\leq \mathsf{E}\left[\sum_{i=1}^{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j} \frac{\alpha(S_i + \mathsf{E}[S])/(\beta + \mu) + 1}{Y(t)}\right]$$

$$= \mathsf{E}\left[\frac{1}{Y(t)} \left(c_1 \frac{\sum_{i=1}^{Y(t)} S_i^2}{\sum_{j=1}^{Y(t)} S_j} + c_2\right)\right],$$

15

where c_1 and c_2 are the appropriate constants.

We now claim that this tends to 0 as t (and $Y(t)$) tends to infinity if $\mathsf{E}[S] < \infty$. To begin with the first term $c_2/Y(t)$ will clearly tend to 0, since $Y(t) \to \infty$ as $t \to \infty$. We rewrite the expression:

$$\mathsf{E}\left[\frac{1}{Y(t)} \frac{\sum_{i=1}^{Y(t)} S_i^2}{\sum_{j=1}^{Y(t)} S_j} \right] = \mathsf{E}\left[\frac{1}{Y(t)} \sum_{i=1}^{Y(t)} \mathsf{E}\left[\frac{S_i^2}{\sum_{j=1}^{Y(t)} S_j} \middle| Y(t) \right] \right] = \mathsf{E}\left[\frac{S_1^2}{\sum_{j=1}^{Y(t)} S_j} \right],$$

where the last equality is true since all terms in the sum to the left are identically distributed and hence have the same mean. If we introduce the sequence of random variables

$$Z_n := \frac{S_1^2}{\sum_{j=1}^{n} S_j} = \frac{S_1^2}{S_1 + \sum_{j=2}^{n} S_j}$$

we hence want to show that $\mathsf{E}[Z_{Y(t)}] \to 0$. But, since $Y(t) \to \infty$ as $t \to \infty$ we only need to show that $Z_n \to Z$ as $n \to \infty$ in order for $Z_{Y(t)} \to Z$ to hold. Thus, we want to show that $Z_n \to 0$ and then use dominated convergence to obtain the desired result. First we observe that $Z_1 = S_1$ so $\mathsf{E}[Z_1] = \mathsf{E}[S_1] = \mu_S < \infty$ by assumption. Second, since all S_j's are positive we have that Z_n is decreasing. Finally, since $S > 0$ $\mathsf{E}[S] > 0$ which in turn implies that $\sum_{j=2}^{n} S_j$ will be close to $n\mu_S$. This together with the fact that S_1 is bounded in probability (like any finite random variable) implies that Z_n will tend to 0 in probability, and we can hence use dominated convergence. This completes the proof that the fraction of edges that are loops or multiple edges is negligible for the original model if $\mathsf{E}[S] < \infty$.

The modified model

We now show the same type of result for the modified model, but now we also have to assume that $\mathsf{E}[S^2] < \infty$ for the result to be true. We can use the same type of reasoning as for the original model, but because new edges are attached to nodes proportionally to their social indices in the modified model we need to keep track of $\mathcal{V}_i$, the indices of node i's neighbours. It is important to note that $|\mathcal{V}_i| \leq_d D_i$, since D_i is the total number of edges (which may be multiple). Hence, the new edge created by i will be multiple if any of the vertices with indices in $\mathcal{V}_i$ is selected. Along the lines of the arguments for the original model we hence get

$$\mathsf{P}(M|\{S_j\}, \{\mathcal{V}_j\}, Y(t)) = \sum_{i=1}^{Y(t)} \mathsf{P}(M|\{S_j\}, \{\mathcal{V}_j\}, Y(t), i \text{ selected}) \mathsf{P}(i \text{ selected}|\{S_j\}, \{\mathcal{V}_j\}, Y(t))$$

$$= \sum_{i=1}^{Y(t)} \left(\frac{S_i}{\sum_{j=1}^{Y(t)} S_j} + \sum_{u \in \mathcal{V}_i} \frac{S_u}{\sum_{j=1}^{Y(t)} S_j} \right) \frac{S_i}{\sum_{j=1}^{Y(t)} S_j}$$

$$= \sum_{i=1}^{Y(t)} \frac{S_i^2}{\left(\sum_{j=1}^{Y(t)} S_j \right)^2} + \sum_{i=1}^{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j} \sum_{u \in \mathcal{V}_i} \frac{S_u}{\sum_{j=1}^{Y(t)} S_j}.$$

But, the social indices amongst the neighbours of node i will typically be larger than a randomly picked social index because neighbours are selected with a probability proportional to their social index. In fact, the probability that a neighbour node has social index (close to) s is proportional to $sf_S(s)$ (assuming S has a continuous distribution). This is true since the probability of selecting a *given* individual having social index s is proportional to s and the fraction of individuals having such social indices is proportional to $f_S(s)$. This implies that the neighbours in $\mathcal{V}_i$ will have independent social indices $\{\tilde{S}_{i,u}\}$, all with density $\tilde{f}_S$ defined as $\tilde{f}_S(s) = sf_S(s)/\int tf_S(t)dt$ (i.e. the size biased distribution of S). This together with that $|\mathcal{V}_i| \le D_i$ for all i gives us

$$\mathsf{P}(M) = \mathsf{E}\left[\sum_{i=1}^{Y(t)} \frac{S_i^2}{\left(\sum_{j=1}^{Y(t)} S_j\right)^2} + \sum_{i=1}^{Y(t)} \frac{S_i}{\sum_{j=1}^{Y(t)} S_j} \sum_{u \in \mathcal{V}_i} \frac{S_u}{\sum_{j=1}^{Y(t)} S_j}\right]$$

$$\le \mathsf{E}\left[\frac{\sum_{i=1}^{Y(t)} S_i^2}{\left(\sum_{j=1}^{Y(t)} S_j\right)^2}\right] + \mathsf{E}\left[\sum_{i=1}^{Y(t)} \frac{S_i \sum_{u=1}^{D_i} \tilde{S}_{i,u}}{\left(\sum_{j=1}^{Y(t)} S_j\right)^2}\right].$$

By using the independence between $\{\tilde{S}_{i,u}\}$ and the other random variables together with that $\mathsf{E}[D_i|S_i] = c_3 S_i + c_4$, for some constants c_3 and c_4, one can after some simplifications obtain

$$\mathsf{P}(M) \le \mathsf{E}\left[\frac{\sum_{i=1}^{Y(t)} S_i}{\left(\sum_{j=1}^{Y(t)} S_j\right)^2}\right]\left(1 + c_3\mathsf{E}[\tilde{S}]\right) + c_4\mathsf{E}[\tilde{S}]\mathsf{E}\left[\frac{1}{\sum_{j=1}^{Y(t)} S_j}\right].$$

By noting that the expressions inside of the first and last expectations are bounded from above by 1, a similar dominated convergence argument as that which was used for the original model can be applied, implying that $\mathsf{P}(M) \to 0$ as $t \to \infty$ if $\mathsf{E}[\tilde{S}] = \mathsf{E}[S^2]/\mu_S^2 < \infty$.

References

[1] BARABÁSI, A.-L., ALBERT, R.: Emergence of scaling in random networks. *Science*, **286**, 509–512 (1999)

[2] BOLLOBÁS, B., SPENCER, J., RIORDAN, O., TUSNÁDY, G.: The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, **18**, 279–290 (2001)

[3] BOLLOBÁS, B., JANSON, S., RIORDAN, O.: The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, **31**, 3–122 (2007)

[4] BOLLOBÁS, B., JANSON, S., RIORDAN, O.: Sparse random graphs with clustering. *Isaac Newton Institute preprint*, NI08030-CSM (2008)

[5] COOPER, C., FRIEZE, A.: A General Model of Web Graphs. *Random Structures and Algorithms*, **22**, 311-335 (2003)

[6] COOPER, C., FRIEZE, A., VERA, J.: Random Deletion in a Scale-Free Random Graph Process. *Internet Mathematics*, **1**(4), 463–483 (2004)

[7] CHUNG, F., LU, L., DEWEY, T.G., GALAS, D.J.: Duplication Models for Biological Networks. *Journal of computational biology*, **10**(5), 677–687 (2003)

[8] DEO, N., CAMI, A.: Preferential deletion in dynamic models of web-like networks. *Information Processing Letters*, **102**, 156–162 (2007)

[9] FELLER, W.: *An introduction to probability theory and its applications, Vol 1*, 3rd edition. Wiley, Singapore (1968)

[10] HACCOU, P., JAGERS, P., VATUTIN, V.A.: *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambridge University Press, Cambridge (2005)

[11] NEWMAN, M.: The structure and function of complex networks, *SIAM Rev.* **45**, 167-256 (2003)

[12] ROSS, S.M.: *Probability Models* 8th edition. Academic Press, New York (2003)

[13] TUROVA, T.S.: Dynamical random graphs with memory. *Physical Review E*, **65**, 066102 (2002)

[14] TUROVA, T.S.: Long Paths and Cycles in Dynamical Graphs. *Journal of Statistical Physics*, **110**(1/2), 385–417 (2003)

[15] TUROVA, T.S.: Phase Transitions in Dynamical Random Graphs. *Journal of Statistical Physics*, **123**(5), 1007–1032 (2006)

[16] TUROVA, T.S.: Continuity of the percolation threshold in randomly grown graphs. *The Electronic Journal of Probability*, **12**(36), 1036–1047 (2007)