Cluster growth in two growing network models

2002 J. Phys. A: Math. Gen. 35 1179

(http://iopscience.iop.org/0305-4470/35/5/305)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 1179–1194 PII: S0305-4470(02)29589-X

Cluster growth in two growing network models

David Lancaster

School of Computer Science, University of Westminster, Harrow HA1 3TP, UK

Received 8 October 2001, in final form 27 November 2001
Published 25 January 2002
Online at stacks.iop.org/JPhysA/35/1179

Abstract

We analyse a simple model for growing tree networks and find that although it never percolates, there is an anomalously large cluster at finite size. We study the growth of both the maximal cluster and the cluster containing the original vertex and find that they obey power laws. This property is also observed through simulations in a nonlinear model with loops and a true percolating phase.

PACS numbers: 64.60.Ak, 82.30.-b, 89.75.-k

1. Introduction

Recently there has been interest in the properties of random networks that are constructed by a growing process. These networks appear to model certain observed systems rather better than the random graphs of Erdős and Rényi [1]. Already, two reviews are available [2, 3].

At first, interest concentrated on the degree of distribution. It has been noticed that networks such as the World Wide Web, the Internet backbone and scientific collaboration graphs have (at least in some range) a power law degree distribution. This is in contrast to the Poisson distribution found in random graphs. Barabási and Albert [4, 5] noticed that a power law distribution could be obtained in a grown network with preferential attachment.

The grown nature of the network creates correlations that affect more than the degree distribution. More recently, these other aspects of grown networks have been studied: in particular, the phenomenon of percolation, that was of great interest in the study of random graphs. Some grown models are devised to be fully connected and percolation cannot be studied, but in a recent paper, Callaway *et al* [6] studied percolation in a very simple network growth model. Their model introduces a new vertex at every time step, and also, with probability δ, makes a link between two existing vertices, chosen at random. Percolation in this model displays some interesting features that distinguish it from percolation in a corresponding random graph with the same degree distribution. For example, the location and order of the phase transition are modified by the correlations present in the grown model. A similar treatment of other models has been performed by Dorogovtsev *et al* [7], who note that below the percolation transition, the cluster size distribution has a power law dependence in contrast to the exponential dependence typical in non-grown models.

In this paper we pursue the investigation of percolation in grown networks concentrating on models in which the new vertex introduced at each time step, is itself the endpoint of the link possibly created in that time step. Models of this type were in fact the original kind proposed by Barabási and Albert [4,5]. Only later did Dorogovtsev and Mendes [8] introduce the other type of model in which vertex and link creation are decoupled, and which often happens to be more convenient for calculation. In general there are families of such models in which more than one link is added per time step. When two or more links are added there seems to be little to distinguish the statistical features of the two families of models and they can be used interchangeably. However, for the particular case of single-link addition, the models do have a different character because in the case where the vertex is attached, only tree networks can be created. For this reason we call this model, that forms the basis of study in this paper, the 'tree growth model'.

Although most of the physical networks motivating the surge of interest in this subject are not tree-like, other examples, such as food webs, would appear not to contain, or at least to have a low probability of containing loops. However, a significant reason for studying a tree growth network comes from past experience: tree graphs have provided a fruitful field for investigating percolation in non-grown networks. They have been studied both in the physics literature and through the mathematical field of branching processes. Tree models provide an infinite-dimensional or mean field model that is often tractable in a way that finite-dimensional models are not. We shall find that the tree growth model that forms the basis of this paper is indeed a simple tractable model that illuminates more complicated scenarios.

The most interesting feature that we shall use this model to expose is the power law growth of cluster size as the network size increases. In numerical work (and for the size of many practical networks) this feature, and the presence of what appear to be anomalously large clusters, mask the lack of strict percolation in the tree growth model. This is because in random graphs, cluster growth below the percolation threshold is only logarithmic. However, as was pointed out by Dorogovtsev *et al* [7], the power law growth based on the underlying power law size distribution, makes the whole phase have scaling characteristics typical of critical behaviour.

We study the properties of clusters in depth and besides investigating the distribution of sizes of clusters chosen at random, we also study the size of the cluster containing the initial point. This illuminates the intuition that there is a highly connected 'old core' that forms the nucleus of the large clusters and turns out to give a useful analytic handle that is not so obvious in nonlinear models with loops. To ensure that the phenomena we are studying are not an artefact of the tree model we introduce a nonlinear extension and perform some numerical simulations.

It is useful to contrast the properties of this grown tree model with a non-grown or static analogue. In the present case, we argue that the appropriate analogue is a branching process rather than a random graph. Callaway *et al* [6] in their paper on percolation, ascribe the cause of the differences between percolation on grown networks and random graphs to correlations between the degrees of vertices at each end of connecting links. We demonstrate that the tree growth model does not have any such correlations.

The paper is organized as follows. After defining the tree growth model, the branching process we use as a static analogue is introduced. The percolative and other properties of these models are then compared. The main results on the tree growth model are contained in the sections describing the growth of the maximal cluster and the cluster containing the origin. A calculation of the vertex degree correlations in this model is the subject of section 4. The final part of the paper concerns a nonlinear generalization of the tree growth model, which is introduced and numerically simulated in order to confirm that the cluster growth properties observed in the tree model are preserved in more complicated models with loops.

2. Tree growth model

In each time step a new vertex is introduced. With probability δ, the new vertex is connected to another vertex, chosen at random from amongst the existing vertices. The vertex remains disconnected with probability $1-\delta$. In numerical simulations we always start with a single vertex at time $t = 1$, but we do not expect this initial condition to affect results at large times. This model can be identified as the $m = 1$ model A variant without preferential attachment, of the scale free BA models discussed in [5]. It differs from the model considered by Callaway *et al* since only the new vertex has an opportunity to make a link.

This model only generates clusters of tree graphs. There is only a single tree cluster for the case $\delta = 1$. These clusters are fragile in the sense that single deletions will always destroy connectivity [2].

2.1. Degree distribution

We commence by investigating the distribution of the vertex degrees, that is, the number of links attached to a given vertex. Following the notation and methods of Callaway *et al*, we denote the expected number of vertices of degree k at time t by $d_k(t)$. Since the total number of vertices at time t is precisely t, the probability of attaching a new link to an existing vertex of degree k is d_k/t, leading to the following evolution equations:

$$d_0(t+1) = d_0(t) - \delta\frac{d_0(t)}{t} + (1-\delta) \tag{1}$$

$$d_1(t+1) = d_1(t) - \delta\frac{d_1(t)}{t} + \delta\frac{d_0(t)}{t} + \delta \tag{2}$$

$$d_k(t+1) = d_k(t) - \delta\frac{d_k(t)}{t} + \delta\frac{d_{k-1}(t)}{t} \qquad k \geqslant 2. \tag{3}$$

Note that the total number of vertices can be written as $\sum_0^\infty d_k(t) = t$ and that the total expected number of links is given by $\frac{1}{2}\sum_0^\infty k d_k(t) = \delta t$. Since both quantities grow linearly in time we search for solutions of the form, $d_k(t) = p_k t$, and find

$$p_0 = \frac{1-\delta}{1+\delta} \tag{4}$$

$$p_k = \frac{2}{1+\delta}\left(\frac{\delta}{1+\delta}\right)^k \qquad k \geqslant 1. \tag{5}$$

This distribution decays exponentially in contrast to random graph models which have a Poisson degree distribution, and the scale free models with power law distribution. As was demonstrated in [5], this type of decay is characteristic of models in which linking is not preferential.

2.2. Static analogue—branching process

Before proceeding to investigate clustering issues we pause to introduce a non-grown or static analogue of this model. The static model should have the same vertex degree distribution as the grown model, but should be constructed to avoid any correlation between the degree of linked vertices that might arise from the growing process. Furthermore, the analogue should preserve the tree-like character of the model, so it cannot be one of the classic random graphs of Erdős and Rényi [1]. An appropriate model is based on an ensemble of Galton and Watson branching processes [11, 12].

A branching process may be regarded as a growth process in its own right, but each vertex is treated identically, thus avoiding any potential correlation between vertex degrees. In order

to reproduce the vertex degree distribution, we choose the probability of k offspring to be proportional to p_{k+1} in equations (4), (5), so

$$p_k = \frac{1}{1+\delta}\left(\frac{\delta}{1+\delta}\right)^k \qquad k \geqslant 0. \tag{6}$$

This choice gives the correct ratios of vertex degrees at all higher levels. However, at the first level, where no link is already present, it is not obvious that the choice correctly weights the vertices with no children at all. We return to this issue when we discuss the ensemble of branching processes.

The properties of the model are then a textbook exercise [11], but for completeness we summarize the main steps. The main concern is the with the cluster sizes, in particular the question of percolation. This approach based on branching processes is identical to the studies of percolation on trees, for example Bethe lattices, which were popular in the 1980s [10] and provided a mean field model for the percolation transition.

Percolation occurs in this model when the extinction probability of the branching process is less than unity. This extinction probability may be calculated using the generating function for the probabilities (6):

$$g(x) = \sum_0^\infty p_k x^k = \frac{1}{1+\delta-\delta x}. \tag{7}$$

The extinction probability is given by the smallest root, x_0, of the equation: $g(x) = x$. This root is 1 for all values of δ so percolation never takes place (though, in the same way as for one-dimensional percolation, $\delta = 1$ may be regarded as a critical point).

The technique above can be extended to find the distribution n_i, of finite clusters in this model. For a single branching process, the generating function, $\rho(x) = \sum_1^\infty n_i^B x^i$, for the quantities n_i^B, which are the probabilities that the process contains i nodes, is given by the solution to, $\rho(x) = xg(\rho(x))$, and is found to be

$$\rho(x) = \frac{(1+\delta)}{2\delta} - \frac{1}{2\delta}\sqrt{(1+\delta)^2 - 4\delta x}. \tag{8}$$

The quantities n_i^B may now be read off, however these are not the cluster numbers n_i, as usually defined. The static model is an ensemble of branching processes, so n_i^B corresponds to the number of clusters of size i per process, but n_i is the number per node. To relate these quantities we compute the average number of nodes in a branching process as $\rho'(1) = 1/(1-\delta)$. In the limit of a large ensemble we then find $n_i = (1-\delta)n_i^B$. A proper discussion of the ensemble would allow a number of isolated nodes besides the clusters based on branching processes, in order to adjust the degree distribution. This more careful discussion leads to the same result:

$$n_1 = \frac{1-\delta}{1+\delta} \tag{9}$$

$$n_2 = \frac{\delta(1-\delta)}{(1+\delta)^3} \tag{10}$$

$$n_3 = \frac{2\delta^2(1-\delta)}{(1+\delta)^5}. \tag{11}$$

A recursion relation may be obtained for higher order terms. These results are used for comparison with the tree growth model.

2.3. Cluster size distribution

The expected number of clusters of size i, N_i, in the tree growth model obey a set of evolution equations that can be obtained by noting that the probability of making a link to a cluster of i

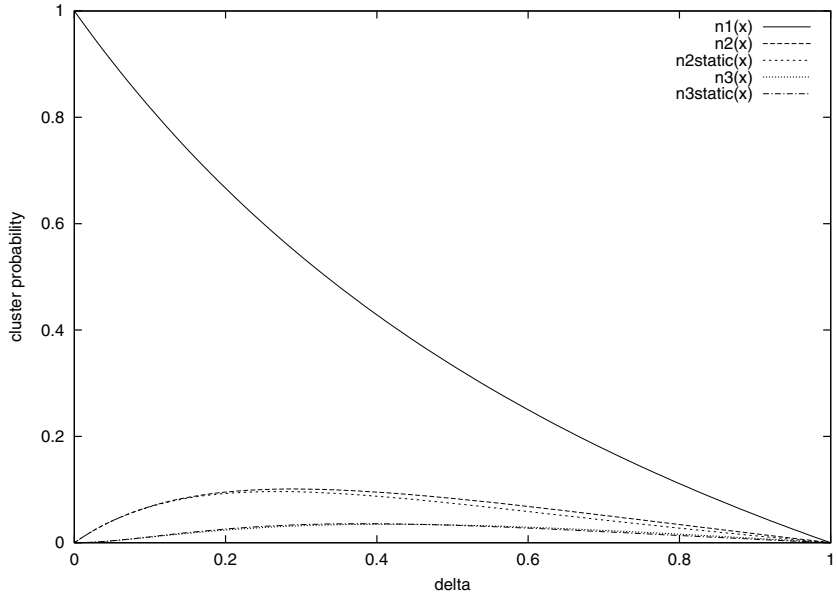

Figure 1. The fraction of vertices in clusters of size 1, 2, 3 (n_1, n_2, n_3) according to the formulae (14), (15) and also the static results (9)–(11). These predictions coincide for clusters of size 1. Simulation results lie on the exact curve but are not shown in this plot.

vertices is iN_i/t. In contrast to the situation in more complicated models, these equations are linear, exact and hold for finite t:

$$N_1(t+1) = N_1(t) - \delta\frac{N_1(t)}{t} + (1-\delta) \tag{12}$$

$$N_i(t+1) = N_i(t) - \delta\frac{iN_i(t)}{t} + \delta\frac{(i-1)N_{i-1}(t)}{t}. \tag{13}$$

The expected total number of clusters $\sum_1^\infty N_i(t)$ grows linearly in time and is given by $(1-\delta)t$, since a new cluster is created whenever a link is not made in a time step. By summing the equations (weighted by i), we also find that the first moment is given by the total number of vertices, $\sum_1^\infty iN_i(t) = t$. These relations also reflect the fact that each cluster is a tree graph, so the number of links is the number of vertices minus one. We search for the cluster size distribution, n_i, of the form, $N_i(t) = n_i t$ and find the following recursion relations:

$$n_1 = \frac{1-\delta}{1+\delta} \tag{14}$$

$$n_i = \frac{(i-1)\delta}{(i\delta+1)}\,n_{i-1} \qquad i \geqslant 2. \tag{15}$$

Although the first term, n_1, is (by design) the same as for the static model, later terms are different. Figure 1 shows the first few terms of the cluster size distribution for both the static model and the growth model. Notice that while the exact result is similar to the static one for small delta, it is smaller for larger delta.

Indeed, the large cluster behaviour of the static and the growth model are completely different. The large cluster behaviour of (15) is power law:

$$n_i \stackrel{i\to\infty}{\longrightarrow} n_1\Gamma(2+1/\delta)i^{-(1+1/\delta)}. \tag{16}$$

That of the branching model is dominated by exponential decay.

The power law decay of the cluster distribution for the growing model has been noticed by Dorogovtsev *et al* [7] in the non-percolating phase of nonlinear growth models. They have termed it a self-organized critical state because the preferential attachment to larger clusters which causes the power law decay occurs automatically. For non-grown networks, exponential decay of the cluster numbers is a common feature (as in the static example). This difference has consequences for the way clusters grow.

3. The largest cluster

Direct numerical simulations of the growing network indicate that for values of δ larger than about $1/2$, there is a cluster of size considerably larger than the others. This cluster often contains the original vertex and suggests that there may be a percolating cluster based on the 'old core' of vertices that are created early in the growth. These numerical simulations are in fact misleading, but expose anomalous finite size effects that are studied below.

An analytic approach to percolation does not take the usual route because the equations (12), (13) are exact and hold for any incipient percolating cluster besides the finite clusters. Ordinarily, the sum, $\sum_1^\infty i n_i$ only accounts for finite clusters and the infinite cluster must be added separately. However, according to the equations, this sum equals t and contains all the vertices, thereby leaving no room for an infinite cluster. The generating function approach used in [6], although pleasantly tractable, merely reproduces this information.

Percolation does not occur in this model, except in the trivial limiting case $\delta = 1$ where the network just consists of a single tree graph. This phase diagram resembles that of ordinary one-dimensional percolation. To understand the reasons why percolation does not take place, yet large clusters do appear at finite size, it is helpful to study the numerical data for the maximum sized cluster. This will then lead us to an investigation of the cluster containing the original site.

3.1. Numerical study

On closer inspection of the numerical data it is found that the fraction of sites contained in the largest cluster suffers from an anomalously slow finite size effect, becoming smaller as the growth process is continued to larger times. For example, at $\delta = 0.8$, the fraction drops from about 0.29 at $t = 10^3$, to 0.17 for a network ten times larger. In figure 2 we show the largest cluster fraction against $\log(t)$ for various δ. The straight lines clearly indicate a power law dependence. The exponent can be determined by fitting, or by noticing that another plot of the same quantity (log) against δ displays linear dependence. In any event, the lack of any transition is clear. The fit suggests the form

$$\text{Fraction of vertices in largest cluster} \sim t^{\delta-1}. \tag{17}$$

This form of scaling behaviour can be deduced from the original growth model. Consider a large isolated cluster, $N_{\bar{i}} = 1$. By treating its size, $\bar{i}$, as a continuous variable, we find that it grows according to the probability that a link will attach the new vertex to this cluster:

$$\bar{i}(t+1) = \bar{i}(t) + \delta\frac{\bar{i}}{t}. \tag{18}$$

There is no solution linear in t, but a form $\bar{i} \sim t^\delta$ solves the equation in the large-time limit. The fraction of sites in this largest cluster, $\bar{i}/t$, therefore follows the scaling behaviour observed numerically in (17). As the system grows very large, the relative size of even the largest cluster decreases and it is apparent that the tree growth model never experiences true percolation.

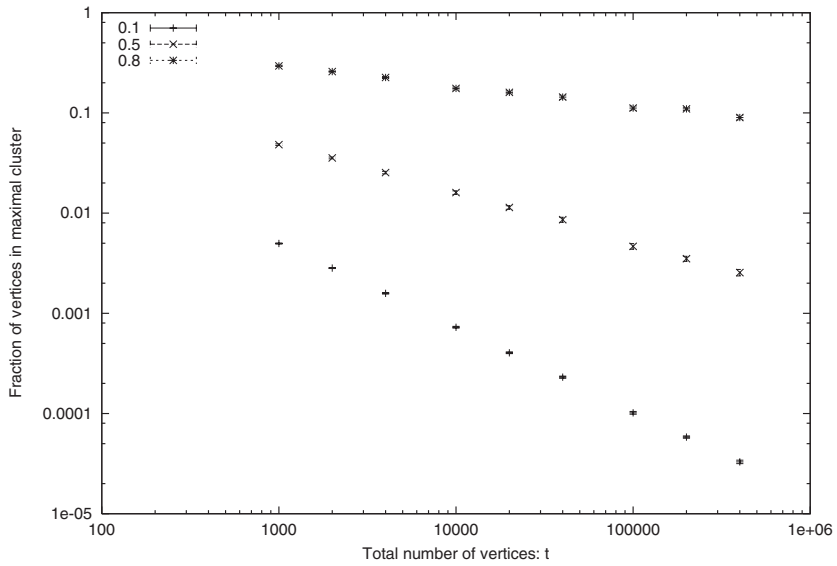

Figure 2. Scaling of largest cluster size fraction against t for $\delta = 0.1, 0.5, 0.8$.

In most static models with percolation, for example random graph models, the finite size scaling of the maximum cluster size is given by $\log(t)$. This is related to the usual exponential decay of the cluster size distribution, and the power law behaviour we see here follows from the distinctive decay (16) in growth models.

3.2. Cluster containing the initial point

The overall distribution of the sizes of randomly chosen clusters (14), (15) does not give any hint of the presence of the large cluster seen in the numerical work above. It is hard to investigate the maximal cluster analytically, but if we rely on the observation that the maximal cluster is likely to be based on one of the oldest vertices, we may approach the problem from a different perspective. The distribution of the size of clusters that contain the original point is amenable to analytic methods and does shed some light on the presence of a large cluster. The possibility of studying this quantity is of course only available in grown networks that have distinguished vertices.

As before, we start by writing evolution equations, this time for the probability $P_i(t)$ that a distinguished cluster has size i at time t ($1 \leqslant i \leqslant t$):

$$P_1(t+1) = P_1(t) - \delta\frac{P_1(t)}{t} \tag{19}$$

$$P_i(t+1) = P_i(t) - \delta\frac{iP_i(t)}{t} + \delta\frac{(i-1)P_{i-1}(t)}{t} \tag{20}$$

$$P_{t+1}(t+1) = \delta P_t(t). \tag{21}$$

These equations are very similar to the ones for the overall cluster size distribution $N_i(t)$ in (12), (13), however, the difference in the first equation prevents any solution $P_i(t) \propto t$. The equations actually hold for *any* distinguished cluster, with the initial condition determining which cluster is selected. Simplest is to choose the cluster distinguished as containing the original point, in which case $P_1(1) = 1$. Other possibilities, for example the cluster containing the second point would be determined by the values at $t = 2$, $P_1(2) = 1 - \delta$ and

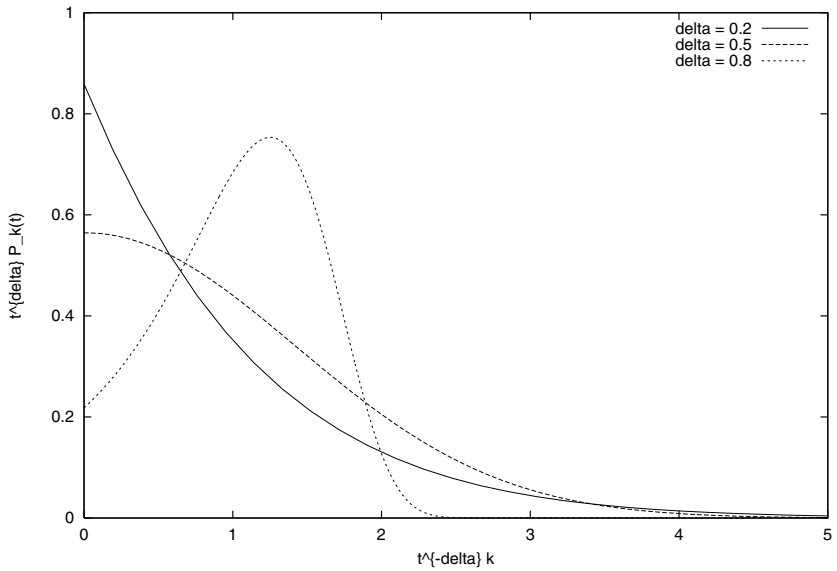

Figure 3. Size distribution of the cluster containing the origin with a scaling plot of $t^{\delta} P_k(t)$ against $t^{-\delta}k$. For a variety of values of δ. The curves are obtained by numerically solving the equations (19)–(21) up to $t = 10^4$.

$P_2(2) = \delta$. This in fact leads to the same distribution as for the first point, but a difference is obtained for the third point which is specified by $P_1(3) = 1 - \delta$, $P_2(3) = \delta(1 - \delta)$ and $P_3(3) = \delta^2$. In the following, we shall only consider the cluster containing the original point.

The sum $\sum_1^t P_k(t)$ is preserved by these equations, and can be set to 1, as expected for a probability, by the initial condition. The average size of the distinguished cluster, $\bar{k}(t) = \sum_1^t k P_k(t)$, obeys $\bar{k}(t+1) = (1 + \delta/t)\bar{k}(t)$. So at large times we expect that $\bar{k}(t) \sim t^{\delta}$. This is essentially the same argument as in (18) of the last section and indeed the evolution equation has the same intuitive origin. In this form the prefactor can be determined from the initial condition. Equations for all the higher moments of the distribution will be considered below.

For large t and k, the continuum version of the evolution equation becomes

$$t\frac{\partial P}{\partial t} = -\delta \frac{\partial (kP)}{\partial k} \tag{22}$$

which has a scaling solution,

$$P(t,k) = t^{-\delta} f(kt^{-\delta}) \tag{23}$$

where $f(u)$ is any function.

This result is confirmed, and the form of the scaling function $f(u)$ determined, by numerically solving the difference equations (19)–(21) and plotting them appropriately as shown in figure 3. No change in the form of the function is visible as t is increased beyond about 2000. A scaling relation of this form is interesting because it is found for all values of δ not just those in the vicinity of the critical point at $\delta = 1$.

The scaled cluster distribution shows a clear change in form around $\delta = 1/2$. Although the mean of the distribution varies smoothly with δ, and is close to 1 on the scaled plot (corresponding to $\bar{k}(t) = t^{\delta}$ before scaling), the mode moves away from zero (cluster size,

$k = 1$, before rescaling) as δ becomes greater than about $1/2$. Eventually, as $\delta \to 1$ the scaling function becomes progressively more peaked around $u = 1$. This provides an argument for the likely presence of a large maximal cluster for $\delta \gtrsim 1/2$.

The form of the scaling function is not easy to determine analytically. Only in the limit of large or small u, can $f(u)$ be determined using the solutions for $P_1(t)$ and $P_t(t)$ obtained from (19) and (21).

For comparison with simulations it is better to compare the moments of the distribution rather than the full form. The moments, defined as,

$$S_n(t) = \sum_1^t k^n P_k(t) \tag{24}$$

obey equations obtained from weighted sums of (19)–(21):

$$S_0(t+1) = S_0(t) \tag{25}$$

$$S_1(t+1) = \left(1 + \frac{\delta}{t}\right) S_1(t) \tag{26}$$

$$S_2(t+1) = \left(1 + \frac{2\delta}{t}\right) S_2(t) + \frac{\delta}{t} S_1(t) \tag{27}$$

$$S_n(t+1) = S_n(t) + \frac{\delta}{t} \sum_1^n \binom{n}{i-1} S_i(t) \tag{28}$$

where the last equation contains a binomial coefficient. By forming suitable linear combinations, these equations can be solved in terms of the following function:

$$R(z,t) = \prod_{i=1}^{t-1} (1 + z/i) = \frac{\Gamma(z+t)}{\Gamma(t)\Gamma(z+1)} \tag{29}$$

$$\xrightarrow{t\to\infty} \frac{t^z}{\Gamma(z+1)}. \tag{30}$$

For example

$$S_0(t) = R(0,t) = 1 \tag{31}$$

$$S_1(t) = R(\delta,t) \tag{32}$$

$$S_2(t) = 2R(2\delta,t) - R(\delta,t) \tag{33}$$

$$S_3(t) = 6R(3\delta,t) - 6R(2\delta,t) + R(\delta,t). \tag{34}$$

Higher terms can be computed iteratively, and it is also possible to treat clusters containing other than the original point. As t becomes large, $R(n\delta,t) \sim t^{n\delta}$, so the leading term dominates and $S_n(t) \to n!R(n\delta,t)$. However, for finite t, the sub-leading terms are large in the region $\delta \lesssim 1/\log(t)$ and must be kept in numerical work.

In figure 4 we show comparisons of these formulae against simulation results for the mean and the second moment. Bearing in mind the scaling behaviour, we plot each moment divided by a power of t^δ. These first moments show excellent agreement.

4. Vertex degree correlations

To conclude our study of the tree growth model we follow the same argument used by Callaway *et al* to determine the correlations between the vertex degree at each end of a randomly chosen link. Such correlations were studied in preferential attachment models in [13]. The number

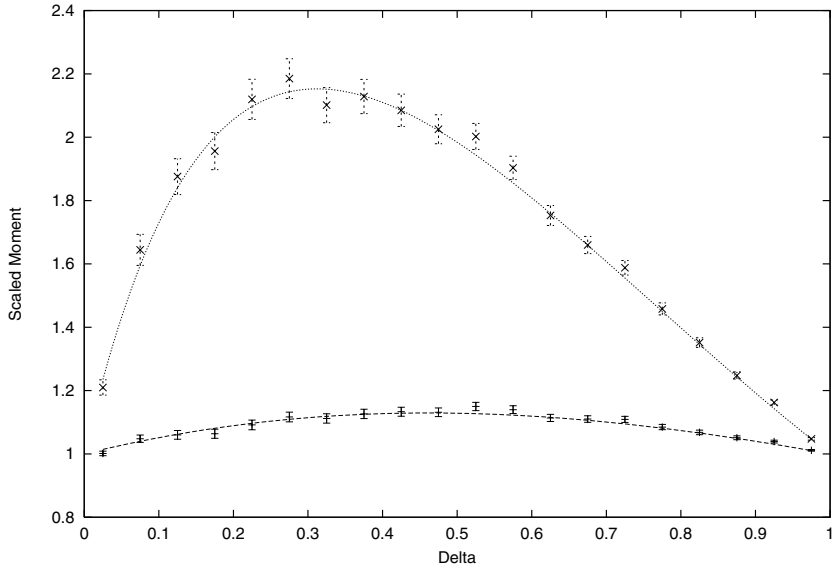

Figure 4. First two moments of the size of cluster containing the origin. The upper curve is for the second moment, and the lower curve is for the mean. Each moment is scaled according to $t^{-n\delta}S_n(t)$ and based on simulations of 4000 samples of networks containing 10^4 vertices. The curves are the theoretical results including the sub-leading finite size terms.

of edges that join vertices of degrees j and k is denoted E_{jk}. This matrix is symmetric. For links that join vertices of the same degree, E_{kk} is defined to be *twice* the number of such links. In this case *exact* evolution equations can be derived by treating the vertices with a single link specially:

$$E_{11}(t+1) = E_{11}(t) + 2\delta\frac{d_0}{t} - 2\delta p_1\frac{E_{11}}{d_1} \tag{35}$$

$$E_{1k}(t+1) = E_{1k}(t) + \delta\frac{d_{k-1}}{t} - \delta\left(p_1\frac{E_{1k}}{d_1} + p_k\frac{E_{1k}}{d_k}\right) \tag{36}$$

$$E_{jk}(t+1) = E_{jk}(t) + \delta\left(p_{k-1}\frac{E_{jk-1}}{d_{k-1}} + p_{j-1}\frac{E_{j-1k}}{d_{j-1}}\right) - \delta\left(p_j\frac{E_{jk}}{d_j} + p_k\frac{E_{jk}}{d_k}\right) \tag{37}$$

where $d_k(t)$ and p_k are the vertex degree numbers and their probabilities as determined earlier in section 2.

The total expected number of links is given by $\frac{1}{2}\sum_{jk}E_{jk}(t)$ and the evolution equations show that it is given by δt as anticipated. We therefore write the probabilities as $E_{jk}(t) = 2\delta t e_{jk}$, and derive the following equations:

$$(1+2\delta)e_{11} = p_0 \tag{38}$$

$$(1+2\delta)e_{1k} = \frac{p_{k-1}}{2} + \delta\, e_{1k-1} \tag{39}$$

$$(1+2\delta)e_{jk} = \delta(e_{jk-1} + e_{j-1k}). \tag{40}$$

By appropriately multiplying these equations and adding, we can find the following relations between the moments:

$$M_0 = \sum_{jk} e_{jk} = 1 \tag{41}$$

$$M_1 = \sum_{jk} j e_{jk} = 1 + \delta + \tfrac{1}{2} \sum_{j=0} j p_j \tag{42}$$

$$M_2 = \sum_{jk} jk e_{jk} = 1 + 2\delta M_1 + \sum_{j=0} j p_j. \tag{43}$$

Using the results of section 2 on the vertex degrees, we find the average degree $\sum_0^\infty k p_k = \bar{k} = 2\delta$. The sum above includes vertices with no links, and the average degree on the end of a randomly chosen link is $\mu = \sum k^2 p_k / \sum k p_k = 1 + 2\delta$.

The covariance between vertex degrees at each end of a randomly chosen link is defined as,

$$C = \sum_{jk} (j - \mu)(k - \mu) e_{jk} \tag{44}$$

$$= M_2 - 2\mu M_1 + \mu^2 M_0. \tag{45}$$

Combining these results we find that C vanishes identically and that there is no correlation between the degrees at the end of randomly chosen links in this model. This result is supported by simulations.

In view of this result, it is slightly surprising that the analogue static model which was specifically designed to avoid these correlations, is not identical to the tree graph model. There is still a distinction as was apparent from the cluster numbers.

5. Two-link growth model

The preceding study of cluster growth in the tree growth network has been reasonably tractable, fundamentally due to the tree property of the network. The question arises as to which features are preserved in more general models.

The most obvious difference in more complicated models is the presence of a percolating phase. The tree growth model has no percolating phase except the trivial one at $\delta = 1$. The physical reason for this deficiency is not directly the tree nature of the network. The cause should rather be sought in the growth itself. There is no mechanism to attach existing clusters to each other. A mechanism of this type was responsible for the percolating properties in the model of Callaway *et al*, and clearly introduces nonlinearities into the model, for example in the equation for the generating function of cluster sizes.

A natural extension of our tree growth model is a model in which at each time step a new vertex is created and then, with probability δ connected by *two* links to the existing vertices. Each new link is assigned a random terminating vertex amongst the existing vertices. The networks grown by this model are not necessarily tree-like and loops can form. The static analogue is therefore a random graph rather than a branching process. Again, in numerical work, we use an initial condition consisting of a single vertex. This model can be identified as the $m = 2$ model A variant without preferential attachment, of the scale free BA models discussed in [5]. A further generalization that we have considered is a model in which there is are fixed probabilities for single- and double-link connections. This leads to a more complicated phase diagram, but not to any significantly new observations. Yet another approach would be to consider cluster growth in a model with a fixed number of vertices. Such models of aggregation were considered by Smoluchowski and applied to polymerization of ARB_{f-1}

branched molecules by Stockmayer [14] and others. The gelation transition found in these models corresponds to percolation.

Below, we present the basic properties of the two-link model closely following the methods of Callaway *et al*. We then perform simulations to study the cluster properties and compare them with what was found in the tree model.

5.1. Two-link growth model—degree distribution

The equations leading to the degree distribution that are obtained by the same means as for the tree growth model:

$$d_0(t+1) = d_0(t) - 2\delta\frac{d_0(t)}{t} + (1-\delta) \tag{46}$$

$$d_1(t+1) = d_1(t) - 2\delta\frac{d_1(t)}{t} + 2\delta\frac{d_0(t)}{t} \tag{47}$$

$$d_2(t+1) = d_2(t) - 2\delta\frac{d_2(t)}{t} + 2\delta\frac{d_1(t)}{t} + \delta \tag{48}$$

$$d_k(t+1) = d_k(t) - 2\delta\frac{d_k(t)}{t} + 2\delta\frac{d_{k-1}(t)}{t}. \tag{49}$$

Note that the total number of vertices, t, can be written as $\sum_0^\infty d_k(t)$ and that the expected number of links, $2\delta t$, is given by $\frac{1}{2}\sum_0^\infty k d_k(t)$. Searching for solutions of the form, $d_k(t) = p_k t$, we find

$$p_0 = \frac{1-\delta}{1+2\delta} \tag{50}$$

$$p_1 = \frac{2\delta(1-\delta)}{(1+2\delta)^2} \tag{51}$$

$$p_k = (1+8\delta)\frac{2^{k-2}\delta^{k-1}}{(1+2\delta)^{k+1}} \qquad (\text{for } k \geqslant 2). \tag{52}$$

Again this distribution decays exponentially after the first couple of terms.

5.2. Two-link growth model—cluster size distribution

The cluster sizes N_i in the two-link growth model obey a set of evolution equations which are now approximate and only valid for finite clusters at large t since processes in which both links end in the same cluster are ignored. This is the same approximation that is made in the Callaway *et al* analysis:

$$N_1(t+1) = N_1(t) - 2\delta\frac{N_1(t)}{t} + (1-\delta) \tag{53}$$

$$N_2(t+1) = N_2(t) - 2\delta\frac{2N_2(t)}{t} \tag{54}$$

$$N_i(t+1) = N_i(t) - 2\delta\frac{iN_i(t)}{t} + \delta\sum_{j=1}^{i-2}\frac{jN_j(t)}{t}\frac{(i-j-1)N_{i-j-1}(t)}{t}. \tag{55}$$

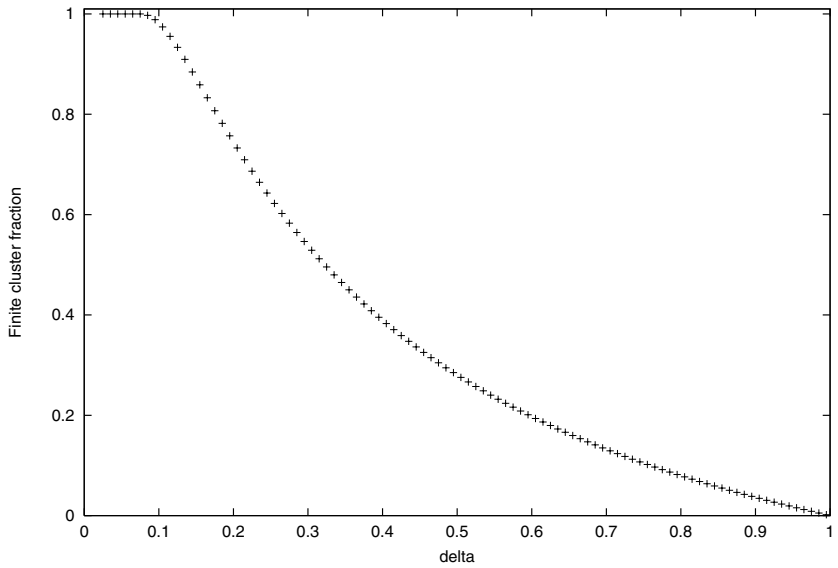

Figure 5. The fraction of the vertices contained within finite clusters. Obtained by numerical integration of the differential equation (59) using a step size of 10^{-6}.

Solutions of the form $N_i(t) = n_i t$ are considered and a recursion relation obtained:

$$n_1 = \frac{\delta}{1+2\delta} \tag{56}$$

$$n_2 = 0 \tag{57}$$

$$n_i = \frac{\delta}{1+2i\delta} \sum_{j=1}^{i-2} j n_j (i-j-1) n_{i-j-1}. \tag{58}$$

Analysis of these cluster numbers is best carried out using the generating function for the cluster sizes, $g(x) = \sum_1^\infty i n_i x^i$, which obeys a nonlinear equation

$$g' = \frac{1}{2\delta}\left(\frac{1-\delta-g/x+\delta g^2}{1-xg}\right). \tag{59}$$

5.3. Two-link growth model—percolation

We compute $g(1)$ by numerically integrating the equation (59) starting from an initial condition ($g(\epsilon) = n_1\epsilon$). Figure 5 shows the results, and we recall that $g(1)$ is the expected fraction of vertices contained in the finite clusters, so when it differs from 1, percolation occurs.

The model percolates for most of the range of δ, but for a range of small δ there is no percolation. It is possible to obtain the critical value δ_c by studying $g'(1)$. In the percolating region $g(1) < 1$, so it is simple to take the $x \to 1$ limit of the right-hand side of equation (59) to obtain,

$$g'(1) = \frac{1}{2\delta}\left(1-\delta(1+g(1))\right). \tag{60}$$

In the case $\delta < \delta_c$, $g(1) = 1$, and this limit must be taken more carefully with the help of L'Hopital's rule. The resulting quadratic equation can be solved to give

$$g'(1) = \frac{1-4\delta \pm \sqrt{1-16\delta+16\delta^2}}{4\delta}. \tag{61}$$

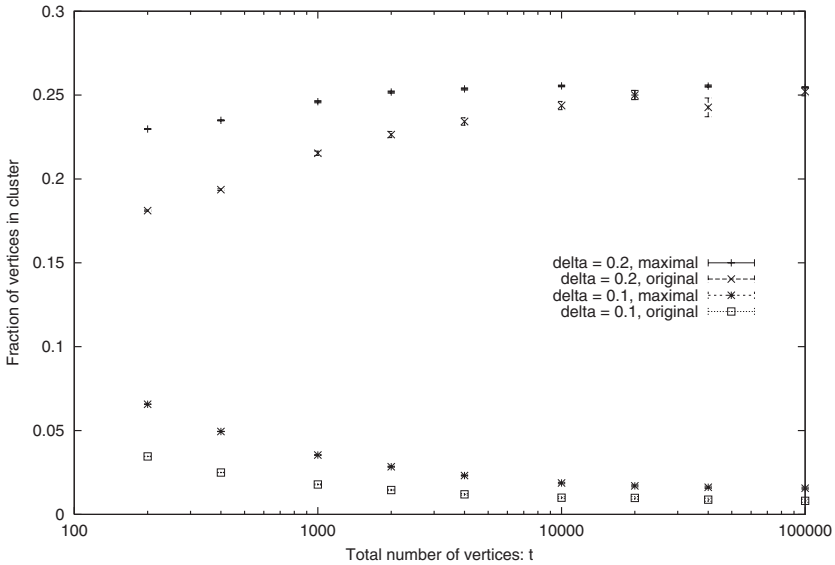

Figure 6. Maximal and original cluster fractions in the two-link model for two values of δ, (0.1, 0.2) within the percolating region. In each case the maximal curve is above the original one. Averages are taken over a large number of samples ranging from 10^2 for the largest networks to 10^5 at the smallest.

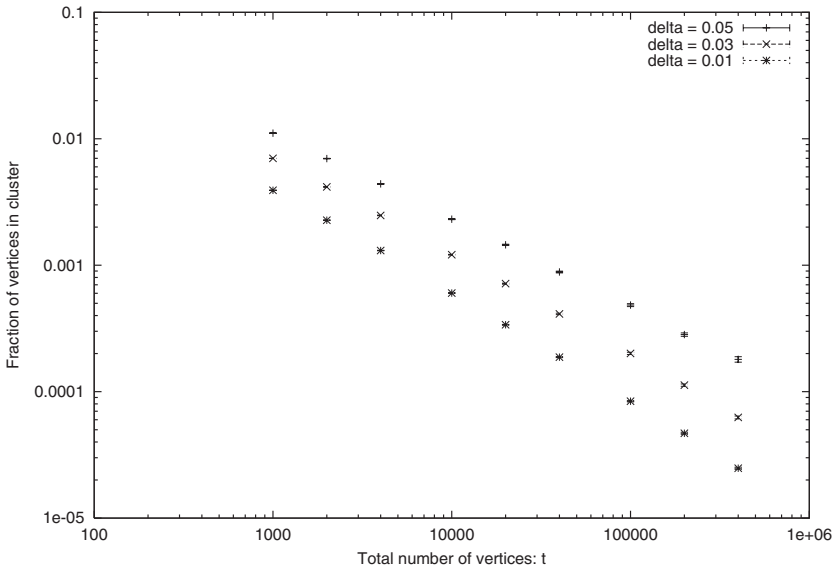

Figure 7. Scaling of maximal cluster in the two-link model below the percolation threshold. The original cluster follows similar curves. For $\delta = 0.01, 0.03, 0.05$. Averages are taken over a large number of samples ranging from 50 for the largest networks to 10^4 for the smallest.

We omit regions where the root is not real, and further require that it be positive. Finally, recognizing that $g'(1) \to 1$ as $\delta \to 0$ since in this limit all clusters have size one, we are able to pick the negative sign as being the only correct branch.

In summary $\delta_c = 1/2 - \sqrt{3}/4 \approx 0.066\,99$. With $g'(1)$ taking different values on each side:

$$g'(1) = \begin{cases} \dfrac{1 - 4\delta - \sqrt{1 - 16\delta + 16\delta^2}}{4\delta} & \text{for} \quad \delta < \delta_c \\ \dfrac{1}{2\delta}\,(1 - \delta(1 + g(1))) & \text{for} \quad \delta > \delta_c. \end{cases} \tag{62}$$

The critical behaviour we have described is very similar to that observed in the model studied by Callaway *et al*. By performing a similar investigation near the critical point, we find the same signals of an infinite order transition with $1 - g'(1) \sim \mathrm{e}^{\alpha/\sqrt{\delta - \delta_c}}$.

6. Cluster growth in the two-link growth model

In this section we describe the results of numerical simulations to find how large clusters grow in this model. We track both the maximal cluster and the cluster containing the original vertex.

In the region above the percolation threshold the maximal cluster naturally grows with t. It is interesting to see how finite size affects influence this and how the cluster containing the original point grows. This is shown in figure 6 which indicates that there is a region where the original cluster is smaller than the maximal one, but as the size of the network increases, this cluster approaches the size of the maximal one. This result supports the intuition that the 'old core' of vertices act as a seed for the percolating cluster. Indeed, the probability that the maximal cluster contains the original vertex appears to grow to 1 for any δ in the percolating phase. Unfortunately the statistics for this analysis are not good for the sizes we have considered and this result should only be taken as suggestive.

The finite size effects are most apparent for $\delta = 0.1$ which is quite close to the critical point. In this case the fraction of sites in the either maximal or original cluster decrease with t in a way reminiscent of the behaviour in the tree growth model. Estimates of a correlation size can be made on the basis of logarithmic plots which show a clear change in slope as the network size exceeds the correlation size at that value of δ.

It is the situation below the percolation threshold that holds more interest for comparison with the tree growth model. In figure 7 we show evidence that the maximal cluster scales with a power law decay in this region. The original cluster behaves in the same way. This is exactly as in the tree growth model, and as emphasized before, quite distinct from the $\log(t)$ behaviour in random graph models.

Having demonstrated that scaling occurs in the same way as in the tree growth model, we postpone any further study of the exponent of the growth. This is because of the difficulty of getting far from the critical point in this particular model.

7. Conclusion

The study of grown networks was originally motivated by real networks which are by nature finite. We have shown in a simple tree growth model, that although an infinite size system does not display percolation, finite systems of sizes that may have relevance to observations, often contain large clusters. These clusters grow with a power law dependence on the system size and provide another manifestation of the critical nature of the whole phase. The power law growth can be analysed carefully in this model, especially by studying clusters with distinguished points, but the pattern of power law growth appears to be general as found in numerical simulations in a nonlinear model.

Acknowledgments

I would like to thank the PDC group at Southampton University for the use of computer facilities. I also acknowledge a helpful discussion with Adam Prügel-Bennett.

References

[1] Bollobás B 1985 *Random Graphs* (New York: Academic)
[2] Albert R and Barabási A L 2001 *Preprint* cond-mat/0106096
[3] Dorogovtsev S N and Mendes J F 2001 *Preprint* cond-mat/0106144
[4] Barabási A L and Albert R 1999 *Science* **286** 509
[5] Barabási A L, Albert R and Jeong H 1999 *Physica* A **272** 173
[6] Callaway D S, Hopcroft J E, Kleinberg J M, Newman M E J and Strogatz S H 2001 *Preprint* cond-mat/0104546
[7] Dorogovtsev S N, Mendes J F F and Samukhin A N 2001 *Preprint* cond-mat/0106141
[8] Dorogovtsev S N and Mendes J F F 2000 *Europhys. Lett.* **52** 33
Dorogovtsev S N and Mendes J F F 2001 *Phys. Rev.* E **63** 56 125
[9] Grimmet G 1999 *Percolation* 2nd edn (Berlin: Springer)
[10] Stauffer D and Aharony A 1992 *Introduction to Percolation Theory* (London: Taylor and Francis)
[11] Feller W 1950 *Introduction to Probability Theory and its Applications* (New York: Wiley)
[12] Athrey K B and Ney P E 1972 *Branching Processes* (Berlin: Springer)
[13] Krapivsky W P L and Redner S 2001 *Phys. Rev.* E **63** 066123
[14] Stockmayer W H 1943 *J. Chem. Phys.* **11** 45