Scientific
Research

Some Models of Reproducing Graphs:
I Pure Reproduction

Richard Southwell, Chris Cannings
School of Mathematics and Statistics, University of Sheffield, Sheffield, UK
E-mail: *bugzsouthwell@yahoo.com*
Received June 26, 2010; *revised August* 10, 2010; *accepted August* 15, 2010

Abstract

Many real world networks change over time. This may arise due to individuals joining or leaving the network or due to links forming or being broken. These events may arise because of interactions between the vertices which occasion payoffs which subsequently determine the fate of the nodes, due to ageing or crowding, or perhaps due to isolation. Such phenomena result in a dynamical system which may lead to complex behaviours, to self-replication, to chaotic or regular patterns, to emergent phenomena from local interactions. They give insight to the nature of the real-world phenomena which the network, and its dynamics, may approximate. To a large extent the models considered here are motivated by biological and social phenomena, where the vertices may be genes, proteins, genomes or organisms, and the links interactions of various kinds. In this, the first paper of a series, we consider the dynamics of pure reproduction models where networks grow relentlessly in a deterministic way.

1. Introduction

There has been much recent interest in the way in which networks such as the World Wide Web grow, and the structures which result from various rules by which new vertices are added and link to the existing vertices. One of the most studied is the so called Preferential Attachment model whereby a new node is added at each time $t \in N^+$ (we use N and N^+ for the non-negative integers and positive integers respectively) and links to some set of existing vertices with probabilities which depend on the degrees of the latter. In the simplest case the probabilities are simply proportional to the degree, a model introduced by Yule [1], again by Simon [2], and then more recently by Barábasi and Albert [3]. The outcome of this process (see [4,5]) is a network in which the degree of a randomly selected node follows a power law distribution (*i.e.*, if X is that degree then the probability $Pr(X = k) = ak^{-b}$), and the network is scale-free in the sense that $Pr(X = l^*c)/Pr(X = c) = Pr(X = l^*d)/Pr(X = d)$ for all l, c and d.

On the other hand there has been relatively little attention paid to the growth of networks through the reproduction of existing vertices and the generation of links between these new vertices and the old vertices, although,

of course, the preferential attachment model where a new vertex is linked to an existing vertex could be regarded as the production of an offspring by the latter. This is clearly a situation which arises in a biological population which reproduces itself and in which we track relatedness. In a population which reproduces asexually, if we join each individual to its parent, then we simply produce a tree for each clone. More interestingly, if in a sexually reproducing population we join each individual to their two parents we obtain a genealogy (see [6] for alternate ways of representing this network).

A further biological example happens when a genome duplication occurs [7]. The genes in the genome each code for some specific protein. If one considers the set of proteins of some organism as vertices in some network and joins any pair of vertices if the corresponding pair of proteins can bind then one obtains the protein-protein-interaction network. In a genome duplication every gene is essentially duplicated, so that there are now two copies producing the protein previously produced by one copy (we assume for simplicity that there is a simple one-one mapping of proteins to genes, ignore post-translational modification and other interactions, and splicing variation). If we then distinguish between the two copies of the genes and the proteins produced by those two copies

then we have a doubling of the set of vertices, and a quadrupling of the number of the links. This is our model 5 below.

More generally suppose that a set of entities are allowed to reproduce and that links which are produced in the new network are defined in terms of the existing links, and the relatedness of new and old vertices. In addition to the gene-duplication example above this might correspond to the establishment of the social network between individuals. For example, taking a gynocentric view, suppose that daughters of mothers who are friends are also always friends, and that mother and daughter also are treated as friends, then we obtain a particular set of relationships in a population as it reproduces, our model 6. Certainly it is well known in some species of apes and monkeys that social relationship is influenced by biological relatedness [8,9], and this is also well known in other groups e.g. spotted hyenas [10].

From now on we switch our terminology to that of graph theory, *i.e.*, refer to a graph rather than a network, and to an edge rather than a link. A graph, denoted $G(V,E)$ or just G for short, is a specification of some set V, whose elements are referred to as vertices, and some set $E \subset V \otimes^* V$ whose elements are called edges. $V \otimes^* V$ denotes the set of unordered pairs of elements (we adopt $\otimes$ for the direct product, *i.e.*, the set of ordered pairs, and elsewhere for the direct product of matrices) from V since we are restricting ourselves to undirected graphs, and we do not exclude the possibility of self-edges, that is choosing the same element of V twice. We will extensively use the notion of a graph product [11]. Suppose we have two graphs $G = (U,C)$ and $H = (V,D)$, then we define a new graph $K(W,E) = G \Delta H$ as a graph product of G and H, where $W = U \otimes V$, and the edge set E contains all $((u_1,v_1),(u_2,v_2))$, where $u_i \in U$ and $v_i \in V$, which satisfy some set of relationships which depend on the identity, adjacency or non-adjacency of u_1 and u_2, and of v_1 and v_2 [11].

We consider the following processes. The current graph is updated by adding to it a new vertex (the offspring) for each existing vertex (the parent). Each edge of the current graph is replaced by a subset of the edges of the complete graph formed on the pair of parent vertices and their two offspring; we always retain the edge between the parent vertices. Thus the "old" graph is always a subgraph of the "new" graph. The eight distinct ways in which this can be done constitute the set of models we consider (defined precisely below). Note further that there is no mortality in this model, all vertices and edges, once created are immortal. We shall discuss models in which the death of a vertex depends on the degree or the age of that vertex, and models in which interactions

(games) between neighbours determine the survival in subsequent papers.

2. The Models

We are interested in a family of sequences of graphs $G_t(V_t,E_t)$, where $t \in N$ which we shall refer to as time, V_t is the set of vertices and $E_t \subseteq V_t \otimes^* V_t$ the set of edges. We define a set of functions $F_i()$ for $i = 0,..,7$, which map graphs to graphs. In general we consider the sequences defined recursively by specifying G_0 and function $F_i(G)$; then $G_{t+1} = F_i(G_t)$. In each case we form G_{t+1} as a graph product of the existing graph G_t with a simple two vertex graph.

Suppose that $H_i = H_i(W_i,K_i) = F_i(G)$ for $G = G(V,E)$. Then H_i has vertex set $W_i = V \otimes \{0,1\}$. Thus each vertex of V gives rise to two vertices in W_i. We shall refer to the vertices $(u,0)$ and $(u,1)$ arising from $u \in V$ as the offspring vertex and parent vertex respectively. H_i has edge set K_i. Now if u and v are distinct elements of V, then $(u,v) \notin E \Rightarrow ((u,j),(v,j)) \notin K_i$ for all j and $(u,v) \in E \Rightarrow ((u,1),(v,1)) \in K_i$. We introduce three indicators (functions taking values 0 or 1), α, β and γ to specify the additional edges which are added to K_i. The index i of the eight functions $F_i(G)$ are written in binary and these define the three indicators for that model e.g. F_6 has $\alpha = 1$, $\beta = 1$ and $\gamma = 0$. Thus $F_i(G)$ for $i = 4\alpha + 2\beta + \gamma$ has edges as follows. If $u \in V$ then $((u,0),(u,1)) \in K_i$ if, and only if, $\beta = 1$. If $(u,v) \in E$ then $((u,0),(v,0)) \in K_i$ if, and only if, $\alpha = 1$. Finally $((u,0),(v,1)) \in K_i$ $((u,1),(v,0)) \in K_i$ if, and only if, $\gamma = 1$.

These models are illustrated in **Figure 1**.

We shall discuss here only the details of the eight models described above by using the eight F_i repeatedly. We could generalize the model in various ways, e.g. by taking some sequence $\{x_t\}$, possibly generated at random, of elements from $\{0,1,2,...,7\}$ and using the function F_{x_t} as the transition from G_t, or we could
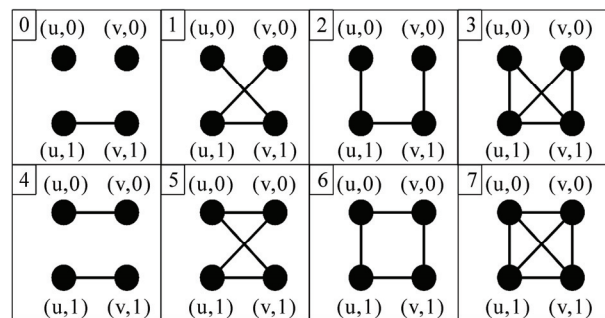

Figure 1. The motif that an edge {u, v} is replaced by under each model. The code for the models is shown at the top left of each panel.

AM

take the α, β and γ themselves to be probabilities that the corresponding edges are included at each stage.

We shall derive results for our models for the number of vertices with degree d, the total number of edges, the chromatic number, diameter and average distance, and comment briefly on the automorphisms. A further paper will explore additional graph entities such as cliques, and cycles, the number of polygons of given size arising from a given polygon at the previous time step, and add further information regarding some of the entities explored here.

As stated above the operations introduced here are all equivalent to taking a graph product of G_T with some simple graph Z. **Table 1** specifies the products for each of the models.

3. Binary String Representation

As mentioned above, a useful, alternative way to define the rules is in terms of the notion of parent and offspring vertices, and binary strings. If we begin with a graph $G_0(V_0, E_0)$, then if $u \in V_0$, this vertex gives rise to $(u,1)$ and $(u,0)$ in G_1, which we write as $u0$ and $u1$, and to $((u,1),1)$, $((u,1),0)$ $((u,0),1)$ and $((u,0),0)$ in G_2, which are written in the obvious way as $u11$, $u10$, $u01$ and $u00$, and so on. In G_t there will be 2^t vertices arising from u. We denote these vertices as strings of length $(t+1)$ written $u\tau$ where τ runs over the binary strings of length t. We refer to these representations as vertex strings.

The eight models, all of which at time t have $|V_0| \times 2^t$ vertex strings, give rise to distinct edge sets. We now specify precisely the edge set for each model at time t. Consider two vertices ux_t and vy_t (possibly identical) at time t, so x_t and y_t are two binary strings of length t, whose I'th elements are denoted by x_t^i and y_t^i. Now we define a third string z_t, where the i'th element of z_t, z_t^i, is determined by the pair (x_t^i, y_t^i). The purpose of z_t is to specify the sequence of edges which need to be added to u and v in order to reach ux_t and vy_t. In specifying the models earlier we introduced α, β and γ, as indicators for the three distinct types of new edge. Here we identify the elements of z_t with α, β, γ and two new terms γ^* and δ. Thus if we have $(x_t^i, y_t^i) = (0,0)$, indicating that an edge must be placed between the offspring of the individuals ux_{t-1} and vy_{t-1}, then we record $z_t^i = \alpha$. Similarly we track the other edges, as is detailed below. Note for ease we introduce a δ corresponding to the choice $(x_t^i, y_t^i) = (1,1)$, and differentiate between $(x_t^i, y_t^i) = (0,1)$ and $(x_t^i, y_t^i) = (1,0)$ by using γ and γ^* respectively. When we use the z_t's to specify which edges exist in G_t, we shall in fact take $\delta = 1$ always, and $\gamma = \gamma^*$.

Table 1. The products are denoted by a single letter K = Kronecker, C = Cartesian, H = Comb, S = Strong = AND, N = non-standard.

Model	α	β	γ	Product	Edges of Z
0	0	0	0	K	$\{(1,1)\}$
1	0	0	1	K	$\{(0,1),(1,1)\}$
2	0	1	0	H	$\{(0,1)\}$
3	0	1	1	N	N
4	1	0	0	K	$\{(0,0),(1,1),\}$
5	1	0	1	K	$\{(0,0),(0,1),(1,1)\}$
6	1	1	0	C	$\{(0,1)\}$
7	1	1	1	S	$\{(0,1)\}$

If
1) $(x_t^i, y_t^i) = (0,0)$ then $z_t^i = \alpha$,
2) $(x_t^i, y_t^i) = (1,1)$ then $z_t^i = \delta$,
3) $(x_t^i, y_t^i) = (0,1)$ or $(x_t^i, y_t^i) = (1,0)$ and $ux_{t-1} = vy_{t-1}$ then $z_t^i = \beta$,
4) $(x_t^i, y_t^i) = (0,1)$ and $ux_{t-1} \neq vy_{t-1}$ then $z_t^i = \gamma$,
5) $(x_t^i, y_t^i) = (1,0)$ and $ux_{t-1} \neq vy_{t-1}$ then $z_t^i = \gamma^*$.

The string $(u,v)z_t$ specifies the start and the sequence of operations which need to take place to progress from (u,v) to (ux_t, vy_t). As examples consider (A) vertices $(u0010101010)$ and $(u0011001110)$, then $(u,v)z_{10} = ((u,u)\alpha\alpha\delta\beta\gamma\alpha\delta\gamma^*\delta\alpha)$, and (B) $(u0011001011)$ and $(v0011001011)$ gives rise to $((u,v), \alpha\alpha\delta\delta\alpha\alpha\delta\alpha\delta\delta)$. Further note that z_t can contain at most one β, and then only if $u = v$.

Now we assert that ux_t and vy_t are linked for a specific model if, and only if, each of the entries in z_t (such as α, β, etc.) is equal to 1. If we start with $u = v$ then we obtain sequences of the form $z_t = ((u,u) < \alpha, \delta >^k \beta < \alpha, \gamma, \gamma^*, \delta >^{t-k-1})$, where the powers of the sets $<>$ are the direct products. Note here that α's in front of the β must be taken as having value 1 in every model since they relate to the same vertex. If we start with $u \neq v$ then we obtain sequences $z_t = ((u,v) < \alpha, \gamma, \gamma^*, \delta >^t)$.

There is one additional complication in the case where we have self-edges in G_0. We need to consider the ambiguities which may arise if $\beta = 1$ and $\alpha = 1$, since the former acting on a vertex, and the latter acting on a self edge at that vertex will result in the same edge. We can deal with this case efficiently by ensuring that any vertex with a self-edge at any time t is only subjected to one of the operators.

For our examples above we have that (A) requires $\alpha = \beta = \gamma = 1$ (recall $\delta = 1$ and $\gamma = \gamma^*$ always) so

there is an edge only in model 7, while (B) requires only that $\alpha = 1$ and $u \neq v$ (note that in the absence of a β, $u = v$ would only lead to two copies of the same vertex) so there is an edge in models 4, 5, 6 and 7.

4. Homogeneity

Definition. Merger of Two Graphs

Given two graphs $G(U, E)$ and $H(V, F)$ then we define the merger of G and H as the graph $J(U \cup V, E \cup F)$, and denote this by $G \uplus H$

Theorem

For each of the models specified above given some $G_0(V_0, E_0)$ then with $G_{t+1} = F(G_t)$ and writing in the obvious way $G_t = F^t(G_0)$ we have that,
$G_t = \uplus_{(u,v) \in E_0} F^t(L(u,v))$ where $L(u,v)$ is the graph with vertices u and v, and one edge (u,v).

Proof

This follows immediately from the definition of the functions $F_i(G)$. It is clear that for any $G(V, E)$ we have $F_i(G \uplus L(u,v)) = F_i(G) \uplus F_i(L(u,v))$ for each of the possible cases 1) $u \in V$, $v \in E$ and $(u,v) \in E$, 2) $u \in V$, $v \in V$ and $(u,v) \notin E$, 3) $u \in V$ and $v \notin V$, and 4) $u \notin V$ and $v \notin V$. Then by induction the result follows.

In view of the theorem above much of the information is captured by considering the case where G_0 consists of a single edge. We consider how a single edge (and sometimes other equally simple structures) evolve under our models.

4.1. Models 0, 1, 4 & 5; Kronecker Products

As **Table 1** shows, four of the models use the Kronecker product, in fact those for which $\beta = 0$. For such products, which we denote by $\otimes$, we have $A(G \otimes H) = A(G) \otimes A(H)$, where $A(G)$ denotes the adjacency matrix of G, and $\otimes$ also denotes the Kronecker product when applied to matrices. The adjacency matrices for the Z's of models are 0, 1, 4 and 5, respectively,

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The t th Kronecker powers of these, $2^t \times 2^t$ matrices, are easy to obtain. That for model 0 has a single 1 in the $(2^t, 2^t)$ position, and zeroes elsewhere, model 4 gives the identity matrix, model 5 gives a matrix of 1's. Only model 1 has an interesting pattern, which is essentially the bitwise AND pattern exhibited in [12], and which is discussed below in the Section 8.

4.2. Model 2

Model 2 is particularly simple as we have a tree structure;

we simply add a new branch at each vertex. Here we can capture all the structure by starting with G_0 as a single isolated vertex. There are various ways to describe the resulting tree, and these will be explored in more detail in a subsequent paper. For the moment we give only one such description. Starting from a single vertex labeled u, we obtain vertices $u1$ and $u0$ which are linked, then vertices $u11$, $u10$, $u01$ and $u00$. After t steps we have a tree with $2^t - 1$ edges on the vertices of the cube of dimension t. This tree is necessarily a spanning tree. As we make an extra step we take the current cube, with its spanning tree, and make a copy of the cube, join vertices of the $t-1$ dimensional cube to the matching vertex in the copy. An alternate way of expressing this is to consider a t-dimensional cube with all edges present. Choose a particular coordinate and remove all the edges from the cube for which this coordinate is 0, then move to the cube which has this coordinate equal to 1, and within this cube repeat the process. At each stage one removes all the edges of a cube, whose dimension is one smaller than at the previous step.

4.3. Models 3, 6 and 7

Model 3 is by far the most complex and interesting of the models. We shall discuss several aspects of this model, along with the other models, but shall postpone a fuller discussion to subsequent papers.

In model 6 the graph arising from a single edge after t steps is the $(t+1)$-dimensional cube.

In model 7 the graph arising from a single edge after t steps is the complete graph on 2^{t+1} vertices.

5. Numbers of Vertices and Edges

For a general $G_0(V_0, E_0)$ we have immediately that the number of vertices at time t is $|V_0| * 2^t$ for all models. The number of edges on the other hand depends on the particular model, and can be relatively easily deduced from the z_t possibilities and the α etc. appropriate for each model. For example, for model 3, we have $\beta = \delta = \gamma = \gamma^* = 1$, so for $u \neq v$ we have $z^t = ((u,v) < \delta, \gamma, \gamma^* >^t)$ so there are clearly 3^t edges for each (u,v). We have $z_t = ((u,u) < \alpha, \delta >^k \beta < \delta, \gamma, \gamma^* >^{t-k-1})$ and this results in $(3^t - 2^t)$ edges for each u. The complete set of formulae are given in **Table 2**.

6. Chromatic Number

A vertex colouring of a graph G is the assignment of a colour to each vertex in such a way that no adjacent vertices in G have the same colour. The minimal number of colours required to achieve this is the chromatic number,

Table 2. Formulae describing the number of edges after *t* time steps within the different models.

Model	Number of Edges
0	$\lvert E_0 \rvert$
1	$3^t \lvert E_0 \rvert$
2	$(2^t - 1)\lvert V_0 \rvert + \lvert E_0 \rvert$
3	$(3^t - 2^t)\lvert V_0 \rvert + 3^t \lvert E_0 \rvert$
4	$2^t \lvert E_0 \rvert$
5	$4^t \lvert E_0 \rvert$
6	$t * 2^{t-1} \lvert V_0 \rvert + 2^t \lvert E_0 \rvert$
7	$(2 * 4^{t-1} - 2^{t-1})\lvert V_0 \rvert + 4^t \lvert E_0 \rvert$

which we denote by $\chi(G)$. A colouring which achieves this minimum will be refered to as a minimal colouring. With the exception of model 7, the chromatic number of the growing graphs are easy to obtain.

6.1. Models 0, 1, 4 and 5

Suppose that we have a minimal colouring for G_0 with $\chi(G_0)$ colours. For models 0, 1, 4 and 5, each offspring can be given the same colour as its parent without violating the condition that we have a colouring, and so the chromatic number remains equal to $\chi(G_0)$.

6.2. Models 2 and 6

For models 2 and 6, suppose we have a minimal colouring of G_0 using the set $\{c_0, c_1, \ldots, c_{\chi(G_0)-1}\}$ of $\chi(G_0)$ colours. Now suppose that the offspring of a vertex coloured with c_i is coloured $c_{(i+1) \bmod \chi(G_0)}$. Then, provided $\chi(G_0) > 1$ this will constitute a minimal colouring for $R_i(G_0)$ for $i = 2$ and for $i = 6$. Thus

$$\chi(G_t) = max\big(2, \chi(G_0)\big).$$

6.3. Model 3

In this model $\chi(G_{t+1}) = \chi(G_t) + 1$. This is because in any minimal, proper colouring of G_t there will be an individual with $\chi(G_t) - 1$ distinct colours amongst its neighbours (actually there will be at least $\chi(G_t)$ such individuals, and since this individual's offspring is joined both to the individual and all its neighbours, this offspring must be given a new colour. This colour can then be given to every offspring.

6.4. Model 7

This is by far the most difficult case, and will be treated more fully elsewhere. We observe only that

$$\chi(G_t) + 1 \le \chi(G_t) \le 2\chi(G_t).$$

The first inequality follows since model 3 produces a subgraph of model 7 when they act on the same G. The latter inequality is evident since giving a minimal, proper colouring of G_t using colours $\{c_0, c_1, \ldots, c_{\chi(G_0)-1}\}$ we can colour the offspring of any vertex coloured c_i with some c_i^*, from a set $\{c_0^*, c_1^*, \ldots, c_{\chi(G_0)-1}^*\}$ of completely new colours. It is clear that if G_t is a clique, or bipartite, then the chromatic number doubles, but this doubling is not general. For example the chromatic number of any polygonal graph Q of odd degree > 3 is 3, while $\chi(F_7(Q)) = 5$.

7. The Distance Structure

We now turn to the details of the distances between vertices. The distance between vertices u and v is denoted by $d(u,v)$, the diameter of a graph g by $\vartheta(G)$. For a graph G_t we denote the numbers of pairs of vertices with distance x as $\eta_t(x)$. For each models we shall derive the recursions for the distances through time. Models 0 and 4 are excluded since they lead to disconnected components for which the notion of distance is inappropriate. We also suppose that our initial graph is connected so that all subsequent graphs are.

7.1. Model 2

We begin with model 2 since this will allow an easy demonstration of our methods. It is clear that $\vartheta(G_t) = \vartheta(G_{t-1}) + 2$. As in every model if $u \in V_t$ then $d(u0, u1) = 1$, while if $(u,v) \in E_t$ with $d(u,v) = d$ then $d(u1, v1) = d$, $d(u1, v0) = d(u1, v0) = d + 1$ and $d(u0, v0) = d + 2$. We can then write

$\eta_{t+1}(0) = 2\eta_t(0)$,

$\eta_{t+1}(1) = \eta_t(0) + \eta_t(1)$,

$\eta_{t+1}(2) = 2\eta_t(1) + \eta_t(2)$ and

$\eta_{t+1}(k) = \eta_t(k-2) + 2\eta_t(k-1) + \eta_t(k)$ if $k \ge 3$.

This enables us to derive closed form expressions for the $\eta_t(i)$, for example

$\eta_t(0) = 2^t \eta_0(0)$,

$\eta_t(1) = \eta_0(1) + (2^t - 1)\eta_0(0)$,

and

$\eta_t(2) = \eta_0(2) + 2t\eta_0(1) + 2(2^t - t - 1)\eta_0(0)$ but the forms rapidly become somewhat unmanageable.

AM

We can specify the total number of distances, for all models, $L_t = (4^t |V_0|^2 + 2^t |V_0|)/2$ being simply the number of pairs of vertices plus the number of vertices, and being the same for all the models. The total of the distances

$$L_t^* = 4L_{t-1}^* + 4L_{t-1} - 3\eta_{t-1}(0)$$
$$= 4^t L_0^* + 2^{2t-1} t \left(\eta_0(0)\right)^2 + \left(2^{t-1} - 2^{2t-1}\right)\eta_0(0).$$

The average distance $d_t = L_t^*/L_t$ and for t large this gives $d_t \approx c + t$ where $c = (2L_0^* - \eta_0(0))/(\eta_0(0))^2$. The average distance thus increases by 1 at each stage.

We can derive this average result in another, more direct way. Suppose at some stage we have n vertices and average distance μ_t. Now we add the extra offspring. The average distance is just the distance between a randomly selected pair of vertices. We consider the possible pairs in the new graph. There are $n(2n+1)$ such pairs, n are of type $(u0, u1)$ and contribute a distance 1. Of the remaining $2n^2$, 1/4 are of type $(u1, v1)$, 1/2 are of type $(u1, v0)$, and 1/4 are of type $(u0, v0)$, which contribute $d(u1, v1)$, $d(u1, v1) + 1$ and $d(u1, v1) + 2$ respectively. The average over these $2n^2$ latter will thus be $\mu_t + 1$ and overall we will therefore have

$$\mu_{t+1} = \frac{n + 2n^2\left(\mu_t + 1\right)}{n(2n+1)} = 1 + 2n\mu_t/(2n+1).$$

A similar argument can be used to obtain an expression for the variance of the distances which asymptotically increases by 1/2 per time step.

7.2. Model 1

In order to derive the appropriate expressions for model 1 we need to track not only the distances but also the number of edges which belong to triangles. Accordingly we define set $\nabla(G)$ of the edges which belong to a triangle (an edge $(u, v) \in E$ belongs to a triangle if there exists some k such that $(i, k) \in E$ and $(j, k) \in E$). Define $\eta_t^\nabla(1) = |\nabla(G)|$ and $\eta_t^*(1) = \eta_t(1) - \eta_t^\nabla(t)$. The necessity of considering triangles arises because the offspring of two linked parents will be distance 3 apart if the parents' link is not part of a triangle, but only 2 apart if there is a triangle since they are linked through the common neighbour of their parents.

In detail we have

For all $u \in V$ we have $d(u0, u1) = 2$,

and for $(u, v) \in E$,

if $d(u, v) > 1$ we have

$d(u0, v0) = d(u0, v1) = d(u1, v0) = d(u1, v1) = d(u, v)$,

if $(u, v) \in \nabla(G)$ then $d(u0, v0) = 2$, $d(u0, v1) = d(u1, v0) = 1$, $(u0, v1)\nabla(G)$ and $(u1, v0) \in \nabla(G)$,

and

if $(u, v) \in V \backslash \nabla(G)$ then $d(u0, v0) = 3$, $d(u0, v1) = d(u1, v0) = 1$, $(u0, v1) \notin \nabla(G)$ and $(u1, v0) \notin \nabla(G)$.

We have immediately that

$$\vartheta(G_t) = max(\vartheta(G_{t-1}), 3) = max(\vartheta(G), 3).$$

From these expressions we obtain recursions for the η's, as follows,

$$\eta_t(0) = 2\eta_{t-1}(0),$$
$$\eta_t^\nabla(1) = 3\eta_{t-1}^\nabla(1)$$
$$\eta_t^*(1) = 3\eta_{t-1}^*(1)$$
$$\eta_t(2) = \eta_{t-1}(0) + \eta_t^\nabla(1) + 4\eta_{t-1}(2)$$
$$\eta_t(3) = \eta_{t-1}^*(1) + 4\eta_{t-1}(3)$$
$$\eta_t(k) = 4\eta_{t-1}(k) \text{ if } k > 3.$$

From the above recursions we can find simple closed form expressions for the η's, as follows,

$$\eta_t(0) = 2^t \eta_0(0),$$
$$\eta_t^\nabla(1) = 3^t \eta_0^\nabla(1)$$
$$\eta_t^*(1) = 3^t \eta_0^*(1)$$
$$\eta_t(2) = (4^t - 2^t)\eta_0(0)/2 + (4^t - 3^t)\eta_0^\nabla(1) + 4^t \eta_0(2)$$
$$\eta_t(3) = (4^t - 3^t)\eta_0^*(1) + 4^t \eta_0(3)$$
$$\eta_t(k) = 4\eta_{t-1}(k) \text{ if } k > 3.$$

The total distance
$$L_t^* = 4^t L_0^* + (4^t - 3^t)(2\eta_0^*(1) + \eta_0^\nabla(1)) + (4^t - 2^t)\eta_0(0),$$
so that the average distance converges to a constant. The variance of the distances also converges to a constant.

7.3. Model 3

We obtain, in a straightforward manner that
$$\eta_t(0) = 2\eta_{t-1}(0),$$
$$\eta_t(1) = \eta_{t-1}(0) + 3\eta_{t-1}(1),$$
$$\eta_t(2) = \eta_{t-1}(1) + 4\eta_{t-1}(2),$$
$$\eta_t(k) = 4\eta_{t-1}(k) \text{ for } k > 2.$$
We have
$$\vartheta(G_t) = max(\vartheta(G_{t-1}), 2) = max(\vartheta(G), 2),$$
and the total distance
$$L_t^* = 4^t(L_0^* + \eta_0(0) + \eta_0(1)) - 3^t(\eta_0(0) + \eta_0(1)) \text{ so that}$$
the average distance converges to a constant. The variance of the distances also converges to a constant.

7.4. Model 5

We obtain $\eta_t(0) = 2\eta_{t-1}(0)$,
$$\eta_t(1) = 4\eta_{t-1}(1),$$
$$\eta_t(2) = \eta_{t-1}(0) + 4\eta_{t-1}(2),$$
$$\eta_t(k) = 4\eta_{t-1}(k) \text{ for } k > 2.$$
We have
$$\vartheta(G_t) = \vartheta(G),$$
and the total distance
$$L_t^* = 4^t(L_0^* + \eta_0(0)) - 2^t \eta_0(0) \text{ so that the average distance converges to a constant. The variance of the distances also converges to a constant.}$$

7.5. Model 6

We obtain $\eta_t(0) = 2\eta_{t-1}(0)$,
$\quad \eta_t(1) = \eta_{t-1}(0) + 2\eta_{t-1}(1)$,
$\quad \eta_t(k) = 2\eta_{t-1}(k-1) + 2\eta_{t-1}(k)$ for $k \geq 2$.
We have
$\quad \vartheta(G_t) = \vartheta(G_{t-1}) + 1$, and
$\quad L_t^* = 4L_{t-1}^* + 2L_{t-1} - \eta_{t-1}(0)$, from which we can easily demonstrate that the average distance L_t^*/L_t increases by 1/2 per time step for large t. The variance of the distances increases by 1/4 per time step for large t.

7.6. Model 7

We obtain
$\quad \eta_t(0) = 2\eta_{t-1}(0)$,
$\quad \eta_t(1) = \eta_{t-1}(0) + 4\eta_{t-1}(1)$,
$\quad \eta_t(k) = 4\eta_{t-1}(k)$ for $k \geq 2$.
From this we have that the diameter is constant, and
$\quad L_t^* = 4_t L_0^* + (4^t - 2^t)\eta_0(0)/2$,
so the average distance converges to a constant. The variance of the distances also converges to a constant.

8. Automorphism

In models 0, 1, 4 and 5 we have Kronecker products and this allows us to determine the automorphisms on vertices in G_t. In these models for a pair of vertices ur and vs, where u and v belong to G_0, and r and s are binary strings, to be automorphic in G_t, they must have u and v automorphic in G_0 and r and s to be automorphic in the Kronecker product of the appropriate Z. This is straightforward in each of the models. The most interesting case is model 1. For this the adjacency matrix is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

The n th Kronecker product, A_k, of this matrix have its rows and columns naturally indexed by the n th Kronecker products of vectors $(0,1)$ and $(0,1)^T$, and will have a zero wherever the row index and column index have a 0 in the same position. It immediately follows that the matrix A_k is invariant under any permutation to the elements of the row and column indices. Accordingly the permutations induce an equivalence relation over the binary strings; two strings being automorphic if they contain the same number of 0's.

9. Generating All Graphs

Now [4] proved that for any graph $H(U,F)$ of order n

there exists a set W of size $\leq n^2/4$ such that one can associate distinct subsets W_i with the n vertices, such that $(i,j) \in F$ if, and only if, $W_i \cap W_j = \phi$. Consider the case where we evolve a single vertex, linked to itself, for t time steps under model 1. The vertex set of the resulting graph will be the set of t length binary strings. Suppose each string x_t is associated with a set $S(x_t) = \{i \mid x_t^i = 0\}$. Then two vertices, x_t and y_t, are joined if, and only, $S(x_t) \cap S(y_t) = \phi$. It follows that every graph with n vertices is isomorphic to some subgraph of the t'th iterate when $t \geq \lceil n^2/4 \rceil$.

Now as pointed out in [4] the bound is exact only when the graph H is bipartite with vertices partitioned equally, when n is even, and differing by 1 when n is odd. Thus we will observe many graphs will appear at earlier stages, for example the graph K_n, the complete graph on n vertices, will appear at time $t = n-1$, since we may take $W_i = \{i\}$. We shall investigate this phenomenon elsewhere.

10. The Degree Distribution

Since we have a deterministic process the degrees of the vertices are well defined. We denote by $D(G,x)$ the number of vertices of degree x in graph G, and refer to the degree of any vertex x as $deg(x)$. We shall refer to the degree distribution by which we mean the distribution of the degree of a randomly chosen vertex.

10.1. Degree Distribution; Models 0, 1, 4, 5

We begin with the models which are Kronecker products. Given two graphs $J = (V,E)$ and $K = (W,F)$ with $y \in V$ with $deg(y)$ and $z \in W$ with $deg(z)$, then for $(y,z) \in J \otimes K$ we have $deg((y,z)) = deg(y)*deg(z)$. It follows that

$$D(J \otimes K, x) = \Sigma_{j \in \zeta(i)} D(J,j) D(K,i/j)$$

where $\zeta(i) = \{j \mid j \in N, j \mid i\}$, $j \mid i$ having the usual meaning that j divides i.

Now in these models we have $G_t = G_0 \otimes Z_t$, where Z_t is the graph obtained by taking the graph Z (as described in **Table 1**) and taking the Kronecker product of it with itself t times. By knowing the degree distribution of Z_t one can easily determine the distribution starting from a generic initial graph. Under model 0, Z_t has $(2^t - 1)$ isolated vertices and a single vertex with degree 1, model 1 has $_tC_r$ vertices with degree 2^r, model 4 has every vertex with degree 1, while model 5 has every vertex with degree 2^t.

10.2. Degree Distribution; Model 2

At each stage all the vertices of G_t increase their de-

gree by 1, and a set of $|G_t|$ vertices of degree 1 are added. Thus $D(G_t, x) = \sum_{r=0}^{t} (_t C_r) D(G_0, x - r)$.

10.3. Degree Distribution; Model 6

For model 6 we have a Cartesian product (which we denote $\odot$). For graphs $J = (V, E)$ and $K = (W, F)$ with $y \in V$ with $deg(y)$ and $z \in W$ with $deg(z)$, then for $(y, z) \in J \odot K$ we have $deg((y, z) = deg(y) + deg(z)$. Thus

$$D(J \odot K, x) = \sum_{j=0}^{i} D(J, j) + D(K, i/j)$$

and since the Cartesian power $\odot^t G(\{0, 1\}, (0, 1))$ is simply the t dimensional cube we have that

$$D(G_t, x) = 2^t D(G_0, x - t).$$

10.4. Degree Distribution; Models 3 and 7

As usual model 3 is the most interesting, and the most difficult model to deal with. A vertex $v \in G_t$ with degree x gives rise to $v1$ with degree $2x + 1$ and to $v0$ with degree $x + 1$. Thus

$$D(G_t, x) = D(G_{t-1}, x - 1) + D(G_{t-1}, (x - 1)/2$$

A plot of the frequency distribution of degrees for t = 17 is shown in **Figure 2**.

This distribution will be explored further in subsequent papers.

In model 7 a vertex $v \in G_t$ with degree x gives rise to two vertices, $v1$ and $v0$, each with degree $2x + 1$. If follows that

$$D(G_t, x) = 2^t D\left(G_0, \left(x + 1 - 2^t\right)\big/2^t\right)$$

11. Discussion

We have presented a variety of models for the growth of networks based on parent-offspring links and suggested that these might be used to describe the growth of interactions between individuals within a population.

This description might well be used to represent the binding of proteins under gene- and or genome-duplications. Alternately we might be describing the interactions within a population of organisms where these interactions depend on the interactions which existed amongst those in the previous generation, such as dominance relations (though these might require a directed graph approach).

The model as formulated has deliberately been kept as simple as possible. Thus the model is deterministic. The deterministic assumption would rarely apply in a biological context, but might in a computer context. We
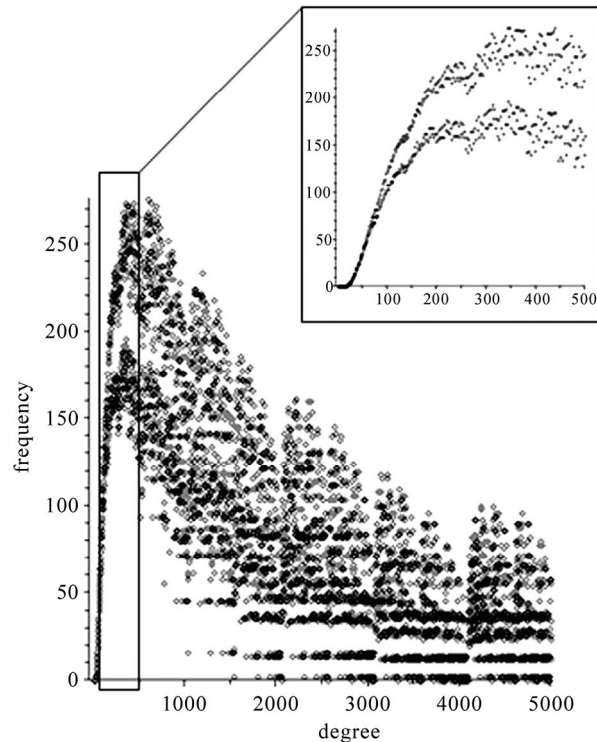

Figure 2. A truncated plot showing the frequencies of vertices with degree ≤ 5000 within the graph obtained by evolving a single edge for 17 time steps under model 3.

have indicated some ways in which a stochastic element can be introduced. One can vary the transition function applied as a stochastic process, or one can vary the links made by making the parameters of the models probabilities, rather than 0 or 1.

The simplicity of our model has allowed us to derive many results. Perhaps the most important of these is the theorem from Section 4. This theorem highlights the fact that the growth of any subgraph is independent of the nature of its exterior surroundings. By using this result and exploiting relations with graph products and binary strings we have derived formulas that describe chromatic number, distance structure and degree distributions.

The model has immortal vertices and edges. In subsequent papers we shall consider models with a similar reproductive structure, but allow for the death of vertices. In the next paper we shall treat the case where individuals have a fixed lifetime, and in the third we shall apply a threshold to the degree of a vertex, nodes which accumulate too high a degree will die. Naturally both of these processes could be made stochastic. In these models edges, once established through the reproductive phase disappear solely because of the death of one of their vertices. Additional features which we shall add in the future include sexual reproduction, and the embedding of the graph in space.

12. Acknowledgements

13. References

[1]   G. U. Yule, "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.," *Philosophical Transactions of the Royal Society of London* (*Series B*), Vol. 213, 1925, pp. 21-87.

[2]   H. A. Simon, "On a Class of Skew Distribution Functions," *Biometrika*, Vol. 42, No. 3-4, 1955, pp. 425-440.

[3]   A. L. Barab´asi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, Vol. 286, No. 5439, 1999, pp. 509-512.

[4]   B. Bollob'as, O. Riordan and G. Spenser, "The Degree Sequence of a Scale-Free Random Graph," *Random Structures and Algorithms*, Vol. 18, No. 3, 2001, pp. 279-290.

[5]   J. Jordan, "The Degree Sequences and Spectra of Scale-Free Random Graphs," *Random Structures and Algorithms*, Vol. 29, No. 2, 2006, pp. 226-242.

[6]   C. Cannings and A. W. Thomas, "Handbook of Statistical Genetics," John Wiley and Sons Ltd, New York, 2007.

[7]   J. S. Taylor and J. Raes, "Duplication and Divergence: The Evolution of New Genes and Old Ideas," *Annual Review of Genetics*, Vol. 38, No. 1, 2004, pp. 615-643.

[8]   A. Widdig, P. Nurnberg, M. Krawczak, W. J. Streich and F. B. Bercovitch, "Paternal Relatedness and Age Proximity Regulate Social Relationships among Adult Female Rhesus Macaques," *Proceedings of the National Academy of Science USA*, Vol. 98, No. 24, 2001, pp. 13769-13773.

[9]   G. Hausfater, "Dominance and Reproduction in Baboons (Papio Cynocephalus)," *Contributions to Primatology*, Vol. 7, 1975, pp. 1-150.

[10] L. Frank, K. Holekamp and L. Smale, "Serengeti II: Dynamics, Management, and Conservation of an Ecosystem," University of Chicago Press, Chicago, 1995.

[11] W. Imrich and S. Klavar, "Product Graphs: Structure and Recognition," John Wiley and Sons Ltd, New York, 2000.

[12] S. Wolfram, "A New Kind of Science," Wolfram Media, Inc., Champaign, 2002.

[13] P. Erd¨os, A. W. Goodman and L. Posa, "The Representation of a Graph by Set Intersections," *Canadian Journal of Mathematics*, Vol. 18, No. 1, 1966, pp. 106-112.