

Further Properties of Reproducing Graphs

Jonathan Jordan, Richard Southwell

School of Mathematics and Statistics, University of Sheffield, Sheffield, United Kingdom

E-mail: bugzsouthwell@yahoo.com

Received July 30, 2010; revised September 16, 2010; accepted September 18, 2010

Abstract

Many real world networks grow because their elements get replicated. Previously Southwell and Cannings introduced a class of models within which networks change because the vertices within them reproduce. This happens deterministically so each vertex simultaneously produces an offspring every update. These offspring could represent individuals, companies, proteins or websites. The connections given to these offspring depend upon their parent's connectivity much as a child is likely to interact with their parent's friends or a new website may copy the links of pre-existing one. In this paper we further investigate one particular model, 'model 3', where offspring connect to their parent and parent's neighbours. This model has some particularly interesting features, including a degree distribution with an interesting fractal-like form, and was introduced independently under the name Iterated Local Transitivity by Bonato *et al.* In particular we show connections between this degree distribution and the theory of integer partitions and show that this can be used to explain some of the features of the degree distribution; we give exact formulae for the number of complete subgraphs and the global clustering coefficient and we show how to calculate the minimal cycle basis.

Keywords: Reproduction, Graph, Network, Deterministic

1. Introduction

Networks are everywhere; wherever a system can be thought of as a collection of discrete elements, linked up in some way, networks occur. With the acceleration of information technology more and more attention is being paid to the structure of these networks, and this has led to the proposal of many models, for example Erdős-Rényi random graphs [1], preferential attachment graphs [2,3], random geometric graphs [4], and small world graphs [5].

In many situations networks *grow* – they expand in size as material is produced from the inside, not added from outside. To study network growth, in [6-8] Southwell and Cannings introduced a class of pure reproduction models, where networks grow because the vertices within reproduce. These models have some similarities to duplication models such as those analysed in [9,10], but in those models only one vertex (chosen randomly) reproduces at every update, whereas in our models every vertex reproduces at every time step.

We suggest that our models, or variations on them, can be applied to many situations where entities are introduced which derive their connections from pre-existing elements. Most obviously they could be used to model

social networks, collaboration networks, networks within growing organisms, the internet and protein-protein interaction networks. The model also captures aspects of the way collaboration networks change over time, with new students collaborating with their supervisors and the people their supervisor collaborates with.

In this paper we focus upon the dynamics of one of these eight pure reproduction models - labelled 'model 3' by the labelling in [7]. In this model, a graph is updated by simultaneously giving each vertex an offspring which is born connected to its parent and its parent's neighbours. This model was introduced independently in [11] under the name of Iterated Local Transitivity (ILT), proposed as a model for the growth of online social networks. The model has some very interesting mathematical properties, most particularly its degree distribution which has an intricate fractal-like form despite being generated by a simple recursion.

In this paper we study this degree distribution and reveal connections with the theory of integer partitions, which allows some of the features of the degree distribution to be explained. We also derive exact formulae for the number of complete graphs and the global clustering coefficient, which we compare with the asymptotic results on the local clustering coefficient in

[11]. We also show how the edge set of our evolving graphs can be described algebraically and how the minimal cycle basis can be calculated.

2. Definitions and Notation

Starting with an initial graph $G_0 = (V_0, E_0)$ we think of our model as generating a sequence G_0, G_1, G_2, \dots of graphs by performing repeated updates. G_{n+1} is obtained by simultaneously giving each vertex of G_n an offspring which is connected to its parent and its parent's neighbours.

When a graph $G_n = (V_n, E_n)$ is updated we use binary notation to keep track of how the vertices reproduce. For each $v \in V_n$ we write v 's offspring, in the updated graph, as $v0$ whilst v itself (surviving on as a parent) is written as $v1$. We refer to $v0$ and $v1$ as the descendants of v . The updated graph, G_{n+1} , will hence have vertex set $V_n \times \{0,1\}$ and $\{ux,vy\}$ will be an edge of G_{n+1} if and only if $u = v, x \neq y$ or $\{u,v\} \in E_0, (x,y) \neq (0,0)$.

As the graph continues to grow the binary prefixes of the vertices describe their life history, for example $v00$ is the grandchild of $v11$. Using this notation we can write any vertex of G_n in the form $v\xi_1, \xi_2, \dots, \xi_n$, where v is a vertex of G_0 and $\xi_i \in \{0,1\}$. Our choice to use binary labels allows us to describe G_n precisely in terms of the initial graph structure. This is described in the appendix.

Our choice to use binary labels allows us to describe G_n precisely in terms of the initial graph structure. In particular G_n has vertex set $V_n = V_0 \times \{0,1\}^n$ and edge set E_n such that $\forall u,v \in V_0 \quad \forall \xi, \pi \in \{0,1\}^n$ we have that $\{u\xi, v\pi\} \in E_n$ if and only if either $\{u,v\} \in E_0, \forall i \in \{1,2,\dots,n\} \quad (\xi_i, \pi_i) \neq (0,0)$ or $u = v, \exists k \in \{1,2,\dots,n\}$ such that $\forall i \in \{1,2,\dots,n\}$ we have $i < k \Rightarrow \xi_i = \pi_i, \quad i = k \Rightarrow \xi_i \neq \pi_i$ and $i > k \Rightarrow (\xi_i, \pi_i) \neq (0,0)$.

This can be shown by induction from the definition of our model.

2.1. The Degree Distribution

The degree distribution of G_n has a remarkably complex form despite being generated by a simple recursion. A degree x vertex $v \in V(G_n)$ will, in G_{n+1} , give a vertex $v1$ with degree $2x+1$ and a vertex $v0$ (the offspring) with degree $x+1$. A plot of the left-hand end of the degree distribution can be seen in **Figure 1**. In this section we will attempt to explain some of the features of this degree distribution, in particular the low frequency of certain degrees, which gives the appearance of parallel lines at the bottom of the degree distribution,

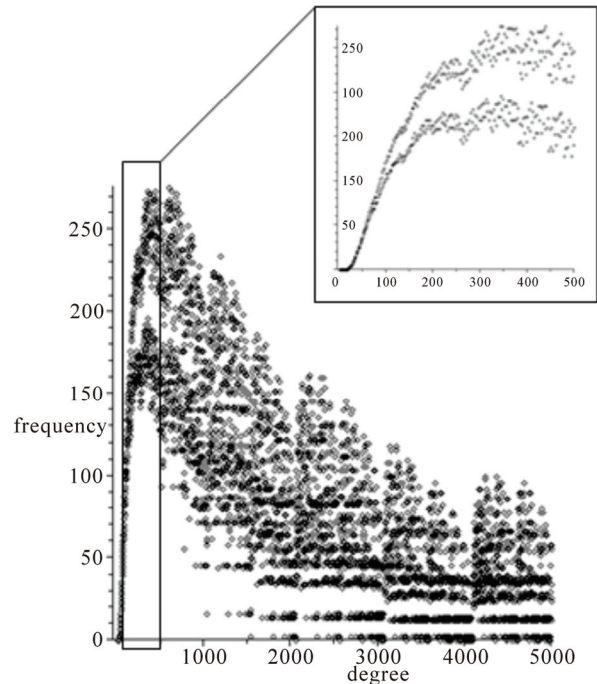


Figure 1. A truncated plot showing the frequencies of vertices with degree less than 5001 within the graph obtained by evolving an isolated vertex for 16 time steps.

and the apparent split in the distribution where odd degrees appear to be more frequent than even ones. We will see that some of these features can be obtained by linking the degree distribution to the theory of binary partitions of integers.

2.2. Links to Binary Partitions

We show that, in the case where G_0 is an isolated vertex, the number of vertices of degree d in G_n can be linked to the numbers of a particular type of binary partition of the integer d into n parts.

Types of partition. A **binary partition** of an integer n is a partition into powers of 2, *i.e.* of the form $n = \sum_{i=1}^k 2^{m_i}$. We will assume $m_i \geq m_{i+1}$, so that the terms are decreasing. The number of binary partitions of n is sequence A018819 in the Online Encyclopedia of Integer Sequences (OEIS, [12]), and the number of partitions of n into k parts is sequence A089052 (read as a triangle) in the OEIS [12].

A **non-squashing partition** of n is a partition $n = \sum_{i=1}^k m_i$ such that $m_j \geq \sum_{i=j+1}^k m_i$, *i.e.* that each part is at least the sum of all smaller parts. It is known that the number of non-squashing partitions of n is equal to the number of binary partitions, with an explicit bijection (which does not preserve the length) given by Sloane and Sellers in [13].

A similar idea to non-squashing partitions is that of a **strongly decreasing partition**, defined by Olsson in [14], which requires a strict inequality: $m_j > \sum_{i=j+1}^k m_i$.

We will define a **non-jumping binary partition** to be a binary partition with the restriction that $m_k = 1$ and that $m_i \leq m_{i+1} + 1$, i.e. that all integer powers of 2 less than 2^m are represented in the partition.

2.3. Recurrences for Partitions and the Degree Distribution

We link the degree distribution to two types of partition, strongly decreasing partitions and the non-jumping binary partitions defined above.

Theorem 1 *Let G_0 be an isolated vertex. Then*

1) *The number of vertices of degree d in G_n is twice the number of non-jumping binary partitions of d into n parts.*

2) *The total number of vertices of degree d in all graphs $G_i, i \geq 1$, is twice the number of strongly decreasing partitions of d (which is equal to the number of non-jumping binary partitions of d).*

Proof. To see this, we show that a bijection between the sets of strongly decreasing partitions and non-jumping binary partitions is naturally related to the representation of the vertices in the form $v_{\xi_1 \xi_2 \dots \xi_n}$. The idea of the bijection is essentially the same as that for binary partitions and non-squashing partitions given by Sloane and Sellers in [13].

Let $p_1 = q_1$ be the partition of 1 given by 1 (which is both a non-jumping binary partition and a strongly decreasing partition). This partition corresponds to the two vertices of degree 1 in G_1 . We then proceed inductively: we assume that we have a non-jumping binary partition p_j of length j and sum x and a strongly decreasing partition q_j with sum x . Then if $\xi_{j+1} = 0$, form a non-jumping binary partition p_{j+1} with sum $x+1$ and length $j+1$ from p_j by adding a new 1 part on the end, and form a new strongly decreasing partition q_{j+1} with sum $x+1$ by adding 1 to the largest part of q_j . If $\xi_{j+1} = 1$, form p_{j+1} by doubling every part of p_j and adding a new 1 part on the end, giving a non-jumping binary partition with sum $2x+1$ and length $j+1$, and form q_{j+1} by adding a new part on the beginning of q_j whose size is the minimum possible, i.e., $x+1$, giving a strongly decreasing partition with sum $2x+1$.

We now show that by working backwards it can be seen that each partition of each type arises in exactly one way from a sequence of $\xi_j, 2 \leq j \leq n$. If we have a non-jumping binary partition p_j with exactly one 1 part, set $\xi_j = 1$ and form p_{j-1} by removing the 1

part and dividing the remaining parts by 2; if p_j has more than one 1 part set $\xi_j = 0$ and form p_{j-1} by simply removing one 1 part. If we have a strongly decreasing partition q_j such that the largest part is equal to one more than the sum of the remaining parts, set $\xi_j = 1$ and form q_{j-1} by removing the largest part; otherwise the largest part will be larger than this, and we can set $\xi_j = 0$ and form q_{j-1} by subtracting one from the largest part.

As each sequence of $\xi_j, 2 \leq j \leq n$ corresponds to a pair of vertices of G_n with the same degree (one with $\xi_1 = 0$ and one with $\xi_1 = 1$) this shows that the number of vertices with degree d in G_n is twice the number of non-jumping binary partitions of d into n parts. The second part follows from the bijection between non-jumping binary partitions of d and strongly decreasing partitions of d .

The total number of strongly decreasing partitions of d (of any length), and hence also the number of non-jumping binary partitions of any length, is sequence A040039 in the OEIS, [12] (this can be seen by using the above recurrence). Hence, with appropriate modification based on the initial graph, this also gives the total number of vertices of degree d in all graphs G_n in total; as mentioned above if we start with an isolated vertex then A040039 needs to be doubled to give the right values, as G_1 contains two vertices of degree 1.

Features of the degree distribution. The plot of the degree distribution of G_n has a number of distinctive features. We attempt to use the connection to binary partitions to explain some of these, starting with the first G_n which contains vertices of degree d . Throughout this section we will assume that G_0 is an isolated vertex, so the results of Theorem 1 apply without modification. We start by finding the first graph G_n which contains a vertex of degree d .

Proposition 2 *Write d in the form $d = 2^r - 1 + s$ for some integers r and s such that $0 \leq s < 2^r$. Then the smallest n such that d occurs as a degree in G_n is r plus the number of ones in the binary representation of s , that is $r + \sum_{m=0}^{r-1} s_m$, where s_m is the digit corresponding to 2^m in the binary representation of s .*

Proof. Given d , the shortest non-jumping binary partition of d can have at most two occurrences of each power of 2. (If 2^m occurs three times, replace them by $2^{m+1} + 2^m$ to obtain a shorter non-jumping binary partition.) A unique shortest non-jumping binary partition of d can then be found by writing $d = 2^r - 1 + s$ for some integers r and s such that $0 \leq s < 2^r$, and forming the partition $d = \sum_{m=0}^{r-1} (s_m + 1)2^m$.

Hence the smallest n for which $d = 2^r - 1 + s$ is a possible value of X_n is r plus the number of ones in

the binary representation of s , that is $r + \sum_{m=0}^{r-1} s_m$.

Then d occurs as a degree for the first time in $G_{r + \sum_{m=0}^{r-1} s_m}$, with multiplicity 2 (as G_1 contains two vertices of degree 1).

We note that the stage at which d appears for the first time is given by A014701($d+1$) from the OEIS, [12]. Furthermore, the smallest d which appears as a degree for the first time in G_{2^r} will be $2^{r+1} - 2$, for which all digits of s are 1, and similarly, the smallest d which appears as a degree for the first time in $G_{2^{r-1}}$ will be $2^r + 2^{r-1} - 2 = 3 \cdot 2^{r-1} - 2$. These numbers give sequence A027383 in the OEIS.

We now show that arguments involving values $d = 2^r - 1 + s$ with small numbers of zeros in the binary representation of s , explain certain features of the plot of the degree distribution. The first part of the following proposition applies to G_n with even n , and the second part to G_n with odd n .

Proposition 3

1) If $d = 2^{r+1} - 2^t - 2$ for some $0 < t \leq r - 2$, there will be $2(r - 2)$ vertices with degree d in G_{2^r} , and if $d = 2^{r+1} - 3$ or $d = 2^{r+1} - 2^{r-1} - 2$ there will be $2(r - 1)$ vertices with degree d in G_{2^r} .

2) If $d = 3 \cdot 2^{r-1} - 2^t - 2$ for some $0 < t \leq r - 2$, there will be $2(r - 2)$ vertices with degree d in $G_{2^{r-1}}$, and if $d = 3 \cdot 2^{r-1} - 3$ or $d = 2 \cdot 2^{r-1} - 2$ there will be $2(r - 1)$ vertices with degree d in $G_{2^{r-1}}$.

Proof. If $d = 2^{r+1} - 2^t - 2$ for some $t \leq r - 1$ then $r + \sum_{m=0}^{r-1} s_m = 2r - 1$, so d appears as a degree for the first time in $G_{2^{r-1}}$. We can also consider the number of non-jumping binary partitions of d of length $2r$: these can be obtained by taking the partition

$d = \sum_{m=0}^{r-1} (s_m + 1)2^m$ and replacing one part of size 2^m by two of size 2^{m-1} , as long as either $s_m = 1$ or $t = r - 1$ so that the non-jumping definition is met. As $s_m = 1$ unless $m = t$, there will be $r - 2$ such partitions, or $r - 1$ in the cases $t = 0$ and $t = r - 1$, so if G_1 has two vertices of degree 1 these degrees will have multiplicity $2(r - 2)$ or $2(r - 1)$, giving the first part.

Similarly, if $d = 3 \cdot 2^{r-1} - 2^t - 2$ there will be $r - 2$ non-jumping binary partitions of length $2r - 1$, or $r - 1$ in the cases $t = 0$ and $t = r - 1$.

Similar arguments can also be constructed involving values $d = 2^r - 1 + s$ with other small numbers of zeros in the binary representation of s .

2.4. The Degree Distribution and Markov Chains

We note that we can choose a random vertex in each G_n by choosing a random vertex v of G_0 and defining a

sequence of independent Bernoulli (1/2) random variables $(\xi_i)_{i \in \mathbb{N}}$; we then let our random vertex be the vertex with label $v_{\xi_1 \xi_2 \dots \xi_n}$. We can then obtain the degree of our random vertex in G_n by letting X_0 be the degree of a random vertex in G_0 and $X_i = (\xi_{ni} + 1)X_{i-1} + 1$, and $(X_n)_{n \in \mathbb{N}}$ can then be considered as a Markov chain on the non-negative integers.

We note that if we let $Y_n = X_n \bmod m$ then $(Y_n)_{n \in \mathbb{N}}$ is a Markov chain on a finite state space $\{0, 1, \dots, m - 1\}$. The most interesting cases are where m is a power of 2; for example if $m = 2$ the transition matrix is

$$\begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

with stationary distribution (1/3, 2/3), implying that the probability that X_n is odd, and hence the proportion of vertices of G_n with odd degree, converges to 2/3 as $n \rightarrow \infty$. Similar results can be obtained for X_n modulo

m	Stationary distribution
2	$\frac{1}{3}(1, 2)$
4	$\frac{1}{15}(3, 4, 2, 6)$
8	$\frac{1}{255}(29, 40, 20, 44, 22, 28, 14, 58)$

other powers of 2, but if we work with X_n modulo an odd prime the transition matrix is doubly stochastic and so we get a uniform distribution as the stationary distribution.

2.5. Lognormal Limit

In this section we discuss some asymptotic properties of the degree distribution. It is shown in [11] that asymptotically the distribution has ‘‘binomial-type’’ behaviour, in the sense that $\binom{n}{j}$ nodes have degree around 2^j ,

which also implies that the degree distribution has a roughly lognormal form for large n ; the following result goes a bit further.

Theorem 4

As $n \rightarrow \infty$, we have

- 1) with probability 1, $\frac{\log X_n}{n} \rightarrow \mathbb{E} \log(1 + \xi) = \log(\sqrt{2})$;
- 2) $\frac{\log X_n - n \log(\sqrt{2})}{\sqrt{n} \log(\sqrt{2})} \rightarrow_d N(0, 1)$.

Proof. For $n \geq 1$, we can write

$$\log(X_{n+1}) = \log(X_n) + \log(1 + \xi_{n+1}) + \log\left(\frac{X_{n+1}}{X_{n+1} - 1}\right)$$

$$= \log(X_n) + \log(1 + \xi_{n+1}) + \log\left(1 + \frac{1}{X_{n+1} - 1}\right)$$

$$= \sum_{j=2}^{n+1} \left(\log(1 + \xi_j) + \log\left(1 + \frac{1}{X_j - 1}\right) \right).$$

This, applying the Strong Law of Large Numbers to $\sum_{j=2}^{n+1} \log(1 + \xi_j)$, is enough to show that

$\liminf_{n \rightarrow \infty} \log(X_n)/n \geq \log(\sqrt{2})$, almost surely. Hence, for any $a < \log(\sqrt{2})$ we have $X_n > \frac{a^n - 1}{a - 1}$ for n large enough, and so

$$\sum_{j=2}^{n+1} \log\left(1 + \frac{1}{X_j - 1}\right) < C + \sum_{j=2}^{n+1} \log\left(1 + \frac{1}{\frac{a^j - 1}{a - 1} - 1}\right)$$

for some C depending on a .

Note that if $z_1 = 1$ and $z_{n+1} = az_n + 1$ then

$$\log z_{n+1} = n \log a + \sum_{j=2}^{n+1} \log\left(1 + \frac{1}{z_j - 1}\right),$$

but $z_n = \frac{a^n - 1}{a - 1}$, so

$$\sum_{j=2}^{n+1} \log\left(1 + \frac{1}{\frac{a^j - 1}{a - 1} - 1}\right) = \log(a^{n+1} - 1) - n \log a - \log(a - 1)$$

$$< \log\left(\frac{a}{a - 1}\right).$$

Hence $\sum_{j=2}^{\infty} \log\left(1 + \frac{1}{X_j - 1}\right) = Y_{\infty}$, say, is finite, and we

can write

$$\log X_{n+1} = \sum_{j=2}^{n+1} \log(1 + \xi_j) + Y_{n+1}, \text{ with } Y_n \rightarrow Y_{\infty}.$$

Hence the Strong Law of Large Numbers tells us that, almost surely,

$$\frac{\log X_n}{n} \rightarrow \mathbb{E} \log(1 + \xi_1) = \log(\sqrt{2}),$$

and the Central Limit Theorem tells us that

$$\frac{\log X_n - n \log(\sqrt{2})}{\sqrt{n} \log(\sqrt{2})} \rightarrow_d N(0, 1).$$

2.6. Number of Complete Graphs

The following theorem gives an exact formula for $k_m^{(n)}$, the number of m vertex complete graphs in G_n , in terms of the number of complete graphs of various sizes within the initial graph.

Theorem 5

$$k_m^{(n)} = \sum_{h=1}^m (h+1)^n \cdot \binom{m-1}{h-1} \cdot (-1)^{m-h} \cdot \sum_{i=1}^h \binom{h-1}{i-1} \cdot k_i^{(0)}$$

Proof. Suppose the subgraph of G_n induced upon some $X \subseteq V_n$ is an m vertex complete graph. In the updated graph G_{n+1} there will be several complete graphs which are induced upon the descendants $X \times \{0, 1\}$ of vertices from X . In particular there will be m complete graphs on $m+1$ vertices (one induced upon $\{v0\} \cup \{u1 : u \in X\}$ for each $v \in X$) together with $m+1$ complete graphs on m vertices (one induced upon $\{u1 : u \in X\}$ and one induced upon $\{v0\} \cup \{u1 : u \in X - \{v\}\}$ for each $v \in X$).

We will now show that every m vertex complete graph in G_{n+1} can be constructed in the above manner. Suppose there is an m vertex complete graph in G_{n+1} induced upon $Y \subseteq V_{n+1}$. Let $X \subseteq V_n$ be minimal such that each vertex of Y is descended from a vertex of X . The subgraph of G_n induced upon X must be a complete graph since if it were not then a pair of its descendants in Y would not be linked. Since $|Y| = m$, we have $|X| \leq m$. Offspring of distinct vertices in X are not adjacent, this means there can be at most one $v \in X$ such that $v0 \in X$. This implies $|X| \geq m - 1$, and also that $Y = \{u1 : u \in X\}$ or $Y = \{v0\} \cup \{u1 : u \in X - \{v\}\}$ for some $v \in X$.

We have hence shown that the number, $k_m^{(n)}$, of m vertex complete graphs in G_n satisfies the difference equation $k_m^{(n+1)} = (m+1)k_m^{(n)} + (m-1)k_{m-1}^{(n)}$.

This system can be solved using linear algebra by writing

$$k^{(n+1)} = (k_1^{(n+1)}, k_2^{(n+1)}, \dots)^T = M \cdot k^{(n)}$$

where the infinite matrix M is such that $M_{i,j} = 0$ unless $i = j$, in which case $M_{i,j} = i + 1$, or $i = j + 1$ in which case $M_{i,j} = i - 1$.

For $h \geq 1$ the h th eigenvalue of this matrix is $\lambda_h = (h + 1)$, and the associated right eigenvector is

$$e^{(h)} = (e_1^{(h)}, e_2^{(h)}, \dots)^T$$

where

$$e_i^{(h)} = \begin{cases} \binom{i-1}{h-1} \cdot (-1)^{h-i} & \text{if } h \leq i \\ 0 & \text{otherwise} \end{cases}$$

Solving

$$k^{(0)} = \sum_{h \geq 1} \alpha_h \cdot e^{(h)}$$

we find that

$$\alpha_h = \sum_{i=1}^h \binom{h-1}{i-1} \cdot k_i^{(0)}.$$

Expanding

$$\sum_{h \geq 1} \alpha_h \cdot e^{(h)} \cdot (\lambda_h)^n$$

yields the result.

2.7. Global Clustering Coefficient

The global clustering coefficient of G_n is $\frac{\Delta_n}{\Delta_n + P_n}$

where Δ_n is the number of closed triplets (note that each triangle gets counted as three closed triples) and P_n is the number of open triples in G_n . We can describe the global clustering coefficient in terms of the numbers of vertices $k_1^{(0)}$, edges $k_2^{(0)}$, triangles $k_3^{(0)}$ and open triples P_0 in the initial graph with the following result.

Theorem 6

The global clustering coefficient of G_n is

$$\frac{9(2^n \cdot k_1^{(0)} - 2 \cdot 3^n (k_1^{(0)} + k_2^{(0)}) + 4^n (k_1^{(0)} + 2 \cdot k_2^{(0)} + k_3^{(0)}))}{8 \cdot 2^n \cdot k_1^{(0)} - 12 \cdot 3^n (k_1^{(0)} + k_2^{(0)}) + 5^n (4 \cdot k_1^{(0)} + 12 \cdot k_2^{(0)} + 9 \cdot k_3^{(0)} + 3 \cdot P_0)}$$

Proof. First, $\Delta_n = 3 \cdot k_3^{(n)}$. We shall derive a recursion for P_n . Let x be a vertex in G_n ; we shall describe the open triples in G_{n+1} which terminate at a descendant of x . Let $Ne(x) = \{y \in V_n : \{x, y\} \in E_n\}$ be x 's neighbourhood. Let $cl(x)$ be the set of $\{y, z\} \in E_n$ such that $\{y, z\} \subseteq Ne(x)$ (this corresponds to the set of closed triplets containing x). Let $op(x)$ be the set of (y, z) such that $y \in Ne(x)$ and $z \in Ne(y) - Ne(x)$ (this corresponds to the set of open triplets terminating at x).

- a) For $y \in Ne(x)$ there will be one open triple $\{x0, x1, y0\}$ in G_{n+1} .
- a') For $y \in Ne(x)$ there will be one open triple $\{x0, y1, y0\}$ in G_{n+1} .
- b) For $(y, z) \in op(x)$ there will be one open triple $\{x1, y1, z1\}$ in G_{n+1} .
- c) For $(y, z) \in op(x)$ there will be one open triple $\{x0, y1, z0\}$ in G_{n+1} .
- d) For $(y, z) \in op(x)$ there will be one open triple $\{x1, y0, z1\}$ in G_{n+1} .
- e) For $(y, z) \in op(x)$ there will be one open triple $\{x0, y1, z1\}$ in G_{n+1} .
- e') For $(y, z) \in op(x)$ there will be one open triple $\{x1, y1, z0\}$ in G_{n+1} .
- f) For $\{y, z\} \in cl(x)$ there will be one open triple $\{x0, y1, z0\}$ in G_{n+1} .

To see that every open triple T in G_{n+1} which terminates at a descendant of x will be of the type a,a',b,c,d,e,e' or f note that every such T consists of descendants of either 2 or 3 path connected vertices.

Suppose T consists of descendants of 2 path connected vertices. There are two possibilities. T could hold both descendants of x and one descendant d_y of $y \in Ne(x)$. In this case we must have $d_y = y0$ because the other possibility ($d_y = y1$) implies that T is the closed triple $\{x0, x1, y1\}$. So in this case we have that $T = \{x0, x1, y0\}$ is of type a. The other possibility is that T holds one descendant d_x of x and both descendants of $y \in Ne(x)$; this implies T is of type a'. in a similar way to before.

Suppose T consists of descendants of 3 path connected vertices, then either 1) $\exists(y, z) \in op(x)$ such that $T = \{x\alpha, y\beta, z\gamma\}$ or 2) $\exists\{y, z\} \in cl(x)$ such that $T = \{x\alpha, y\beta, z\gamma\}$, where $\alpha, \beta, \gamma \in \{0, 1\}$.

Let us consider case 1). We must have $(\alpha, \beta) \neq (0, 0)$ and $(\beta, \gamma) \neq (0, 0)$ in order for T to be connected. Let us consider the remaining possible values of (α, β, γ) . When $(\alpha, \beta, \gamma) = (1, 1, 1)$ we have case b. When $(\alpha, \beta, \gamma) = (0, 1, 0)$ we have case c. When $(\alpha, \beta, \gamma) = (1, 0, 1)$ we have case d. When $(\alpha, \beta, \gamma) = (0, 1, 1)$ we have case e. When $(\alpha, \beta, \gamma) = (1, 1, 0)$ we have case e'. These are all the possibilities under case 1).

Let us consider case 2). We require $\{x\alpha, y\beta\}, \{y\beta, z\gamma\} \in E_{n+1}$ and $\{y\gamma, z\alpha\} \notin E_{n+1}$. Again we must have $(\alpha, \beta) \neq (0, 0)$ and $(\beta, \gamma) \neq (0, 0)$ in order for T to be connected. In addition we require $\alpha = \gamma = 0$ to ensure $\{y\gamma, z\alpha\} \notin E_{n+1}$. These conditions force $(\alpha, \beta, \gamma) = (0, 1, 0)$, which is case f.

This means that we can enumerate the number of open triples in G_{n+1} by counting the number of a,a',b,c,d,e,e' and f which are generated by each $x \in V_n$. This yields $|E_n|$ open triples of type a, $|E_n|$ of type a', P_n of type b, P_n of type c, P_n of type d, P_n of type e, P_n of type e' and $\Delta_n = 3 \cdot k_3^{(n)}$ of type f.

This gives us the recursion $P_{n+1} = 2 \cdot |E_n| + 5 \cdot P_n + 3k_3^{(n)}$. Adding this to the recursion from the proof of theorem 5 we gain the matrix equation

$$\begin{pmatrix} k_1^{(n+1)} \\ k_2^{(n+1)} \\ k_3^{(n+1)} \\ P_{n+1} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 2 & 3 & 5 \end{pmatrix} \begin{pmatrix} k_1^{(n)} \\ k_2^{(n)} \\ k_3^{(n)} \\ P_n \end{pmatrix},$$

which we can solve to find

$$P_n = 2 \cdot 3^n (k_1^{(0)} + k_2^{(0)}) + 5^n \left(\frac{4}{3} k_1^{(0)} + 4k_2^{(0)} + 3 \cdot k_3^{(0)} + P_0 \right) - \frac{2^n \cdot k_1^{(0)}}{3} - 3 \cdot 4^n (k_0^{(0)} + 2 \cdot k_1^{(0)} + k_3^{(0)}),$$

and

$$k_3^{(n)} = 2^n \cdot k_1^{(0)} - 2 \cdot 3^n (k_1^{(0)} + k_2^{(0)}) + 4^n (k_1^{(0)} + 2 \cdot k_2^{(0)} + k_3^{(0)}).$$

Substituting these results into the formula, $\frac{3 \cdot k_3^{(n)}}{3 \cdot k_3^{(n)} + P_n}$, for the global clustering coefficient yields the stated result.

The global clustering coefficient decreases asymptotically as $(4/5)^n$ making it smaller than the local clustering coefficient, which decreases asymptotically as $(7/8)^n$, as shown in [11].

2.8. Minimal Cycle Basis

A (simple) cycle of a graph $G=(V,E)$ is a subset $C \subseteq E$ such that the graph with vertex set $\{u: \exists v, \{u,v\} \in C\}$ and edge set C is connected with each vertex of degree 2.

The set $\text{cyc}(G)$ of all cycles in a graph forms a vector space with symmetric difference as addition. When G is connected with a spanning tree T the fundamental cycles of T form a cycle basis (that is, a subset of $\text{cyc}(G)$ from which G can be generated). The dimension $\text{cyc}(G)$ is $|E| - |V| + 1$; a cycle basis is minimal when it holds this many elements.

Theorem 7 *If B is a minimal cycle basis of $G_n = (V_n, E_n)$ then*

$B' = B \times \{1\} \cup \{\{u1, u0\}, \{u0, v1\}, \{v1, u1\}\} : \{u, v\} \in E\}$
is a minimal cycle basis of $G_{n+1} = (V_{n+1}, E_{n+1})$

Proof. $|V_{n+1}| = 2|V_n|$ and $|E_{n+1}| = 3|E_n| + |V_n|$ so the dimension of the cycle space of G_{n+1} is $2|E_n|$ plus the dimension of G_n 's cycle space. This is equal to the number of elements in B' because B' consists of the minimal cycle basis of B plus two triangles, $\{\{u1, u0\}, \{u0, v1\}, \{v1, u1\}\}$ and $\{\{v1, v0\}, \{v0, u1\}, \{u1, v1\}\}$, for each edge $\{u, v\} \in E_n$. Let

$$\Delta = \{\{u1, u0\}, \{u0, v1\}, \{v1, u1\}\}$$

be a triangle in B' . Δ cannot be expressed as a linear combinations of other elements from B' since it holds an edge $\{u0, v1\}$ that occurs nowhere else. Elements $c \in B \times \{1\}$ cannot be expressed as a linear combinations of other elements from B' since they cannot be written as a linear combination of other elements of $B \times \{1\}$ (because this would invalidate the assumption that B is a minimal cycle basis of G_n). Suppose a linear combination equal to c contains a triangle, such as Δ , then the edge $\{u0, v1\}$ will be in c (since it cannot be removed by the addition of any other element from B'). This is a contradiction since each edge of c connects two parent nodes. Hence we have shown that B' is linearly independent with dimension equal to the cycle space of G_{n+1} . It follows that B' is a minimal

cycle basis of G_{n+1} .

3. Acknowledgements

The authors gratefully acknowledge support from the EPSRC (Grant EP/D003105/1) and helpful input from Dr. Seth Bullock and others in the Amorph Research Project (www.amorph.group.shef.ac.uk).

4. References

- [1] P. Erdős and A. Rényi, "On Random Graphs. I," *Publicationes Mathematicae*, Vol. 6, 1959.
- [2] 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*, London, Vol. B213, 1925, pp. 21-87.
- [3] R. Albert, A. Barabasi and H. Jeong, "Mean-Field Theory for Scale-Free Random Networks," *Physica A*, Vol. 272, 1999, pp. 173-187.
- [4] M. Penrose, "Random Geometric Graphs," Oxford University Press, Oxford, 2003.
- [5] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, Vol. 393, No. 6684, 1998, pp. 409-410.
- [6] R. Southwell and C. Cannings, "Games on Graphs that Grow Deterministically," In: *Proceedings of International Conference on Game Theory for Networks GameNets'09*, Istanbul, 2009, pp. 347-356.
- [7] R. Southwell and C. Cannings, "Some Models of Reproducing Graphs. 1 Pure Reproduction," *Applied Mathematics*, Vol. 1, 2010, pp. 137-145.
- [8] R. Southwell and C. Cannings, "Some Models of Reproducing Graphs. 2 Age Capped Vertices," *Applied Mathematics*, Vol. 1, No. 4, 2010, pp. 251-259.
- [9] F. Chung, L. Lu, T. Dewey and D. Gales, "Duplication Models for Biological Networks," *Journal of Computational Biology*, Vol. 10, No. 5, 2003, pp. 677-687.
- [10] N. Cohen, J. Jordan and M. Voliotis, "Preferential Duplication Graphs," *Journal of Applied Probability*, Vol. 47, No. 2, 2010, pp. 572-585.
- [11] A. Bonato, N. Hadi, P. Horn, P. Pralat and C. Wang, "Models of on-Line Social Networks," *Internet Mathematics*, 2010.
- [12] N. J. A. Sloane, "The On-Line Encyclopedia of Integer Sequences," 2009. <http://www.research.att.com/njas/sequences/>
- [13] N. J. A. Sloane and J. A. Sellers, "On Non-Squashing Partitions," *Discrete Mathematics*, Vol. 294, No. 3, 2005, pp. 259-274.
- [14] J. B. Olsson, "Sign Conjugacy Classes in Symmetric Groups," *Journal of Algebra*, Vol. 322, No. 8, 2009, pp. 2793-2800.