INFINITE LIMITS OF STOCHASTIC GRAPH MODELS - DRAFT

ANTHONY BONATO AND JEANNETTE JANSSEN

ABSTRACT. We present new generalized copying models for massive self-organizing networks such as the web graph, and analyze their limit behaviour. Our model is motivated by a desire to unify common design elements of the copying models of the web graph, and the partial duplication model for biological networks. In these models, new nodes copy (with some error) the link structure of existing nodes, and a certain number of random links may be added to the new node that can link to any of the existing nodes. In our new models, a function ρ parameterizes the number of random links, and thereby allows for the analysis of threshold behaviour. We study the infinite limits of graphs generated by our model, and compare and contrast these limits for different choices of ρ with properties of the infinite random graph.

1. INTRODUCTION AND BACKGROUND

Much recent attention has focused on stochastic models of real-world massive self-organizing networks; see for example, [1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18]. One of the most widely studied such network is the web graph, where the nodes represent web pages or sites, and the edges represent the links between pages, and certain biological networks, such as the network of protein interactions in a cell. In self-organizing networks, each node acts as an independent agent, which will base its decision on how to link to the existing network on local knowledge. As a result, the link neighbourhood of a new node will often be similar to that of an existing node.

The *copying models* of [1, 17] were proposed as models of the web graph, while the *partial duplication model* of [11] was designed to model biological networks. It has been shown that, with probability 1 as time tends to infinity, the copying model generates graphs whose in-degree distribution satisfies a *power law* with exponent $\gamma \geq 2$: for a positive

1991 *Mathematics Subject Classification.* 05C80, 68R10, 94C15.

The authors gratefully acknowledges support from NSERC research grants, and from a MITACS grant. The first author acknowledges support from a Wilfrid Laurier University Senior Research Fellowship.

integer k, the proportion of nodes of in-degree k is approximately $k^{-\gamma}$. With probability 1 as time tends to infinity, the partial duplication model generates graphs whose degree distribution satisfies a power law with an exponent in the range $1 < \gamma < 2$.

In the natural sciences, stochastic models are often studied by taking the infinite limit. Limiting behaviour can clarify the similarities and differences between different models, and show the consequences of the choices made in the design of the model. In some sense, the infinite limit magnifies the properties of the finite graphs that lead to it.

In [7], a number of *infinite* limits of finite graphs are studied, defined by properties derived from real-world self-organizing networks. An example of such a limit is the so-called *infinite random graph*, or R. If we fix any $p \in (0,1)$ and consider the limit of the graphs in $G(n,p)$ as n tends to infinity, then the resulting graph will with probability 1 be isomorphic to R. A view of convergence of R is converging to *white noise* (which is the combination of all frequencies of sound). For example, each countable graph is isomorphic to an induced subgraph of R. In this paper, we examine the limiting behaviour of a generalized copying model, concentrating especially on the conditions that cause convergence to R.

The copying and duplication models, although based on similar principles, produce graphs with different properties; in particular, they produce distinct ranges of power law exponents (which mirrors known experimental data). We introduce a generalized model, written $G(p, \rho, H)$, which serves to unify both models by placing them into a single framework.

The three parameters of the model $G(p, \rho, H)$ are a *copying probability* $p \in (0,1)$, a monotone increasing *random link function* $\rho : \mathbb{N} \to \mathbb{N}$, and a fixed finite *initial graph* H.

(1) At $t = 0$, set $G_0 = H$.
(2) For a fixed $t \geq 0$, assume that G_t has been defined, is finite, and contains G_0 as an induced subgraph. The nodes of G_t are the *existing nodes*. To form G_{t+1}, add node v_{t+1} named the *new node* to G_t and choose its neighbours as follows.
   (a) Choose an existing node u from G_t uniformly at random (u.a.r.). The node u is called the *copying node*.
   (b) For each neighbour w of u, independently add an edge from v_{t+1} to w with probability p. In addition, choose $\rho(t)$-many nodes from $V(G_t)$ u.a.r., and add edges from v_{t+1} to each of these nodes. The latter edges are called *random links*.

Note that in general the graphs G_t are not simple. We note that if $\rho(t) = 0$, then the graphs G_t generated by $G(p, \rho, H)$ correspond exactly to the graphs generated by the partial duplication model. If $\rho(t)$ is a constant, then the graphs G_t are undirected analogues of graphs generated by the copying model.

Infinite limits of graphs generated by the models $G(p, \rho, H)$ satisfy an adjacency property named locally e.c.; see Theorem 1. In [7], we showed that the locally e.c. defines an interesting infinite class of graphs, which have many properties in common with R. (In [7], the locally e.c. property was named property (B).) For example, a suggestive property true for all graphs which are locally e.c. is *inexhaustibility*: if any node is deleted, then the remaining graph is isomorphic to the original.

At each time-step t, our model adds approximately $\rho(t)$-many random links between the new node and the existing nodes. Theorem 2 shows that we need $\theta(t)$-many random links for the limit to be isomorphic to the infinite random graph. Our suspicion is that the case when $\rho(t) \in \theta(t^s)$, where $0 < s < 1$, represents a "grey area". As s tends to 1, the limit obtains more and more characteristics of R.

2. Limits and the $G(p, \rho, H)$ models

Before we state the main results for this section, we require a few definitions. If $(G_t : t \in \mathbb{N})$ is a sequence of graphs with G_t an induced subgraph of G_{t+1}, then define the *limit* of the G_t, written

$$G = \lim_{t\to\infty} G_t,$$

by

$$V(G) = \bigcup_{t\in\mathbb{N}} V(G_t),$$

and

$$E(G) = \bigcup_{t\in\mathbb{N}} E(G_t).$$

An analogous definition holds for limits of finite directed graphs. For example, the countably infinite clique is isomorphic to the limit of the sequence consisting of all finite cliques.

Let S be a finite set of nodes of $G = \lim_{t\to\infty} G_t$. We say that S is *born at time* t_0 if t_0 is the least positive integer t such that $S \subseteq V(G_t)$.

The infinite random graph R is the unique (up to isomorphism) countable graph satisfying the *existentially closed* or *e.c.* adjacency property. We say that the node $z \in V(G) \setminus (X \cup Y)$ is *correctly joined* or *c.j.* to X and Y, if z is joined to each node of X and no node of Y.

e.c. property: A graph G is e.c. if for each pair of finite disjoint subsets X and Y of nodes of G, there exists a node z c.j. to X and Y.

A logically weaker adjacency property introduced in [7] is *locally e.c.*, defined below. (In [7], locally e.c. is referred to as property (B).) If y is a node of G, then

$$N(y) = \{z \in V(G) : yz \in E(G)\}$$

is the *neighbour set* of y in G. Locally e.c. is a variant of the e.c. property that applies only to sets contained in the neighbour set of a node.

Locally e.c.: A graph G is locally e.c. if for each node y of G, for each finite $X \subseteq N(y)$, and each finite $Y \subseteq V(G) \setminus X$, there exists a node $z \neq y$ which is c.j. to X and Y.

Our model always generates limits satisfying the locally e.c. property.

Theorem 1. *Fix $p \in (0,1)$, $\rho(t) = \lfloor \alpha t^s \rfloor$ with α and s real numbers with $\alpha + p < 1$ and $s \in [0,1]$, and H a finite graph. With probability 1, the limit*

$$G = \lim_{t\to\infty} G_t$$

of graphs generated by the model $G(p, \rho, H)$ is locally e.c.

Before we prove Theorem 1, we first prove the following lemma that will be used several times in the rest of the paper. *We make the following simplification that will be used throughout the rest of the paper.* Note that if $|V(G_0)| = m$, then $|V(G_{t-1})| = m + t - 1$, where $t \geq 0$. Since $m+t \in \Theta(t)$ and we are only interested in asymptotic calculations with t, we will assume without loss of generality that

$$|V(G_t)| = t.$$

Lemma 1. *Let X be a fixed set of x nodes with the property that $X \subseteq V(G_t)$. If $t \geq 1$, then the probability $p_{ran}(t)$ that v_t is joined to all of X by random links satisfies*

$$\left(\frac{\rho(t) - x + 1}{t}\right)^x \leq p_{ran}(t) \leq \left(\frac{\rho(t)}{t - x + 1}\right)^x. \tag{2.1}$$

Proof. Since the $\rho(t)$ random links at time t are chosen uniformly at random from $V(G_{t-1})$, we have that

$$p_{ran}(t) = \frac{\binom{t-x}{\rho(t)-x}}{\binom{t}{\rho(t)}} = \frac{\rho(t)(\rho(t)-1)\cdots(\rho(t)-x+1)}{(t)(t-1)\cdots(t-x+1)}.$$

The proof now follows by estimating the fraction by the largest and smallest terms in the numerator and denominator. □

Proof of Theorem 1. Since a countable union of measure 0 subsets has measure 0, it suffices to show that for a fixed $y \in V(G)$, finite $X \subseteq N(y)$, and finite $Y \subseteq V(G) \setminus X$, the probability that there is no node c.j. to X and Y is 0 (since there only countably many choices for y, X, and Y in G). *We will apply this observation repeatedly to the proofs of all our theorems on adjacency properties of infinite limits, and so we tacitly assume it from now on.*

Let t_0 be the time that y, X, and Y are born in G, and fix $t > t_0$. If u is a node in $V(G_t)$, then we abuse notation and let $N(u)$ represent the nodes joined to u at time t. Define $Y_1 = Y \cap N(y)$ and $Y_2 = Y \setminus N(y)$. Let $|X| = x$, $|Y \cap N(y)| = y_1$ and $|Y \setminus N(y)| = y_2$. The probability that y is chosen as copying node in G_t equals $\frac{1}{t}$.

Assume that at time t, the node y is chosen as the copying node. Then v_t is joined to X with probability at least p^x. (Note the "at least" derives from the facts that an edge to a node of X may either happen with probability p if X is in $N(y)$, or from a random link.) By Lemma 1, with $S = Y_1$ and $s = y_1$, the probability that node v_t is joined to a fixed node of S by a random link is at most $\frac{\rho(t)}{t}$. Thus, the probability that v_{t+1} is joined to X and not Y_1 is at least

$$p^x \left(1 - \left(p + \frac{\rho(t)}{t}\right)\right)^{y_1}.$$

(Note that we need that $p + \frac{\rho(t)}{t} < 1$, which follows since $p + \alpha < 1$ by hypothesis.) The only way v_t may be joined to a node of Y_2 is by a random link, since Y_2 is not in the neighbourhood of the copying node y. Therefore,

$$\begin{aligned}
\mathbb{P}(v_t \text{ is c.j. to } X \text{ and } Y) &\geq p^x \left(1 - \left(p + \frac{\rho(t)}{t}\right)\right)^{y_1} \left(1 - \left(\frac{\rho(t)}{t}\right)\right)^{y_2} \\
&\geq p^x \left(1 - \left(p + \frac{\rho(t)}{t}\right)\right)^{y} \left(1 - \left(\frac{\rho(t)}{t}\right)\right)^{y} \\
&= p(t),
\end{aligned}$$

where $y = |Y|$. (Note that y_1 and y_2 will depend on as t increases. The last inequality removes this dependence on t.) We now simplify the term $p(t)$ depending on values of s.

Case 1. $s = 1$.

In this case, $\lfloor \alpha t \rfloor$. Hence,

$$\begin{aligned} p(t) &\geq p^x \, (1-(p+\alpha))^y \, (1-\alpha)^y \\ &\geq p^x \, (1-(p+\alpha))^{2y} \\ &= \beta, \end{aligned}$$

where β is a constant in $(0,1)$ that does node depend on t. (Recall that by hypothesis in our model, $p+\alpha<1$).

Case 2. $s<1$.

In this case, if $t>t_0$, then

$$\begin{aligned} p(t) &\geq p^x \left(1-\left(p+\frac{\lfloor \alpha t^s \rfloor}{t}\right)\right)^y \left(1-\left(\frac{\lfloor \alpha t^s \rfloor}{t}\right)\right)^y \\ &\geq p^x \left(1-\left(p+\frac{\alpha t^s}{t}\right)\right)^{2y} \\ &= p^x(1-p-o(1))^{2y} \\ &= p^x(1-p)^{2y}(1-o(1)). \end{aligned}$$

It follows that

$$\mathbb{P}(\text{no node of } G \text{ is c.j. to } X,Y) \leq \prod_{t=t_0}^{\infty}\left(1-\frac{q(t)}{t}\right), \tag{2.2}$$

where

$$q(t) = \begin{cases} \beta & \text{if } s=1, \\ p^x(1-p)^{2y}(1-o(1)) & \text{else.} \end{cases}$$

In either case, the infinite product in (2.2) converges to 0. □

The problem of determining when the limits of graphs generated by copying models converge to R was left open in [7]. The following theorem addresses this problem. For a positive integer n, a graph is n-e.c. if for each pair of disjoint subsets X and Y of nodes of G with $|X\cup Y|=n$, there exists a node z c.j. to X and Y. Observe that a graph is e.c. if and only if it is n-e.c. for all positive integers n.

Theorem 2. *Fix $p\in(0,1)$, H, and $\rho=\lfloor \alpha t^s \rfloor$, where α and s are non-negative real numbers with $\alpha, s\in[0,1]$ and $\alpha+p<1$. Let $G=\lim_{t\to\infty} G_t$ be generated according to the model $G(p,\rho,H)$.*

1. *If $s=1$ and $\lfloor \alpha t^s \rfloor \geq 1$ for all $t>0$, then with probability 1 G is isomorphic to R.*
2. *If $s\in[0,1)$ and $\lfloor \alpha t^s \rfloor \geq 1$ for all $t>0$, then with probability 1 G is $\lfloor \frac{1}{1-s} \rfloor$-e.c.*
3. *If $s\in[0,1)$, then with positive probability G is not isomorphic to R.*

Theorem 2 presents an example of *threshold* behaviour for convergence to R: with high probability, as s tends to 1, the limit G acquires more and more properties of R, but with positive probability is not itself isomorphic to R. At $s = 1$, we obtain R with high probability.

Proof of Theorem 2. For the proof of (1) we will use an adjacency property, introduced in [7] (and originally referred to as property (B,1)), which is an extension of the locally e.c. property to sets that have at most one node *not* in the neighbourhood of a node:

1-locally e.c.: For each node u of G, for each finite $X \subseteq N(u)$, for each finite $Y \subseteq V(G) \setminus X$, and for each node $w \in V(G) \setminus (N(u) \cup \{u\} \cup Y)$, there is a node z c.j to $X \cup \{w\}$, Y.

It was proved in Theorem 3 of [7] that any graph with property 1-locally e.c. is isomorphic to R. Hence, it is sufficient for us to prove that with probability 1, G satisfies the 1-locally e.c. property.

Fix nodes u and w, and sets X and Y as in the statement of the 1-locally e.c. property. Let t_0 be the time that u, w, X, and Y are born. Let $x = |X|$, $y_1 = |Y \cap N(u)|$, and $y_2 = |Y \setminus N(u)|$. For each $t \geq t_0$, let $B_{X,Y,w}(t)$ be the event that the new node v_t of G_{t+1} is correctly joined to $X \cup \{w\}$ and Y. Let $A_u(t)$ be the event that u is chosen as the copying node in G_t. By Lemma 1, if $t \geq 1$ the probability that a node w becomes a neighbour of v_t in step t is at most $\frac{\rho(t)}{t} \leq \alpha$ if $w \notin N(u)$, and at most $p + \alpha$ if $w \in N(u)$. The probability that an existing node w gets a new edge is at least the probability that w gets a random link, which is at least $\frac{\rho(t)}{t} \geq \alpha - \frac{1}{t}$ if $t > 0$. Choose $t_1 > t_0$ large, so for all $t > t_1$, we have that $\alpha - \frac{1}{t} \geq \frac{\alpha}{2}$.

Using these facts, we have that for $t > t_1$

$$\begin{aligned}
\mathbb{P}(B_{X,Y,w}(t) \mid A_u(t)) &\geq p^x \frac{\alpha}{2}(1-(p+\alpha))^{y_1}(1-\alpha)^{y_2} \\
&\geq p^x \frac{\alpha}{2}(1-p-\alpha))^{y} \\
&= \beta, \qquad (2.3)
\end{aligned}$$

where $y = y_1 + y_2$, and β is a suitable non-negative constant in $(0, 1)$ that does not depend on t. (We tacitly use here the fact that $p+\alpha < 1$.)

Thus, by (2.3) we have that

$$\begin{aligned}
& \mathbb{P}(\text{There is } \textit{no} \text{ node c.j. to } X \cup \{w\} \text{ and } Y \text{ in } G) \\
\leq & \prod_{t=t_0}^{\infty} 1 - \mathbb{P}(B_{X,Y,w}(t)) \\
= & \prod_{t=t_0}^{\infty} 1 - \mathbb{P}(B_{X,Y,w}(t) | A_u(t))\mathbb{P}(A_u(t)) \\
= & \prod_{t=t_0}^{\infty} \left(1 - \frac{\beta}{t}(1 + o(1))\right) \\
= & 0.
\end{aligned}$$

We now prove item (2). Let U be a set of x nodes born at time t_0. Choose $t_1 > t_0$ so that if $t > t_1$, then

$$\alpha t^{(s-1)} - \frac{1}{t} \geq \frac{\alpha}{2} t^{(s-1)}. \tag{2.4}$$

By Lemma 1, the probability that for $t > t_1$ the new node v_t is joined to all of U is at least

$$\begin{aligned}
\left(\frac{\rho(t) - x + 1}{t}\right)^x &= \left(\frac{\rho(t)}{t} - \frac{x-1}{t}\right)^x \\
&\geq \left(\frac{\rho(t)}{t}\right)^x - o(1) \\
&\geq \frac{\alpha}{2} t^{(s-1)x}(1 - o(1)), \qquad (2.5)
\end{aligned}$$

where the second inequality follows by the inequality $\frac{\rho(t)}{t} \geq \frac{\alpha t^s - 1}{t}$ and (2.4).

Let $x = \lfloor \frac{1}{1-s} \rfloor$. By (2.5), for each $t > t_1$, the probability that v_{t+1} is joined to all nodes in U is at least

$$\frac{\alpha}{2t}(1 - o(1)).$$

Then

$$\begin{aligned} & \mathbb{P}(\text{There is no node joined to all of } U) \\ \leq & \prod_{t=t_1}^{\infty} (1 - \mathbb{P}(B_{X,Y,w}(t)) \\ = & \prod_{t=t_0}^{\infty} \left(1 - \frac{\alpha}{2t}(1 - o(1))\right) \\ = & \; 0. \end{aligned}$$

Now, let X and Y be disjoint sets of nodes with $|X \cup Y| = \lfloor \frac{1}{1-s} \rfloor$, born at time t_0. By above, with probability 1, there is a node v^* born at time $t_2 > t_1$ joined to all of $U = X \cup Y$. If v^* is chosen as the copying node at time $t+1$, then the probability that the new node is c.j. to X, Y is $p^{|X|}(1-p)^{|Y|}$. Hence,

$$\begin{aligned} & \mathbb{P}(\text{There is no node c.j. to } X, Y) \\ \leq & \prod_{t=t_2}^{\infty} \left(1 - \frac{1}{t} p^{|X|}(1-p)^{|Y|}\right) \\ = & \; 0. \end{aligned}$$

We now turn to the proof of item (3). As $t - 1 \in \Theta(t)$, we may assume without loss of generality that $\rho(t) = \lfloor \alpha(t-1)^s \rfloor$. For a fixed subset $X \subseteq V(G)$, let $B_X(t)$ be the event that v_t is joined to X in step t, and let $\delta(X, t)$ be the number of nodes in G_t that are joined to all nodes in X. Note that $\mathbb{P}(B_X(t))$ and $\delta(X, t)$ are random variables, whose value depends on G_t. Let

$$p_{X,t} = \mathbb{E}(\mathbb{P}(B_X(t)))$$

and let

$$d(X, t) = \mathbb{E}(\delta(X, t)).$$

The proof of our theorem will depend on the following lemma, which bounds the number of nodes joined to all of X.

Claim: For all $t_0 > 0$, there exist constants c_1 and c_2 (depending only on t_0, α and s) so that for each set $X \subseteq V(G_{t_0})$, and for all $t \geq 2t_0$,

$$d(X, t) \leq c_1^{|X|} t^{(s-1)|X|+1},$$

and

$$p_{X,t} \leq c_2^{|X|}(t-1)^{(s-1)|X|}.$$

The proof of the claim is by induction on t. Let

$$c_3 = \max\{2\alpha, 1 + \left(\frac{1-p}{p}\right)(2\alpha)t_0^{s-1}\}.$$

Choose constants c_1 and c_2 so that the bounds hold for $t = t_0$ and all subsets of $V(G_{t_0})$, and so that $c_1 \leq c_2$, and

$$c_3 + pc_1 \leq c_1.$$

Fix $t \geq t_0$, and assume the bounds hold for all subsets of $V(G_{t_0})$ at time t. Fix $X \subseteq G_{t_0}$, and let $k = |X|$. The notation $N(w)$ refers to the neighbourhood of node w in G_t, while u denotes the copying node at time $t+1$.

We will break down the probability $\mathbb{P}(B_X(t))$ into cases, depending on the overlap of the neighbourhood of the node u with X. Namely,

$$\begin{aligned}\mathbb{P}(B_X(t)) &= \sum_{S\subseteq X} \mathbb{P}(B_X(t) \cap (N(u) \cap X = S)) \\ &= \sum_{S\subseteq X} \mathbb{P}(B_X(t) \,|\, N(u) \cap X = S)\mathbb{P}(N(u) \cap X = S)\end{aligned}$$

For any time $t > t_0$, and any set $S \subseteq V(G_{t_0})$ of cardinality r, let $A_S(t)$ be the event that every node of S is joined by a random link to the new node v_t.

For all $t \geq 2t_0$,

$$\begin{aligned}\mathbb{P}(A_S(t+1)) &= \frac{\binom{t-r}{\rho(t+1)-r}}{\binom{t}{\rho(t+1)}} \\ &\leq \frac{\frac{\lfloor \alpha t^s \rfloor !}{(\lfloor \alpha t^s \rfloor - r)!}}{\frac{t!}{(t-r)!}} \\ &\leq \left(\frac{\alpha t^s}{t-r}\right)^r \\ &= (\frac{t}{t-r})^r \alpha^r t^{(s-1)r} \leq (2\alpha)^r t^{(s-1)r}. \qquad (2.6)\end{aligned}$$

The last inequality follows from the fact that $r \leq t_0$ and $t \geq 2t_0 \geq 2r$, so $t - r \geq \frac{t}{2}$.

Let S be a fixed subset of X of cardinality r. (Note that $S \subseteq V(G_{t_0})$.) We obtain that for all $t \geq 2t_0$,

$$
\begin{aligned}
& \mathbb{P}(B_X(t+1) \mid N(u) \cap X = S) \\
= \quad & \sum_{Y \subseteq S} \binom{r}{|Y|} p^{|Y|}(1-p)^{r-|Y|}\mathbb{P}(A_{X-Y}(t+1)) \\
\leq \quad & \sum_{i=0}^{r} \binom{r}{i} p^i (1-p)^{r-i} (2\alpha)^{k-i} t^{(s-1)(k-i)} \\
= \quad & (2\alpha)^{k-r} t^{(s-1)(k-r)} (p + (1-p)(2\alpha)t^{s-1})^r \\
= \quad & (2\alpha)^{k-r} t^{(s-1)(k-r)} p^r \left(1 + \left(\frac{1-p}{p}\right)(2\alpha)t^{s-1}\right)^r \\
\leq \quad & c_3^k t^{(s-1)(k-r)} p^r. \qquad (2.7)
\end{aligned}
$$

The first inequality follows from (2.6), the second equation follows by the binomial theorem, and the last inequality follows from the definition of c_3. Now

$$
\mathbb{P}(N(u) \cap X = S) \leq \mathbb{P}(S \subseteq N(u)) = \frac{\delta(S,t)}{t}, \qquad (2.8)
$$

where the probabilities are considered to be conditioned on G_t; that is, the random variable representing the graph produced by the process up to time t.

Using (2.7) and (2.8) we obtain that

$$
\begin{aligned}
& \mathbb{P}(B_X(t+1)) \\
= \quad & \sum_{S \subseteq X} \mathbb{P}(B_X(t+1) \mid N(u) \cap X = S)\mathbb{P}(N(u) \cap X = S) \\
\leq \quad & \sum_{r=0}^{k} \binom{k}{r} c_3^{k-r} (t)^{(s-1)(k-r)} p^r \left(\frac{\delta(S,t)}{t}\right). \qquad (2.9)
\end{aligned}
$$

As before, probabilities are conditioned on G_t.

We complete the induction step for $p_{X,t+1}$ as follows.

$$\begin{aligned}
p_{X,t+1} &= \sum_{S\subseteq X} \mathbb{P}(B_X(t+1)\mid N(u)\cap X = S)\mathbb{P}(N(u)\cap X = S) \\
&\leq \sum_{r=0}^{k} \binom{k}{r} c_3^{k-r} t^{(s-1)(k-r)} p^r \left(\frac{d(S,t)}{t}\right) \\
&\leq \sum_{r=0}^{k} \binom{k}{r} c_3^{k-r} t^{(s-1)(k-r)} p^r c_1^r t^{(s-1)r}. \\
&= t^{(s-1)k} \sum_{r=0}^{k} \binom{k}{r} c_3^{k-r} (pc_1)^r \\
&= t^{(s-1)k}(c_3 + pc_1)^k \\
&\leq c_1^k t^{(s-1)k}.
\end{aligned}$$

The first inequality follows from (2.9), the monotonicity of expectation, and the linearity of expectation. The second inequality follows by the induction hypothesis. The third equality follows by the binomial theorem, and the last inequality follows by the choice of c_1.

In order to prove the bound on $d(X,t)$, we use the following difference equation:

$$\mathbb{E}(\delta(X,t+1)|G_t) = \delta(X,t) + \mathbb{P}(B_X(t+1)).$$

By taking the expectation of this equation, we have (by properties of conditional expectation and linearity of expectation) that

$$\begin{aligned}
d(X,t+1) &= d(X,t) + p_{X,t+1} \\
&\leq c_2^k t^{(s-1)k+1} + c_1^k t^{(s-1)k} \\
&\leq c_2^k (t+1)^{(s-1)k+1}.
\end{aligned}$$

The proof of the Claim follows.

Now let X be a fixed set of nodes satisfying

$$|X| > \lfloor \frac{1}{1-s} \rfloor.$$

Then $\lim_{t\to\infty} d(X,t) = 0$ as t goes to infinity. Therefore, with positive probability, there are no nodes joined to all of X in G, and so G is not e.c. □

References

[1] A. Adler, M. Mitzenmacher, Towards Compressing Web Graphs, In: *Proceedings of the Data Compression Conference*, 2001.

[2] W. Aiello, F. Chung, L. Lu, Random evolution in massive graphs, Handbook on Massive Data Sets, (Eds. James Abello et al.), Kluwer Academic Publishers, (2002) 97-122.
[3] R. Albert, A. Barabási, Emergence of scaling in random networks, *Science* **286** (1999) 509-512.
[4] B. Bollobás, O. Riordan, Mathematical results on scale-free graphs, Handbook of graphs and networks, S. Bornholdt, H. Schuster (eds) Wiley-VCH, Berlin (2002).
[5] B. Bollobás, C. Borgs, T. Chayes, O. Riordan, Directed scale-free graphs, submitted.
[6] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, The degree sequence of a scale-free random graph process, *Random Structures Algorithms* **18** (2001) 279-290.
[7] A. Bonato, J. Janssen, Infinite limits of copying models of the web graph, *Internet Mathematics* **1** (2003) 193-213.
[8] A. Bonato, J. Janssen, Limits and power laws of models for the web graph and other networked information spaces, accepted to the *Proceedings of Combinatorial and Algorithmic Aspects of Networking*, 2004.
[9] P.G. Buckley, D. Osthus, Popularity Based Random Graph Models Leading to a Scale-Free Degree Sequence, submitted.
[10] P.J. Cameron, The random graph, In: *Algorithms and Combinatorics* **14** (R.L. Graham and J. Nešetřil, eds.), Springer Verlag, New York (1997) 333-351.
[11] F. Chung, G. Dewey, D.J. Galas, Duplication models for biological networks, *Journal of Computational Biology*, **10** (2003) 677-688.
[12] F. Chung, L. Lu, The small world phenomenon in hybrid power law graphs, *Complex Networks*, (Eds. E. Ben-Naim et. al.), Springer-Verlag (2004) 91-106.
[13] F. Chung, L. Lu, Coupling on-line and on-line analyses for random power law graphs, submitted.
[14] C. Cooper, A. Frieze, On a general model of web graphs, *Random Structures Algorithms* **22** (2003) 311-335.
[15] C. Cooper, A. Frieze, J. Vera, Random deletions in a scale free random graph process, submitted.
[16] A. Fabrikant, E. Koutsoupias, C. Papadimitriou, Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet, In: *Proceedings of the 34th Symposium on Theory of Computing*, 2002.
[17] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal, Stochastic models for the web graph, *Proceedings of the 41th IEEE Symp. on Foundations of Computer Science*, 2000.
[18] S.H. Strogatz, D.J. Watts, Collective dynamics of 'small-world' networks, *Nature* **393** (1998) 440-42.

Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada, N2L 3C5
E-mail address: abonato@rogers.com

Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada, B3H 3J5
E-mail address: janssen@mathstat.dal.ca