Internet Mathematics Vol. 5, No. 1–2: 173–193

A Spatial Web Graph Model with Local Influence Regions

W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prałat

Abstract. We present a new stochastic model for complex networks, based on a spatial embedding of the nodes, called the *spatial preferred attachment* (*SPA*) model. In the SPA model, nodes have influence regions of varying size, and new nodes may link to a node only if they fall within its influence region. The spatial embedding of the nodes models the background knowledge or identity of the node, which will influence its link environment. In our model, nodes can determine their link environment based only on local knowledge of the network. We prove that our model gives a power-law in-degree distribution, with exponent in $[2, \infty)$ depending on the parameters, and with concentration for a wide range of in-degree values.

1. Introduction

Current stochastic models for complex networks, such as those described in [Bonato 08, Chung and Lu 06], aim to reproduce a number of graph properties observed in real-world networks such as the web graph. On the other hand, experimental and heuristic treatments of real-life networks operate under the tacit assumption that the network is a visible manifestation of an underlying hidden reality. For example, it is commonly assumed that communities in a social network can be recognized as densely linked subgraphs, or that web pages with many common neighbors contain related topics. Such assumptions imply that there is an a priori community structure or relatedness measure of the nodes, which is reflected by the link structure of the graph.

A common method to represent relatedness of objects is by an embedding in a metric space, so that related objects are placed close together, and communities are represented by clusters of points. Following a common text-mining

1542-7951/08 $0.50 per page

technique, web pages are often represented as vectors in a word-document space. Using latent semantic indexing, these vectors can then be embedded in a Euclidean topic space, so that pages on similar topics are located close together. Experimental studies [Menczer 04] have confirmed that similar pages are more likely to link to each other. On the other hand, experiments also confirm a large amount of topic drift: it is possible to move to a completely different topic in a relatively short number of hops. This points to a model in which nodes are embedded in a metric space, and the edge probability between nodes is influenced by their proximity.

The *spatial preferred attachment* (*SPA*) model proposed in this paper combines the above considerations with the often-used preferential attachment principle: pages with high in-degree are more likely to receive new links. In the SPA model, each node is placed in space and surrounded by an influence region. The volume of the influence region is determined by the in-degree of the node. The volume of each region is scaled by time, so the influence regions of nodes that do not gain new links will steadily decrease in size. The decrease in the volume of influence regions is motivated by the fact that the topic space grows over time. A new node v can link to an existing node u only if v falls within the influence region of u. If v falls within the influence region of u, then v will link to u with probability p. Thus, the model is based on the preferential attachment principle, but only implicitly: nodes with high in-degree have a large influence region, and therefore are more likely to attract new links.

A random graph model with certain similarities to the SPA model is the geometric random graph; see [Penrose 03]. In that model, all influence regions have the same size, and the link probability is $p = 1$. Flaxman, Frieze, and Vera supply an interesting geometric model in which nodes are embedded on a sphere and the link probability is influenced by the relative positions of the nodes [Flaxman et al. 07]. This model is a generalization of a geometric preferential attachment models presented by the same authors in [Flaxman et al. 06], which influenced our model. Other geometric models for complex models are now emerging, such as the inner product model; see, for example, [Young and Scheinerman 07].

There are at least two features that distinguish the SPA model from previous models. First, a new node can choose its links purely based on local information. Namely, the influence region of a node can be seen as the region where the associated entity (such as a web page or scholarly paper) is visible: only entities that are close enough (in topic) to fall within the influence region will be aware of its existence, and thus have a possibility to link to it. Moreover, a new node links independently to each node visible to it. Consequently, the new node needs no knowledge of the invisible part of the graph (such as in-degree of other nodes, or total number of nodes or links) to determine its neighborhood. Second, since a new node links to each visible node independently, the out-degree is not a constant nor chosen according to a predetermined distribution, but arises naturally from the model.

1.1. The SPA Model

We formally define the SPA model as follows. Fix parameters $m \in \mathbb{N}$, the dimension, and $p \in [0,1]$, the link probability. In addition, fix three positive constants A_1, A_2, and A_3 such that $pA_1 \leq 1$. Let S be the unit hypercube in $\mathbb{R}^m$, with the torus metric $d(\cdot,\cdot)$ derived from the L_∞ metric. In particular, for any two points x and y in S,

$$d(x,y) = \min\left\{\|x - y + u\|_\infty : u \in \{-1,0,1\}^n\right\}.$$

The torus metric is chosen so that there are no boundary effects, and altering the metric will not significantly affect the main results of the paper. The L_∞ norm is chosen so that every point on the boundary of the unit cube has equal distance 1/2 to the center of the hypercube. However, the norm could be easily replaced by any of the L_p norms, with changes only to some of the constants in our main results.

For each positive real number $\alpha \leq 1$, and $u \in S$, define the *ball around* u *with volume* α as

$$B_\alpha(u) = \{x \in S : \, d(u,x) \leq r_\alpha\},$$

where $r_\alpha = \alpha^{1/m}/2$, so r_α is chosen such that B_α has volume α.

The SPA model generates stochastic sequences of graphs $(G_t : t \geq 0)$, where $G_t = (V_t, E_t)$, and $V_t \subseteq S$. Let $d^-(v,t)$ be the in-degree of node v in G_t, and $d^+(v,t)$ its out-degree. We define the *influence region* of node v at time $t \geq 1$, written $R(v,t)$, to be the ball around v with volume

$$|R(v,t)| = \frac{A_1 d^-(v,t) + A_2}{t + A_3},$$

or $R(v,t) = S$ if the right-hand side is greater than 1.

The process begins at $t = 0$, with G_0 being the empty graph. Time step t, for $t \geq 1$, is defined to be the transition between G_{t-1} and G_t. At the beginning of each time step t, a new node v_t is chosen uniformly at random (uar) from S, and added to V_{t-1} to create V_t. Next, independently, for each node $u \in V_{t-1}$ such that $v_t \in R(u, t-1)$, a directed edge (v_t, u) is created with probability p. Thus, the probability that a link (v_t, u) is added in time step t equals $p|R(u,t-1)|$. See Figure 1 for a drawing of a simulation of the SPA model.

Because new nodes choose independently whether to link to each visible node, and the size of the influence region of a node depends only on the edges from younger nodes, the distribution of the random graph G_n produced by the SPA model with parameters A_1, A_2, A_3, p, m is equivalent to the graph G_{n+A_3} produced by the SPA model with the same values for A_1, A_2, p, m, but with $A_3 = 0$, where the first A_3 nodes have been removed. Since the results presented in this paper do not depend on the first nodes, we will assume throughout that $A_3 = 0$. In the rest of the paper, $(G_t : t \geq 0)$ refers to a sequence of random graphs

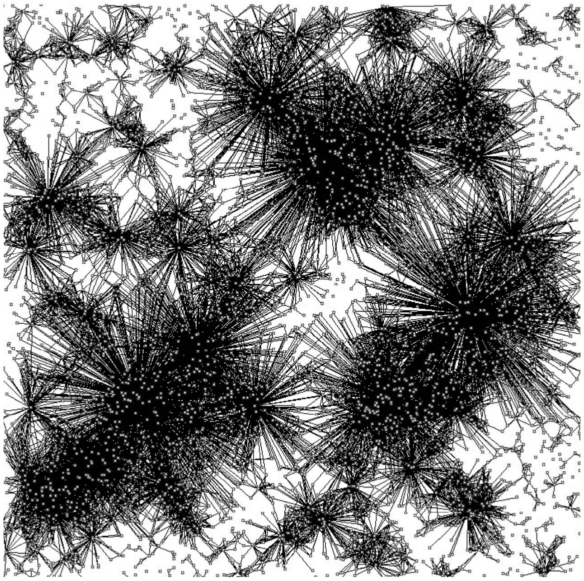

Figure 1. A simulation of the SPA model on the unit square with $t = 5000$, $p = 1$, and $A_1 = 1, A_2 = 0$.

generated by the SPA model with parameters A_1, A_2, p, and m, and we assume that $A_3 = 0$. We use the notation $[n]$ for $\{0, 1, \ldots, n\}$. All logarithms are in base e.

1.2. Main Results

We now state our main results on the SPA model, with proofs deferred to the next section. We first prove that with high probability a graph G_n generated by the SPA model has an in-degree distribution that follows a power law. See Figure 1 for the in-degree distribution of a simulation of the SPA model. We say that an event holds *asymptotically almost surely* (*aas*) if it holds with probability tending to one as $n \to \infty$. An event holds *with extreme probability* (*wep*) if it holds with probability at least $1 - \exp(-\Theta(\log^2 n))$ as $n \to \infty$. We will often use the stronger notion of wep in favor of the more commonly used aas, since it simplifies some of our proofs. If we consider a polynomial number of events that each holds wep, then all events hold wep. Let $N_{i,t}$ denote the number of nodes of in-degree i in G_t. For an integer $n \geq 0$, define

$$i_f = i_f(n) = \left(\frac{n}{\log^8 n} \right)^{\frac{pA_1}{4pA_1+2}}.$$

Theorem 1.1. *Fix $p \in (0,1]$. Then for any $i \geq 0$,*

$$\mathbb{E}(N_{i,n}) = (1+o(1))c_i n,$$

where

$$c_0 = \frac{1}{1+pA_2}, \tag{1.1}$$

and for $1 \leq i \leq n$,

$$c_i = \frac{p^i}{1+pA_2+ipA_1} \prod_{j=0}^{i-1} \frac{jA_1+A_2}{1+pA_2+jpA_1}. \tag{1.2}$$

For $i = 0, \ldots, i_f$, wep

$$N_{i,n} = (1+o(1))c_i n.$$

Since $c_i = (1+o(1))ci^{-(1+\frac{1}{pA_1})}$ for some constant c, this shows that for large i, the expected proportion $N_{i,n}/n$ follows a power law with exponent $1+\frac{1}{pA_1}$, with concentration for all values of i up to i_f. If $pA_1 = 10/11$, then the power law in-degree exponent is 2.1, the same as observed in the web graph (see [Bonato 08, Chung and Lu 06]).

The previous result characterizes the distribution of in-degrees in the graph. The total number of nodes of a given in-degree (smaller than i_f) is tightly concentrated around its mean. In the next result, we give a precise expression for the probability distribution of the in-degree of the individual node v_i born at time i, in the case that $pA_1 < 1$. No concentration result can be obtained here, but part (c) does give a bound on the maximum value that the in-degree of any particular vertex can reach.

For v_j the node added at time step j, let $d^-(v_j,n)$ be the in-degree of this node at the end of time step n.

Theorem 1.2. *If $0 < pA_1 < 1$, then the following hold:*

(a) *For $1 \leq j \leq n(1-\log^{-1} n)$ and $0 \leq l \leq \sqrt{j}\log^{-1} n$ or for $n(1-\log^{-1} n) < j < n$ and $l = 0,1$,*

$$\mathbb{P}(d^-(v_j,n) = l) = (1+O(\log^{-1} n))\binom{l+(A_2/A_1)-1}{l}\left(\frac{j}{n}\right)^{pA_2} \times \left(1-\left(\frac{j}{n}\right)^{pA_1}(1+O(\log^{-1} n))\right)^{l}.$$

(b) *For* $n(1-\log^{-1} n) < j < n$ *and* $l \geq 2$,

$$\mathbb{P}(d^{-}(v_j, n) = l) = O\left(\frac{l^{(A_2/A_1)-1}}{(\log n)^l}\right).$$

(c) *For all* $K > 0$,

$$\mathbb{P}\left(\text{There exists } j \leq n : d^{-}(v_j, n) \geq K(\log n)^2 (n/j)^{pA_1}\right) = O\left(n^{-Ke^{-18}}\right).$$

Theorem 1.2(c) implies (taking $K = \log^2 n$) that wep every node v_j has in-degree at most $(n/j)^{pA_1} \log^4 n$. If we are interested in an event that holds aas, then every node v_j has in-degree $O((n/j)^{pA_1} \log^2 n)$. Conditional on this, items (a) and (b) characterize the distribution of $d^{-}(v_j, n)$ for all $j \geq \log^8 n$ when $pA_1 \leq 1/2$, and for $j \geq n^{pA_1 - 1/2} \log^8 n$ when $pA_1 > 1/2$.

Let $M_t = |E_t|$, the number of edges in G_t, and let $m_t = \mathbb{E}(M_t)$. Then we have that

$$\mathbb{E}(M_{t+1} \mid M_t) = M_t + \sum_{j=1}^{t} p \frac{A_1 d^{-}(v_j, t) + A_2}{t} = M_t + \frac{pA_1 M_t}{t} + pA_2,$$

and so $m_1 = 0$, and for $t \geq 1$,

$$m_{t+1} = m_t \left(1 + \frac{pA_1}{t}\right) + pA_2.$$

The (first-order) solutions of this recurrence are

$$m_n \sim \begin{cases} \frac{pA_2}{1-pA_1} n, & pA_1 < 1, \\ n \log n, & pA_1 = 1. \end{cases}$$

Theorem 1.3. *If* $pA_1 < 1$, *then aas the number of edges is concentrated around its expected value:*

$$M_n = (1 + o(1)) m_n.$$

An important difference between the SPA model and many other models is that the out-degree is not a parameter of the model, but is the result of a stochastic process. Using the expression for m_n above, we can easily derive the expected out-degree of a vertex v_j. For example, this out-degree equals $pA_2/(1-pA_1) + o(1)$ if $pA_1 < 1$. Since the expected out-degree is small, we do not expect concentration. The next result gives bounds for the maximum out-degree in the graph.

Theorem 1.4. *Asymptotically almost surely,*

$$\max_{0\leq i\leq n} \deg^+(v_i, n) \geq (1+o(1))p\frac{\log n}{\log\log n}.$$

However, aas all nodes have out-degree $O(\log^2 n)$.

Theorem 1.5. *Asymptotically almost surely,* $\deg^+(v_n, n) = O(\log^2 n)$.

From Theorem 1.2, the number of nodes of in-degree zero in a graph generated by the SPA model in G_n is linear in n. In addition, with positive probability a new node will land in a part of S not covered by any influence regions, and thus have out-degree zero. Therefore, the underlying undirected graph of G_n is not connected. In fact, we expect that for the majority of distinct pairs u, v, there will not be a directed path from u to v.

2. Proofs of Results

This section is devoted to the proofs of the theorems outlined in the previous section.

2.1. Proof of Theorem 1.1

The equations relating the random variables $N_{i,t}$ are described as follows. Since G_1 consists of one isolated node, $N_{0,1} = 1$, and $N_{i,1} = 0$ for $i > 0$. For all $t > 0$, we derive that

$$\mathbb{E}(N_{0,t+1} - N_{0,t}|G_t) = 1 - N_{0,t}p\frac{A_2}{t}, \tag{2.1}$$

$$\mathbb{E}(N_{i,t+1} - N_{i,t}|G_t) = N_{i-1,t}p\frac{A_1(i-1)+A_2}{t} - pN_{i,t}\frac{A_1 i + A_2}{t}. \tag{2.2}$$

Recurrence relations for the expected values of $N_{i,t}$ can be derived by taking the expectation of the above equations. To solve these relations, we use the following lemma on real sequences, which is [Chung and Lu 06, Lemma 3.1].

Lemma 2.1. *If* (α_t), (β_t), *and* (γ_t) *are real sequences satisfying the relation*

$$\alpha_{t+1} = \left(1 - \frac{\beta_t}{t}\right)\alpha_t + \gamma_t,$$

and $\lim_{t\to\infty}\beta_t = \beta > 0$ *and* $\lim_{t\to\infty}\gamma_t = \gamma$, *then* $\lim_{t\to\infty}\alpha_t/t$ *exists and equals* $\gamma/(1+\beta)$.

Applying this lemma with $\alpha_t = \mathbb{E}(N_{0,t})$, $\beta_t = pA_2$, and $\gamma_t = 1$ gives that $\mathbb{E}(N_{0,t}) = c_0 t + o(t)$ with c_0 as in (1.1). For $i > 0$, the lemma can be inductively applied with

$$\alpha_t = \mathbb{E}(N_{i,t}), \quad \beta_t = p(A_1 i + A_2), \quad \text{and} \quad \gamma_t = \mathbb{E}(N_{i-1,t})\frac{A_1(i-1)+A_2}{t}$$

to show that $\mathbb{E}(N_{i,t}) = c_i t + o(t)$, where

$$c_i = c_{i-1} p \frac{A_1(i-1)+A_2}{1+p(A_1 i + A_2)}.$$

It is straightforward to verify that the expression for c_i as defined in (1.1) and (1.2) satisfies this recurrence relation.

We prove concentration for $N_{i,t}$ when $i \leq i_f$ using a relaxation of Azuma–Hoeffding martingale techniques. The random variables $N_{i,t}$ do not a priori satisfy the c-Lipschitz condition: it is possible that a new node may fall into many overlapping regions of influence. Nevertheless, we will prove that deviation from the c-Lipschitz condition occurs with exponentially small probability. The following lemma gives a bound for $|N_{i,t+1} - N_{i,t}|$ that holds with extreme probability.

Lemma 2.2. *With extreme probability, the following inequality holds for all* $0 \leq t \leq n-1$:

$$|N_{i,t+1} - N_{i,t}| \leq 2(A_1 i + A_2)\log^2 n, \quad \textit{for } 0 \leq i \leq t.$$

Proof. Fix t, let $i, j \leq t$, and let $X_j(i,t)$ denote the indicator variable for the event that v_j has degree i at time t and v_{t+1} links to v_j. It follows that

$$N_{i,t+1} - N_{i,t} = \sum_{j=1}^{t} X_j(i-1,t) - \sum_{j=1}^{t} X_j(i,t),$$

and so

$$|N_{i,t+1} - N_{i,t}| \leq \max\Big(\sum_{j=1}^{t} X_j(i-1,t), \sum_{j=1}^{t} X_j(i,t)\Big). \tag{2.3}$$

Let $Z_j(i,t)$ denote the indicator variable for the event that v_{t+1} is chosen in the ball of volume $(A_1 i + A_2)/t$ around node v_j. Clearly, if $X_j(i,t) = 1$, then $Z_j(i,t) = 1$ as well, so $X_j(i,t) \leq Z_j(i,t)$. Thus, to bound $|N_{i,t+1} - N_{i,t}|$ it suffices to bound the values of $Z(i,t)$, where

$$Z(i,t) = \sum_{j=1}^{t} Z_j(i,t).$$

The variables $Z_j(i,t)$ for $j = 1, \ldots, t$ are mutually independent. To see this, we can assume the position of v_{t+1} to be fixed. Then, the value of $Z_j(i,t)$ depends only on the position of v_j. Since the position of each node is chosen independently and uniformly, the value of $Z_j(i,t)$ is independent of the value of any other $Z_{j'}(i,t)$ where $j \neq j'$. Therefore, $Z(i,t)$ is the sum of independent Bernoulli variables with probability of success equal to

$$\mathbb{P}(Z_j(i,t) = 1) = \frac{A_1 i + A_2}{t}.$$

Using Chernoff's inequalities (see, for instance [Janson et al. 00, Theorem 2.1]), we can show that wep $Z(i,t) < A_1 i + A_2 + (A_1 i + A_2)\log^2 n < 2(A_1 i + A_2)\log^2 n$. Using these bounds, the proof now follows, since by (2.3),

$$|N_{i,t+1} - N_{i,t}| \leq \max(Z(i-1,t), Z(i,t)). \qquad \square$$

We mention that Theorem 1.5 can be used to improve the upper bound for $|N_{i,n} - N_{i,n-1}|$ to $O(\log^2 n)$, since the maximum change cannot be greater than the out-degree of vertex v_n.

To sketch the technique of the proof of Theorem 1.1, we consider $N_{0,t}$, the number of nodes of in-degree zero. We use the supermartingale method of [Pittel et al. 96], as described in [Wormald 99].

Lemma 2.3. *Let $G_0, G_1, \ldots, G_n$ be a random graph process and X_t a random variable determined by $G_0, G_1, \ldots, G_t$, $0 \leq t \leq n$. Suppose that for some real β and constants γ_i,*

$$\mathbb{E}(X_t - X_{t-1} \mid G_0, G_1, \ldots, G_{t-1}) < \beta$$

and

$$|X_t - X_{t-1} - \beta| \leq \gamma_i$$

for $1 \leq t \leq n$. Then for all $\alpha > 0$,

$$\mathbb{P}\big(\text{for some } t \text{ with } 0 \leq t \leq n : X_t - X_0 \geq t\beta + \alpha\big) \leq \exp\Big(-\frac{\alpha^2}{2\sum \gamma_j^2}\Big).$$

Theorem 2.4. *With extreme probability, for every $1 \leq t \leq n$, we have that*

$$N_{0,t} = \frac{t}{1 + A_2 p} + O(n^{1/2}\log^3 n) = c_0 t + O(n^{1/2}\log^3 n).$$

Proof. We first transform $N_{0,t}$ into something close to a martingale. It provides some insight if we define a real function $f(x)$ to model the behavior of the scaled random variable $\frac{1}{n}N_{0,xn}$. If we presume that the changes in the function correspond to the expected changes of the random variable (see (2.1)), we obtain the differential equation

$$f'(x) = 1 - f(x)\frac{pA_2}{x}$$

with the initial condition $f(0) = 0$. The general solution of this equation can be put in the form

$$f(x)x^{pA_2} - \frac{x^{1+pA_2}}{1+pA_2} = C.$$

Consider the real-valued function

$$H(x,y) = x^{pA_2}y - \frac{x^{1+pA_2}}{1+pA_2} \tag{2.4}$$

(note that we expect $H(t, N_{0,t})$ to be close to zero). Let $\mathbf{w}_t = (t, N_{0,t})$, and consider the sequence of random variables $(H(\mathbf{w}_t) : 1 \leq i \leq n)$. The second-order partial derivatives of H evaluated at $\mathbf{w}_t$ are all $O(t^{pA_2-1})$. Therefore, we have

$$H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) = (\mathbf{w}_{t+1} - \mathbf{w}_t) \cdot \operatorname{grad} H(\mathbf{w}_t) + O(t^{pA_2-1}), \tag{2.5}$$

where "$\cdot$" denotes the inner product and $\operatorname{grad} H(\mathbf{w}_t) = (H_x(\mathbf{w}_t), H_y(\mathbf{w}_t))$.

Observe that from our choice of H, we have that

$$\mathbb{E}(\mathbf{w}_{t+1} - \mathbf{w}_t \mid G_t) \cdot \operatorname{grad} H(\mathbf{w}_t) = 0.$$

Hence, taking the expectation of (2.5) conditional on G_t, we obtain that

$$\mathbb{E}(H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t) \mid G_t) = O(t^{pA_2-1}).$$

From (2.5), noting that

$$\operatorname{grad} H(\mathbf{w}_t) = \left(pA_2 t^{pA_2-1}N_{0,t} - t^{pA_2}, t^{pA_2}\right),$$

and using Lemma 2.2 (and the comment after the lemma) to bound the change in $N_{0,t}$, we have that wep

$$|H(\mathbf{w}_{t+1}) - H(\mathbf{w}_t)| \leq t^{pA_2}O(\log^2 n) + O(t^{pA_2}) = O(t^{pA_2}\log^2 n).$$

Now we may apply Lemma 2.3 to the sequence $(H(\mathbf{w}_t) : 1 \leq i \leq n)$, and symmetrically to $(-H(\mathbf{w}_t) : 1 \leq i \leq n)$, with $\alpha = n^{1/2+pA_2}\log^3 n$, $\beta = O(t^{pA_2-1})$, and $\gamma_t = O(t^{pA_2}\log^2 n)$, to obtain that wep

$$|H(\mathbf{w}_t) - H(\mathbf{w}_0)| = O(n^{1/2+pA_2}\log^3 n)$$

for $1 \leq t \leq n$. Since $H(\mathbf{w}_0) = 0$, this implies from the definition (2.4) of the function H, that wep

$$N_{0,t} = \frac{t}{1+pA_2} + O(n^{1/2}\log^3 n)$$

for $1 \leq t \leq n$, which finishes the proof of the theorem. □

We may repeat (recursively) the argument as in the proof of Theorem 2.4 for $N_{i,t}$ with $i \geq 1$. Since the expected change for $N_{i,t}$ is slightly different now (see (2.2)), we obtain our result by considering the following function:

$$H(x,y) = x^{p(A_1 i + A_2)} y - c_{i-1} \frac{p(A_1(i-1) + A_2)}{1 + p(A_1 i + A_2)} x^{1+p(A_1 i + A_2)}.$$

Using this function, we may show by similar arguments as in the case $i = 0$ that wep

$$N_{i,n} = c_i n + O(i n^{1/2} \log^3 n).$$

We therefore obtain concentration for all degrees i up to

$$i_f = \left(\frac{n}{\log^8 n} \right)^{\frac{pA_1}{4pA_1+2}},$$

since

$$\begin{aligned} i_f n^{1/2} \log^3 n &= n^{\frac{3pA_1+1}{4pA_1+2}} \log^{\frac{4pA_1+6}{4pA_1+2}} n \\ &= o\left(n^{\frac{3pA_1+1}{4pA_1+2}} \log^{\frac{4pA_1+6}{4pA_1+2}+1} n \right) \\ &= o\left(i_f^{-(1+\frac{1}{pA_1})} n \right) = o(c_{i_f} n). \end{aligned}$$

2.2. Proof of Theorems 1.2 and 1.3

We present the proofs of the results on the in-degrees of individual nodes and the number of edges.

Proof of Theorem 1.2. To simplify notation let $\eta = A_1 p$, $\nu = A_2 p$, and $\xi = A_2/A_1$. Let the node added at time step v be denoted by v, and treat the current time step (given as n above) as t. Let $\mathbb{P}(d^-(v,t) = l)$ denote the distribution of the in-degree of node v at the end of time step t.

The indicator variable $X(t+1)$ for an increase in $d^-(v,t)$ by receiving a link from v_{t+1} is a Bernoulli random variable with parameter $p(A_1 d^-(v,t) + A_2)/t$. Thus,

$$\mathbb{P}(X(t+1) = 0 \mid d^-(v,t) = j) = 1 - \frac{\eta j + \nu}{t}, \tag{2.6}$$

$$\mathbb{P}(X(t+1) = 1 \mid d^-(v,t) = j) = \frac{\eta j + \nu}{t}. \tag{2.7}$$

Let v, t be fixed, suppose $d^-(v,t) = l$, and let $\mathbf{T} = (T_j,\ j = 1, \ldots, l)$ denote the time steps T_j (if any) at which the degree of v changed. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_l)$

denote a particular value of $\mathbf{T}$, so that τ_j is the time step at which $d^-(v, \tau_j)$ changed from $j-1$ to j. For $v < \tau \leq t$ let

$$J = \{\boldsymbol{\tau} : \tau_1 < \tau_2 < \cdots < \tau_l\}$$

be the sequences of possible transitions. Hence,

$$\mathbb{P}(d^-(v,t) = l) = \sum_{\boldsymbol{\tau} \in J} \mathbb{P}(\mathbf{T} = \boldsymbol{\tau}).$$

Let

$$\Psi_j = \mathbb{P}(X(T) = 0, \text{ for all } \tau_j < T < \tau_{j+1}),$$

with $\tau_l < T \leq \tau_{l+1} = t$ when $j = l$. If $\tau_{j+1} = \tau_j + 1$, then let $\Psi_j = 1$. If $\tau_{j+1} \geq \tau_j + 2$, then from (2.6) we have that

$$\Psi_j = \prod_{\tau_j < T < \tau_{j+1}} \left(1 - \frac{\eta j + \nu}{T}\right).$$

Define $\omega = \log t$. Since $l \leq \sqrt{v}/\omega$, then $(\eta j + \nu)/t \leq (\eta l + \nu)/v = o(1)$, so that

$$1 - \frac{\eta j + \nu}{t} = e^{-\frac{\eta j + \nu}{t} - O\left(\frac{j^2}{t^2}\right)}.$$

Let

$$\delta(\tau, j) = j^2/\tau. \tag{2.8}$$

Then

$$\begin{aligned}\Psi_j &= \exp\left(\left(-(\eta j + \nu) \sum_{\tau_j < T < \tau_{j+1}} \frac{1}{T}\right) - O(\delta(\tau_j, j))\right)\\ &= \left(\frac{\tau_j}{\tau_{j+1}}\right)^{\eta j + \nu} (1 + O(\delta(\tau_j, j))).\end{aligned}$$

For $0 \leq j \leq l-1$, let $\Phi_j(t+1) = \mathbb{P}(X(t+1) = 1 \mid d^-(v,t) = j)$. Thus from (2.7), we have

$$\Phi_j(t+1) = \frac{\eta j + \nu}{t}.$$

Let $\Phi_j = \Phi(\tau_{j+1})$, and let $\Phi_l = 1$. Let $F(\boldsymbol{\tau})$ denote $\mathbb{P}(d^-(v,t) = l \text{ and } \boldsymbol{\tau})$. Let $\mathbb{P}(\boldsymbol{T}_j = \tau_j \mid \boldsymbol{T}_{j-1} = \tau_{j-1})$ be the probability that the transition to j occurs at τ_j given the transition to $j-1$ at τ_{j-1}. Hence,

$$F(\boldsymbol{\tau}) = \Psi_l \prod_{j=1}^{l} \mathbb{P}(\boldsymbol{T}_j = \tau_j \mid \boldsymbol{T}_{j-1} = \tau_{j-1}) = \prod_{j=0}^{l} \Psi_j\, \Phi_j.$$

Ignoring for the moment the multiplicative error terms, we see that $F(\boldsymbol{\tau})$ is given by

$$\left(\frac{v}{\tau_1}\right)^{\nu} \frac{\nu}{\tau_1} \left(\frac{\tau_1}{\tau_2}\right)^{\eta+\nu} \frac{\eta+\nu}{\tau_2} \cdots \left(\frac{\tau_{l-1}}{\tau_l}\right)^{\eta(l-1)+\nu} \frac{\eta(l-1)+\nu}{\tau_l} \left(\frac{\tau_l}{t}\right)^{\eta l+\nu}.$$

Recall that $\xi = \nu/\eta$. We cancel repeated values of τ_j to give

$$F(\boldsymbol{\tau}) = (1 + O(\delta(t,l))) \frac{\Gamma(l+\xi)}{\Gamma(\xi)} \left(\frac{v}{t}\right)^{\nu} \prod_{j=1}^{l} \frac{\eta\, \tau_j^{\eta-1}}{t^{\eta}} (1 + O(\delta(\tau_j, j))).$$

Thus,

$$\mathbb{P}(d^{-}(v,t) = l) = (1 + O(\delta(t,l))) \frac{\Gamma(l+\xi)}{\Gamma(\xi)} \left(\frac{v}{t}\right)^{\nu} P_1, \tag{2.9}$$

where

$$P_1 = \sum_{\boldsymbol{\tau} \in J} \prod_{j=1}^{l} \frac{\eta\, \tau_j^{\eta-1}}{t^{\eta}} (1 + O(\delta(\tau_j, j))).$$

For $b_j \geq 0$ we have that

$$\begin{aligned}(b_v + \cdots + b_t)^k - ({b_v}^2 + \cdots + {b_t}^2)\binom{k}{2}(b_v + \cdots + b_t)^{k-2} \\ \leq k! \sum_{i_1 < \cdots < i_k} b_{i_1} \cdots b_{i_k} \leq (b_v + \cdots + b_t)^k.\end{aligned}$$

Replace the term $\delta(\tau_j, j)$ in $F(\boldsymbol{\tau})$ with $\delta(\tau_j, l)$ and let

$$b_\tau = (1 + O(\delta(\tau, l))) \frac{\eta \tau^{\eta-1}}{t^{\eta}},$$

so that

$$P_1 = \frac{1}{l!} \left\{ (b_v + \cdots + b_t)^l - O(l^2)({b_v}^2 + \cdots + {b_t}^2)(b_v + \cdots + b_t)^{l-2} \right\}.$$

Using (2.8) and recalling that $l \leq \sqrt{v}/\omega$,

$$\begin{aligned} b_v + \cdots + b_t &= \sum_{v \leq \tau \leq t} \frac{\eta \tau^{\eta-1}}{t^{\eta}} (1 + O(\delta(\tau, l))) \\ &= 1 - \left(\frac{v}{t}\right)^{\eta} \left(1 - \left(\frac{v}{t}\right)^{\eta} O\left(\frac{l^2}{v}\right)\right) \\ &= 1 - \left(\frac{v}{t}\right)^{\eta} \left(1 + O\left(\tfrac{1}{\omega}\right)\right). \end{aligned}$$

An upper bound for P_1, and hence $\mathbb{P}(d^{-}(v,t) = l)$, follows.

For $1 \leq v \leq t(1-1/\omega)$ and $l \leq \sqrt{v}/\omega$ we prove below that

$$\sum_{\tau=v}^{t} b_\tau^2 = O\left(\frac{1}{\omega l^2}\right)\left(\sum_{\tau=v}^{t} b_\tau\right)^2. \tag{2.10}$$

We therefore have that

$$P_1 = \left(1+O\left(\tfrac{1}{\omega}\right)\right)\frac{1}{l!}\left(1-\left(\frac{v}{t}\right)^{\eta}\left(1+O\left(\tfrac{1}{\omega}\right)\right)\right)^{l}. \tag{2.11}$$

Inserting this estimate for P_1 into (2.9) completes the proof of Theorem 1.2(a). As remarked in the previous paragraph, (2.11) is an upper bound for P_1 for any $l \leq \sqrt{v}/\omega$, which completes the proof of Theorem 1.2(b).

Returning to the proof of (2.10), let

$$g(v,t) = \begin{cases} \frac{\eta^2}{1-2\eta}\left(\frac{1}{v}\left(\frac{v}{t}\right)^{2\eta}-\frac{1}{t}\right), & \eta < \frac{1}{2} \\ \frac{1}{4t}\log(t/v), & \eta = \frac{1}{2} \\ \frac{\eta^2}{2\eta-1}\left(\frac{1}{t}-\frac{1}{v}\left(\frac{v}{t}\right)^{2\eta}\right), & \eta > \frac{1}{2}. \end{cases}$$

Using $\delta = O(1/\omega)$ we have that

$${b_v}^2 + \cdots + {b_t}^2 = \left(1+O\left(\tfrac{1}{\omega}\right)\right)g(v,t).$$

It follows by direct examination that $vg(v,t) = O(1)$. Since $l \leq \sqrt{v}/\omega$, we have $l^2 g(v,t) = O(1/\omega^2)$. However, $\sum_i b_i \geq \Theta(1/\omega)$ for $v \leq t(1-1/\omega)$, and the result follows.

We now prove Theorem 1.2(c). Let $X_t = d^-(v,t)$. By Markov's inequality, for $h > 0$,

$$\mathbb{P}(X_t \geq \alpha) = \mathbb{P}(e^{hX_t} \geq e^{h\alpha}) \leq e^{-h\alpha}\mathbb{E}e^{hX_t}. \tag{2.12}$$

Let Y_t be an indicator variable for the increase of in-degree of v at time step $t+1$. Then $X_{t+1} = X_t + Y_t$, where

$$\mathbb{P}(Y_t = 1) = \frac{p(X_t+1)}{t+1},$$

and

$$\mathbb{E}\left(e^{hY_t} \mid X_t\right) = 1 + \frac{p(X_t+1)}{t+1}\left(e^h - 1\right).$$

Assume that $0 < h \leq 1$ (proved below in (2.14), so that $e^h \leq h + h^2$. Then

$$\begin{aligned} \mathbb{E}\left(e^{hX_{t+1}}\right) &= \mathbb{E}\left(e^{hX_t}e^{hY_t}\right) \\ &\leq \mathbb{E}\left(e^{hX_t}e^{\frac{p(X_t+1)}{t+1}\left(e^h-1\right)}\right) \\ &\leq e^{\frac{ph}{t+1}(1+h)}\mathbb{E}\left(e^{hX_t\left(1+\frac{p}{t+1}(1+h)\right)}\right). \end{aligned} \tag{2.13}$$

Let $\epsilon = 9/\omega$, and let

$$h = \frac{1}{\omega}\left(\frac{v}{t}\right)^{p(1+2\epsilon)}.$$

Let $h_t = h$ and for $v+1 \le s \le t$ define h_{s-1} by

$$h_{s-1} = h_s\left(1 + \frac{p}{s}(1+h_s)\right),$$

so that

$$h_s = h\prod_{\tau=s+1}^{t}\left(1+\frac{p}{\tau}(1+h_\tau)\right).$$

Let $\epsilon_\tau = \max(h_\tau : \tau = v, \dots, h)$ and assume (proved below in (2.14)) that $\epsilon_\tau < \epsilon < 1$.

Iterating expression (2.13) and noting that $\mathbb{E}e^{h_v X_v} = 1$ as $X_v = 0$, we have

$$\mathbb{E}e^{hX_t} \le \exp\left(p\sum_{s=v}^{t}\frac{h_s(1+h_s)}{s}\right) \le \exp\left(p(1+\epsilon)\sum\frac{h_s}{s}\right).$$

However, since $1/s + \cdots + 1/t \le 1/s + \log t/s$, we have

$$h_s \le h\exp\left(\sum_{\tau=s+1}^{t}\frac{p(1+\epsilon)}{\tau}\right) \le he^2\left(\frac{t}{s}\right)^{p(1+\epsilon)} \le \frac{e^2}{\omega} < 1, \tag{2.14}$$

for $t \ge 9$.

We therefore have that

$$\begin{aligned}
\mathbb{E}e^{hX_t} &\le \exp\left(hp(1+\epsilon)e^2t^{p(1+\epsilon)}\sum_{s=v}^{t}\frac{1}{s^{1+p(1+\epsilon)}}\right)\\
&\le \exp\left(h\left(\frac{t}{v}\right)^{p(1+\epsilon)}e^2\left(1+\frac{p(1+\epsilon)}{v}\right)\right)\\
&\le \exp\left(\frac{e^4}{\omega}\left(\frac{v}{t}\right)^{\epsilon p}\right)\\
&= 1 + O\left(\frac{1}{\omega}\right).
\end{aligned}$$

Let $\alpha = K\omega^2(t/v)^p$. By (2.12) and (2.14) we have that

$$\begin{aligned}
\mathbb{P}(X_t \ge \alpha) &= (1+o(1))e^{-h\alpha}\\
&= O(1)\exp\left(-K\omega\left(\frac{v}{t}\right)^{2p\epsilon}\right)\\
&= O\left(t^{-Ke^{-18}}\right).
\end{aligned}$$

This completes the proof of item (c), and completes the proof of Theorem 1.2. □

Proof of Theorem 1.3. We count the number of edges by counting the in-degree of nodes. Our approach is as follows: by Theorem 1.1, wep for $i \leq i_f$ the number of nodes $N_{i,n}$ of in-degree i at time n is concentrated.

Let a be the solution of $(n/a)^{pA_1} = i_f$ and let $\omega' = (K \log^2 n)^{1/(pA_1)}$ be the solution of

$$\left(\frac{n}{a\omega'}\right)^{pA_1} K \log^2 n = \left(\frac{n}{a}\right)^{pA_1},$$

where $K \geq 4e^{18}$. From Theorem 1.2(c), with probability $1 - O(n^{-3})$ no node $v \geq a\omega'$ has degree exceeding i_f. Let

$$\mu(n) = \sum_{i \leq i_f} \mathbb{E} N_{i,n} = (1 + o(1)) \sum_{i \leq i_f} N_{i,n},$$

and let

$$\lambda(n) = \sum_{j=1}^{a\omega'} d^-(v_j, n).$$

We prove, conditional on Theorem 1.2(c), that $\lambda(n) = o(m_n)$, and thus the number of edges is concentrated around m_n. We have that for $pA_1 < 1$,

$$\begin{aligned}
\lambda(n) &= \sum_{j=1}^{a\omega'} d^-(v_j, n) \\
&\leq K\omega^2 \sum_{j=1}^{a\omega'} \left(\frac{n}{j}\right)^{pA_1} \\
&= O\left(\omega^2 \left(\frac{n}{a\omega'}\right)^{pA_1} a\omega'\right) \\
&= O\left(n \left(\frac{n}{a}\right)^{pA_1 - 1} \log^{2/(pA_1)} n\right) \\
&= O\left(n \left(\frac{n}{\log^8 n}\right)^{(pA_1-1)/(4pA_1+2)} \log^{2/(pA_1)} n\right) \\
&= O\left(n^{(5pA_1+1)/(4pA_1+2)} \log^{2/(pA_1)} n\right) \\
&= o(n).
\end{aligned}$$

However, $\mu(n) \geq cn$ for some constant $c > 0$, so $\lambda(n) = o(\mu(n))$, and the assertion follows. □

2.3. Proof of Theorems 1.4 and 1.5

We now give the proofs of the results on out-degrees in the SPA model.

Proof of Theorem 1.4. Partition the interval $[0,1]$ into $\lceil 2(n/A_2)^{1/m} \rceil$ subintervals of equal length. Hence, the unit hypercube is partitioned into

$$h = 2^m n/A_2 + O(n^{(m-1)/m}) = (1+o(1))2^m n/A_2$$

identical hypercubes. (We expect each hypercube to contain a constant number of nodes.) We will show that aas there is a hypercube containing $\frac{\log n}{\log\log n}$ nodes.

Fix $c \in \mathbb{R}$ and suppose that

$$k = k(n) = \frac{\log n}{\log\log n}(1+c_n)$$

such that

$$\lim_{n\to\infty}(k+1/2)(\log k + m\log 2 - \log A_2 - 1) = \log n + c.$$

Note that $k = \frac{\log n}{\log\log n}(1 + O(\log\log\log n/\log\log n)) = (1+o(1))\frac{\log n}{\log\log n}$.

The probability q that any fixed hypercube contains exactly k nodes is equal to

$$\begin{aligned} q &= \binom{n}{k}\left(\frac{1}{h}\right)^k\left(1-\frac{1}{h}\right)^{n-k} \\ &= (1+o(1))\frac{n^k}{k!}\left(\frac{A_2}{2^m n}\right)^k \exp\left(-\frac{A_2}{2^m}\right) \\ &= (1+o(1))\frac{1}{k!}\left(\frac{A_2}{2^m}\right)^k \exp\left(-\frac{A_2}{2^m}\right). \end{aligned}$$

Using Stirling's formula $k! = (1+o(1))\sqrt{2\pi k}(k/e)^k$, we obtain that

$$\begin{aligned} q &= (1+o(1))\sqrt{\frac{2^m}{2\pi A_2}}\left(\frac{eA_2}{2^m k}\right)^{k+1/2} \exp\left(-\frac{A_2}{2^m}-\frac{1}{2}\right) \\ &= (1+o(1))\sqrt{\frac{2^m}{2\pi A_2}} \exp\left(-(k+1/2)(\log k + m\log 2 - \log A_2 - 1) - \frac{A_2}{2^m} - \frac{1}{2}\right) \\ &= (1+o(1))\sqrt{\frac{2^m}{2\pi A_2}} \exp\left(-\log n - c - \frac{A_2}{2^m} - \frac{1}{2}\right) \\ &= (1+o(1))\frac{1}{n}\sqrt{\frac{2^m}{2\pi A_2}} \exp\left(-c - \frac{A_2}{2^m} - \frac{1}{2}\right). \end{aligned}$$

It follows that the expected number of hypercubes with exactly k nodes is tending to

$$\lambda = hq = \frac{1}{\sqrt{2\pi}}\left(\frac{2^m}{A_2}\right)^{3/2} \exp\left(-c - \frac{A_2}{2^m} - \frac{1}{2}\right).$$

Now let A_i $(1 \leq i \leq h)$ denote the event that the ith hypercube contains exactly k nodes, and let $S_h = \sum_{i=1}^{h} I_{A_i}$ be the number of events that actually occur (S_h is a random variable). Finally, let

$$B_l^h = \sum_{1 \leq j_1 < \cdots < j_l \leq h} \mathbb{P}\left(\bigcap_{i=1}^{l} A_{j_i}\right).$$

We already showed that $\lim_{h\to\infty} B_1^h = \lambda$. It is also not difficult to see that for a fixed value of l,

$$\lim_{h\to\infty} B_l^h = \frac{\lambda^l}{l!}.$$

Therefore, S_h is tending to a random variable with Poisson distribution; that is,

$$\lim_{h\to\infty} \mathbb{P}(S_h = l) = \frac{\lambda^l}{l!} e^{-\lambda}.$$

In particular,

$$\lim_{h\to\infty} \mathbb{P}(S_h = 0) = e^{-\lambda}.$$

Since $c \to -\infty$ for $k = k_0 = \frac{\log n}{\log\log n}$, aas there is a hypercube K with k_0 points.

Since all nodes have the volume of the ball of influence at least A_2/n during the whole process up to time n (deterministically), the last node v added to K falls into balls of influence of all other nodes inside K (observe that the volume of K is at most $2^{-m}A_2/n$, so this holds even if v lies on the boundary of K). Thus, $\mathbb{E}\deg^+(v,n) \geq pk_0$.

To finish the proof, we use the fact that a sum of independent random variables with large enough expected value is not too far from its mean (see, for example, [Janson et al. 00, Theorem 2.8]). It follows that if $\epsilon \leq 3/2$, then

$$\mathbb{P}\big(|\deg^+(v,n) - \mathbb{E}\deg^+(v,n)| \geq \epsilon\mathbb{E}\deg^+(v,n)\big) \leq 2\exp\left(-\frac{\epsilon^2}{3}\mathbb{E}\deg^+(v,n)\right). \tag{2.15}$$

Setting

$$\epsilon = 1\Big/\sqrt[3]{\mathbb{E}\deg^+(v_i,n)},$$

we obtain that aas

$$\deg^+(v_i,n) = \big(1 + O(\epsilon)\big)\mathbb{E}\deg^+(v_i,n),$$

and the assertion follows. □

Proof of Theorem 1.5. Since the node v_n is chosen uar from the unit hypercube (note that the history of the process does not affect this distribution) with the torus metric, without loss of generality, we may assume that v_n lies in the center of the hypercube. For $1 \leq i < n$, let X_i denote the indicator random variable of the event that v_i lies in the ball around v_n (or vice versa) with volume

$$\alpha = 2i^{-pA_1} n^{pA_1 - 1} \log^2 n.$$

By Theorem 1.2(c), we have that aas

$$d^-(v_i, n) \leq (n/i)^{pA_1} \log^2 n,$$

for all $i \in [n]$. Hence, aas for all $i \in [n-1]$, $X_i = 0$ implies that v_n is not in the influence region of v_i and there is no directed edge from v_n to v_i. Therefore, aas we have that

$$\deg^+(v_n, n) \leq \sum_{i=1}^{n-1} X_i.$$

Since

$$\begin{aligned} \mathbb{E}\Big(\sum_{i=1}^{n-1} X_i \Big) &= \sum_{i=1}^{n-1} O(i^{-pA_1} n^{pA_1 - 1} \log^2 n) \\ &= O\Big(n^{pA_1 - 1} \log^2 n \sum_{i=1}^{n-1} i^{-pA_1} \Big) \\ &= O(\log^2 n), \end{aligned}$$

the assertion follows from the Chernoff bound (see (2.15)). □

3. Generalizations

Several variants of the SPA model may be proposed, and for each variant, it would be interesting to pursue a rigorous analysis of the degree distributions. One such variation is the generalized SPA (or GSPA) model, which allows more control of the out-degree. In the GSPA model, nodes are distributed on the hypercube as in the SPA model, but now receive two regions of influence. Each node v at time step t is assigned both an in-degree influence region with volume

$$\frac{A_- + B_- d^-(v,t)}{t},$$

where A_- and B_- are nonnegative constants and $d^-(v,t)$ is the in-degree of v at time t, and an out-degree influence region with volume

$$\frac{A_+ + B_+ d^+(v,t)}{t},$$

where A_+ and B_+ are nonnegative constants and $d^+(v,t)$ is the out-degree of v at time t.

Edges are now added with probability p between any pair of nodes whose regions interact by a predetermined rule. An important difference between this and the SPA model is that at every time step all nodes can potentially receive out- and in-edges. This implies that graphs generated by the GSPA model can have cycles, and edges that go from younger to older nodes.

We describe three rules for the generation of edges.

Intersection rule. If the in-degree influence region of node v has a nonempty intersection with the out-degree influence region of node u, then the directed edge (u, v) is added.

Disjunction rule. If node u is contained in the in-degree influence region of v, or node v is contained in the out-degree influence region of u, then the directed edge (u, v) is added.

Conjunction rule. If node u is contained in the in-degree influence region of v, and node v is contained in the out-degree influence region of u, then the directed edge (u, v) is added.

One of the rules is chosen (or some combination of them, depending on the motivating application), and edges are added according to the rules. Observe that the disjunction rule is the closest to the SPA model with $A_{\mathrm{out}} = B_{\mathrm{out}} = 0$. Note that edges may well be added between pairs of older nodes in a given time step, not just between the new node and the older nodes. The SPA model also has a fairly small bound B on the out-degree with high probability (see Theorem 1.5). This implies that the graphs so generated have tree width at most B with high probability, which does not accurately model the large tree width observed in the web graph (see [Aiello et al. 01]). The GSPA model may be converted into an undirected model. In this model, there is an influence region based on degree. An edge is added between two nodes according to an overlap rule. The overlap rules above are easily modified to the undirected case.

Acknowledgments. The authors gratefully acknowledge support from NSERC and MITACS grants.

References

[Aiello et al. 01] W. Aiello, F. R. K. Chung, and L. Lu. "A Random Graph Model for Massive Graphs." *Experimental Mathematics* 10:1 (2001), 53–66.

[Bonato 08] A. Bonato. *A Course on the Web Graph*. Providence, RI: American Mathematical Society, 2008.

[Chung and Lu 06] F. R. K. Chung and L. Lu. *Complex Graphs and Networks*. Providence, RI: American Mathematical Society, 2006.

[Cooper 06] C. Cooper. "The Age Specific Degree Distribution of Web-Graphs." *Combinatorics, Probability and Computing* 15 (2006), 637–661.

[Flaxman et al. 06] A. Flaxman, A. M. Frieze, and J. Vera. "A Geometric Preferential Attachment Model of Networks." *Internet Mathematics* 3:2 (2006), 187–205.

[Flaxman et al. 07] A. Flaxman, A. M. Frieze, and J. Vera. "A Geometric Preferential Attachment Model of Networks. II." In *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007, San Diego, CA, USA, December 11–12, 2007, Proceedings*, Lecture Notes in Computer Science 4863, pp. 41–55. Berlin: Springer, 2007.

[Janson et al. 00] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. New York: Wiley, 2000.

[Menczer 04] F. Menczer. "Lexical and Semantic Clustering by Web Links." *Journal of the American Society for Information Science and Technology* 55:14 (2004), 1261–1269.

[Penrose 03] M. Penrose. *Random Geometric Graphs*. Oxford, UK: Oxford University Press, 2003.

[Pittel et al. 96] B. Pittel, J. Spencer, and N. Wormald. "Sudden Emergence of a Giant k-Core in a Random Graph." *Journal of Combinatorial Theory, Series B* 67 (1996), 111–151.

[Wormald 99] N. Wormald. "The Differential Equation Method for Random Graph Processes and Greedy Algorithms." In *Lectures on Approximation and Randomized Algorithms*, edited by M. Karoński and H. J. Prömel, pp. 73–155. Warsaw: PWN, 1999.

[Young and Scheinerman 07] S. J. Young and E. R. Scheinerman. "Random Dot Product Graph Models for Social Networks." In *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007, San Diego, CA, USA, December 11–12, 2007, Proceedings*, Lecture Notes in Computer Science 4863, pp. 138–149. Berlin: Springer, 2007.

W. Aiello, Department of Computer Science, University of British Columbia, 201-2366 Main Mall, Vancouver, BC, V6T 1Z4, Canada (aiello@cs.ubc.ca)

A. Bonato, Department of Mathematics, Ryerson University, Toronto, ON, M5B 2K3, Canada (abonato@ryerson.ca)

C. Cooper, Department of Computer Science, King's College, London, WC2R 2LS, United Kingdom (colin.cooper@kcl.ac.uk)

J. Janssen, Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, B3H 3J5, Canada (janssen@mathstat.dal.ca)

P. Prałat, Department of Mathematics and Statistics, Dalhousie University, Halifax NS, B3H 3J5, Canada (pralat@mathstat.dal.ca)

Received February 14, 2008; accepted May 6, 2008.