Preferred Attachment in Affiliation Networks

Mindaugas Bloznelis · Friedrich Götze

Received: 13 February 2014 / Accepted: 26 May 2014 / Published online: 19 June 2014
© Springer Science+Business Media New York 2014

Abstract Vertices of an affiliation network are linked to attributes and two vertices are declared adjacent whenever they share a common attribute. For example, two customers of an internet shop (or video-sharing website) are called adjacent if they have purchased (or downloaded) the same or similar items. Assuming that each newly arrived customer is linked preferentially to already popular items we obtain a preferred attachment affiliation network that evolves in time. We show that the fraction of customers having \(i\) neighbours scales as \(i^{-2-\alpha}\ln i\) for large \(i\). Here \(\alpha > 0\) is the ratio between the two intensities: intensity of the flow of customers and that of the newly arriving items.

Keywords Preferential attachment · Affiliation network · Degree sequence · Power law · Random intersection graph

1 Introduction and Results

A preferential attachment model of evolving network assumes that each newly arrived vertex is attached preferentially to already well connected sites, [3]. The rich get richer principle is usually implemented by setting the probability of an edge between the new vertex \(v'\) and an old vertex \(v\) to be an increasing function of the degree of \(v\) (the number of neighbours of \(v\)), [6,7,10].

An important class of social networks are affiliation networks: members of a network tend to establish relations if they share some common interests or attributes [8,13]. For example, for marketing purposes customers of an internet shop (users of a video-sharing website, like, e.g., YouTube) are considered related to each other if they have purchased the same or similar items (downloaded the same movie). Here the rich get richer principle affects
customers and items simultaneously: a newly arrived customer $v'$ is linked preferentially to already highly popular items, thus, further increasing their popularity. Similarly, by selecting a highly popular item, the new customer $v'$ becomes adjacent to already highly connected customers (those linked to this item), thus further, increasing the number of their neighbours.

In the present study we consider a preferred attachment model of an affiliation network, where vertices (customers) are linked to items independently at random, and the probability of a link between the new vertex $v'$ and an item $w$ is proportional to the number of vertices already linked to $w$. Two vertices of the network are declared adjacent whenever there is an item linked to both of them. We show that the network has asymptotic degree distribution and give explicit formula for the limiting degree probabilities, see (1) below. In particular, it is shown that the fraction of vertices having degree $i$ scales as $i^{-2-\alpha} \ln i$ for large $i$, where $\alpha > 0$ is a model parameter, see (4) below. We remark, that the logarithmic factor does not show up in related preferred attachment models [6, 7, 10, 11] that do not exploit the bipartite structure of affiliation networks.

1.1 Model

Given $\lambda > 0$ and integer $k > 0$, let $l \geq 0$ be an integer such that $\lambda \leq k + l$. Consider an internet library which contains $w_1, \ldots, w_l$ items at the beginning. Every item $w_j$ is prescribed an initial score $s(w_j) = 1$. On the first step new items $w_{l+1}, \ldots, w_{l+k}$ arrive to the library, each having initial score 1. Then the first customer $v_1$ visits the library and downloads items independently at random: An item $w$ is selected with probability $p_{1,s(w)} = \lambda s(w) (l + k)^{-1}$. Every item selected by $v_1$ increases its score by one.

The collection of items of the library after $n$ steps is denoted $W_n = \{w_1, \ldots, w_{l+nk}\}$. On the $n+1$ step $k$ new items arrive to the library, each having initial score 1. Then the customer $v_{n+1}$ enters the library and downloads items of the library independently at random: An item $w$ is downloaded with probability $p_{n+1,s(w)} = \lambda s(w)(l + (n+1)k + n\lambda)^{-1}$ proportional to the score $s(w)$ of $w$. Here $s(w) - 1$ is the number of customers from $V_n = \{v_1, \ldots, v_n\}$ that have downloaded the item $w$. Every item chosen by $v_{n+1}$ increases its score by one. Finally, two customers are adjacent if there is at least one item selected (downloaded) by both of them. We are interested in the graph $G_n$ on the vertex set $V_n$ defined by this adjacency relation.

For convenience, we may represent items as bins. Each newly arrived bin contains a single ball. A new customer $v_{n+1}$ throws balls into bins $w_1, \ldots, w_{l+(n+1)k}$ at random: Each bin $w$ receives a ball with probability $p_{n+1,s(w)}$ and independently of the other bins. The score $s(w)$ counts the (current) number of balls in the bin $w$. This number may increase with $n$. It measures the popularity of the bin (item) $w$. Hence, popular bins (items) have higher chances to be selected.

1.2 Results

In the present note we address the question about the degree sequence of $G_n$. We shall show that for every $i = 0, 1, \ldots$, the fraction of vertices (customers) $v \in V_n$ of $G_n$ having degree $d(v) = i$ converges in probability to a limit and identify this limit. Namely, we have as $n \to +\infty$
\[
\frac{\#\{v \in V_n : d(v) = i\}}{n} \xrightarrow{p} (1 + \alpha) \mathbb{E}_{\{Z \leq i, \Lambda \geq 1\}} \frac{\Gamma(i + 2\Lambda)}{\Gamma(i + 2\Lambda + \alpha + 2)} \times \frac{\Gamma(Z + 2\Lambda + \alpha + 1)}{\Gamma(Z + 2\Lambda)}, \quad i \geq 1,
\]
\[
\frac{\#\{v \in V_n : d(v) = 0\}}{n} \xrightarrow{p} \mathbb{E}_{\{Z = 0\}} \frac{1 + \alpha}{2\Lambda + 1 + \alpha}.
\]

By \( \xrightarrow{p} \) we denote the convergence in probability. Here \( \alpha = k/\lambda \). \( \Gamma \) denotes Euler’s Gamma function. \( \Lambda \) denotes a Poisson random variable with mean \( \lambda \). \( Z \) is a compound Poisson random variable
\[
Z = \sum_{i=1}^{\Lambda} T_i,
\]
where \( T_1, T_2, \ldots \) are independent random variables independent of \( \Lambda \) and having the same probability distribution
\[
\mathbb{P}(T_1 = j) = x_{j+1}, \quad x_{j+1} = (1 + \alpha)\Gamma(2 + \alpha)\frac{\Gamma(j + 1)}{\Gamma(3 + \alpha + j)}, \quad j = 0, 1, 2, \ldots.
\]
From (1) we find the tail behaviour of the limiting degree distribution. Let \( y_i \) denote the quantity on the right hand side of (1). We have as \( i \to +\infty \)
\[
y_i \sim \lambda(1 + \alpha)^2\Gamma(2 + \alpha)i^{-2-\alpha} \ln i.
\]
Here and below we write \( z_i \sim q_i \) whenever \( z_i/q_i \to 1 \) as \( i \to +\infty \).

Numbers \( x_i, i = 1, 2, \ldots \), have interesting interpretation. They are limits (in probability) of the fractions of the number of items having score \( i \):
\[
\frac{\#\{w \in W_n : s(w) = i\}}{nk} \xrightarrow{p} x_i, \quad \text{as} \quad n \to +\infty.
\]
From the properties of Gamma function (formula (6.1.46) of [2]) we conclude that the sequence \( \{x_i\}_{i \geq 1} \) obeys a power law with exponent \( 2 + \alpha \),
\[
x_i \sim (1 + \alpha)\Gamma(2 + \alpha)i^{-2-\alpha}, \quad \text{as} \quad i \to +\infty.
\]

1.3 Related Work

Results of an empirical study of an evolving coauthorship network (an affiliation network, where authors are declared adjacent if they have a joint publication) are reported in [12]. The model considered in the present paper seems to be new. The idea of such a model has been suggested by Colin Cooper. The extra logarithmic factor in (4) indicates that the degree distribution of the preferred attachment affiliation model has a slightly heavier tail in comparison to that of the “usual” preferential attachment model, see [6,7,10,11], and that of the related spatial preferred attachment model [1]. Let us briefly compare the spatial and affiliation preferred attachment models. In the spatial model each vertex is placed in space (multi-dimensional thorus) and surrounded by an influence region. With probability \( p \) a new vertex \( u \) sends a directed edge to an existing vertex \( v \) whenever \( u \) falls within the influence region of \( v \). Furthermore, the volume of each region is an affine function of the in-degree of the vertex and it is scaled by time. Note that both models represent vertices by subsets (influence regions and collections of items selected by users, respectively) and edges are witnessed by intersections. One significant difference between them is that in the affiliation model the sets are finite and their elements have individual scores that vary very much. It
seems that this variability contributes also the “extra“ log \( i \) factor when compared to the pure power law \( i^{-2-\alpha} \) observed in the spatial preferred attachment model. It is worth mentioning here that affiliation network models, where scores (6) are prescribed to items independently of the choices of vertices, have even heavier tails: The proportion of vertices of degree \( i \) scales as \( i^{-1-\alpha} \) as \( i \to +\infty \), see [4,5].

An important property of real affiliation networks is that they admit a non-vanishing clustering coefficient, [13]. Clustering characteristics of the preferred attachment affiliation model will be considered elsewhere. Another interesting lines of further research were an improvement of (1) by providing concentration bounds, c.f., [1], and an extension of the model by prescribing individual fitness factors to items and vertices.

The paper is organized as follows. A heuristic argument explaining (1) and (5) is given in Sect. 2. A rigorous proof of (1), (4) and (5) is given in Sect. 3. Technical details are collected in Appendix.

Before proceeding further we introduce some notation.

1.4 Notation

We call \( w \in W_n \) and \( v \in V_n \) related whenever \( w \) contains a ball produced by \( v \). The number of balls produced by \( v \) is called the activity of \( v \). A vertex \( v \in V_n \) is called regular in \( G_n \) if every vertex adjacent to \( v \) in \( G_n \) shares with \( v \) a single bin. Let \( d_{i,r}^{(n)} \) denote the probability of the event \( V_{i,r} = \{v_{n+1} \text{ has activity } r, \text{ it has degree } i \text{ in } G_{n+1}, \text{ and it is a regular vertex of } G_{n+1}\} \). We call \( v \in V_n \) an \([i, r]\) vertex if its activity is \( r \) and its degree in \( G_n \) is \( d(v) = i \).

By \( s_{i,r}(v) \) we denote the (current) number of balls contained in the bins related to an \([i, r]\) vertex \( v \) of \( G_n \). We note that any regular \([i, r]\) vertex \( v \) of \( G_n \) has \( s_{i,r}(v) = i + 2r =: s_{i,r} \).

Let \( Y_i^{(n)} \) denote the number of regular vertices of \( G_n \) of degree \( d(v) = i \), and let \( Y_{i,r}^{(n)} \) denote the number of regular \([i, r]\) vertices in of \( G_n \). Given \( w \in W_n \), we denote by \( s_n(w) \) the score of \( w \) after the \( n \)th step. By \( X_i^{(n)} \) we denote the number of bins \( w \in W_n \) of score \( s_n(w) = i \). We put \( X_1^{(0)} = 1 \) and \( X_i^{(0)} = 0, \text{ for } i \geq 2 \).

2 Heuristic

We start with explaining formula (5). Assume for a moment that for each \( i \) the ratios \( X_i^{(n)}/(nk) \) converge to some limit, say \( \bar{x}_i \), as \( n \to +\infty \). So that for large \( n \) we have \( X_i^{(n)} \approx \bar{x}_i n k \). Then from the relations describing approximate behaviour of the numbers \( X_i^{(n)} \),

\[
X_1^{(n+1)} \approx (X_1^{(n)} + k)(1 - p_{n+1,1}), \\
X_2^{(n+1)} \approx X_2^{(n)} (1 - p_{n+1,2}) + (X_1^{(n)} + k)p_{n+1,1}, \\
X_i^{(n+1)} \approx X_i^{(n)} (1 - p_{n+1,i}) + X_i^{(n)} p_{n+1,i-1}, \quad i = 3, 4, \ldots,
\]

we obtain, by neglecting \( O(n^{-1}) \) terms, the equations

\[
\bar{x}_1(n+1)k = (\bar{x}_1nk + k) \left(1 - \frac{1}{n+1} \frac{1}{1+\alpha}\right), \\
\bar{x}_i(n+1)k = \bar{x}_i nk \left(1 - \frac{1}{n+1} \frac{i}{1+\alpha}\right) + \bar{x}_{i-1}k \frac{i-1}{1+\alpha}, \quad i \geq 2.
\]
Solving these equations we arrive to the sequence \( \{x_i\}_{i \geq 1} \) given by formula (3). We remark that \( \{x_i\}_{i \geq 1} \) is a sequence of probabilities having a finite first moment. More precisely, we have
\[
\sum_{i \geq 1} x_i = 1, \quad \sum_{i \geq 1} i x_i = 1 + \alpha^{-1}. \tag{7}
\]

In particular, the common probability distribution of random variables \( T_i \) is well defined. We note that identities (7) are simple consequences of the well known properties of the Gamma function and hyper-geometric series (formulas (6.1.46), (15.1.20) of [2]).

Next we explain (1). We observe that, given \( X_1^{(n)}, X_2^{(n)}, \ldots \), the conditional probability of the event \( V_{i,r} \) is
\[
q_r^{(n)} = \sum_{u_1 + \cdots + u_i+1 = r, \atop 1u_2 + 2u_3 + \cdots + iu_{i+1} = i} \frac{\left( X_1^{(n)}+k \right)_{u_1} \left( X_2^{(n)} \right)_{u_2} \cdots \left( X_{i+1}^{(n)} \right)_{u_{i+1}}}{\left( X_1^{(n)}+X_2^{(n)}+\cdots \right)_{r}} \frac{r!}{u_1! \cdots u_{i+1}!} + o(1). \tag{8}
\]

Here we use notation \( (x)_u = x(x-1)\ldots(x-u+1) \), \( u_s \) counts those bins \( w \in W_{n+1} \) of score \( s_n(w) = s \) that have received a ball from \( v_{n+1} \), and \( q_r^{(n)} \) is the conditional probability, given \( X_1^{(n)}, X_2^{(n)}, \ldots \), of the event that \( v_{n+1} \) has produced \( r \) balls. The remainder \( o(1) \) accounts for the probability that \( v_{n+1} \) is not a regular vertex of \( G_{n+1} \). Now, using the approximations \( X_i^{(n)} \approx x_i/nk \), \( i \geq 1 \), and identities (7) we, firstly, approximate the first fraction of (8) by \( x_1^{u_1} \cdots x_{i+1}^{u_{i+1}} \) and, secondly, we approximate \( q_r^{(n)} \) by the Poisson probability \( e^{-\lambda} \lambda^r / r! \). We obtain that
\[
q_r^{(n)} \approx e^{-\bar{\lambda}} \frac{\lambda^r}{r!} \sum_{u_1 + \cdots + u_i+1 = r, \atop 1u_2 + 2u_3 + \cdots + iu_{i+1} = i} x_1^{u_1} x_2^{u_2} \cdots x_{i+1}^{u_{i+1}} \frac{r!}{u_1! \cdots u_{i+1}!} =: c_{i,r}. \tag{9}
\]

Furthermore, we observe that the probability that \( v_{n+1} \) sends a ball to a bin related to a regular \([i, r] \) vertex \( v \) of \( G_n \) is \( s_{i,r} p_{n+1,1} + O(n^{-2}) \).

Now, assume for a moment that for each \( i, r \) the ratios \( Y^{(n)}_i/n \) converge to some limit, say \( \tilde{y}_i \), and \( Y^{(n)}_{i,r}/n \) converge to some limit, say \( \tilde{y}_{i,r} \), as \( n \to +\infty \). So that for large \( n \) we have \( Y^{(n)}_i \approx \tilde{y}_i n \) and \( Y^{(n)}_{i,r} \approx \tilde{y}_{i,r} n \). Invoking these approximations in the relations describing approximate behaviour of numbers \( Y^{(n)}_{i,r} \),
\[
\begin{align*}
Y_{0,0}^{(n+1)} & \approx Y_{0,0}^{(n)} + q_{0,0}^{(n)}, \\
Y_{0,r}^{(n+1)} & \approx Y_{0,r}^{(n)} (1 - s_{0,r} p_{n+1,1}) + q_{0,r}^{(n)}, \quad r \geq 1, \\
Y_{i,r}^{(n+1)} & \approx Y_{i,r}^{(n)} (1 - s_{i,r} p_{n+1,1}) + Y_{i-1,r}^{(n)} s_{i-1,r} p_{n+1,1} + q_{i,r}^{(n)}, \quad i, r \geq 1,
\end{align*}
\]
we obtain, by neglecting $O(n^{-1})$ terms and using the approximation $q_{i,r}^{(n)} \approx c_{i,r}$, the equations

\begin{align*}
\bar{y}_{0,0} &= c_{0,0}, \\
\bar{y}_{0,r} &= \frac{1 + \alpha}{1 + \alpha + 2r} c_{0,r}, \quad r \geq 1, \\
\bar{y}_{i,r} &= \frac{2r + i - 1}{1 + \alpha + 2r + i} \bar{y}_{i-1,r} + \frac{1 + \alpha}{1 + \alpha + 2r + i} c_{i,r}, \quad i, r \geq 1.
\end{align*}

(10) (11)

Solving these equations we arrive to the sequence \{y_{0,0}, y_{i,r}, i \geq 0, r \geq 1\} given by the formulas

\begin{align*}
y_{0,0} &= c_{0,0}, \\
y_{i,r} &= (1 + \alpha) \sum_{j=0}^{i} \frac{(2r + i - 1)_{i-j}}{(1 + \alpha + 2r + i)_{i-j+1}} c_{j,r}.
\end{align*}

(12) (13)

Next we use the identity $c_{j,r} = P(Z = j, \Lambda = r) = E[I_{\Lambda=r}]I_{Z=j}$ and write (13) in the form

$$y_{i,r} = (1 + \alpha) E[I_{\Lambda=r}]I_{Z\leq i} \frac{(2\Lambda + i - 1)_{i-Z}}{(1 + \alpha + 2\Lambda + i)_{i-Z+1}}.$$

Hence we obtain, for $i \geq 1$,

$$\bar{y}_{i} = \sum_{r \geq 1} \bar{y}_{i,r} \approx \sum_{r \geq 1} y_{i,r}$$

$$= (1 + \alpha) E[I_{\Lambda \geq 1}]I_{Z \leq i} \frac{(i + 2\Lambda - 1)_{i-Z}}{(i + 2\Lambda + \alpha + 1)_{i-Z+1}}$$

$$= (1 + \alpha) E[I_{\Lambda \geq 1}]I_{Z \leq i} \frac{\Gamma(i + 2\Lambda)}{\Gamma(i + 2\Lambda + \alpha + 2)} \frac{\Gamma(Z + 2\Lambda + \alpha + 1)}{\Gamma(Z + 2\Lambda)}$$

$$= y_{i}$$

and

$$\bar{y}_{0} = \sum_{r \geq 0} \bar{y}_{0,r} \approx \sum_{r \geq 0} y_{0,r}$$

$$= P(\Lambda = 0) + E[I_{\Lambda \geq 1}]I_{Z=0} \frac{1 + \alpha}{2\Lambda + \alpha + 1} = E[I_{Z=0}] \frac{1 + \alpha}{2\Lambda + \alpha + 1}$$

$$= y_{0}.$$ 

We conclude that the fraction of regular vertices of degree $i = 0, 1, \ldots$ satisfies (1). Now, the result follows from the fact that the fraction of non regular vertices of degree $i$ is $o(1)$ as $n \to +\infty$.

3 Rigorous Results

Let $\bar{Y}_{i,r}^{(n)}$ denote the number of non regular \{i, r\} vertices of $G_n$. We recall that the numbers $c_{i,r}$ and $y_{i,r}$ are defined in (9) and (13) above.

**Theorem 1** Let $n \to +\infty$. Then relations (1), (4), (5) hold.
Proof of Theorem 1 Let us prove \((1)\). Let \(S_n\) denote the total number of balls in the network after the \(n\)-th step. A simple induction argument shows that \(E S_n = l + nk + n \lambda\), see \((53)\).

Let \(\bar{Y}_r^{(n)}\) denote the number of vertices \(v \in V_n\) with activity at least \(r\). We observe that for any \(0 < \varepsilon < 1\)
\[
\sup_n P(n^{-1} \bar{Y}_r^{(n)} > \varepsilon) \to 0 \tag{14}
\]
as \(r \to \infty\). Indeed, vertices of \(V_n\) with activity at least \(r\) contribute at least \(r \bar{Y}_r^{(n)}\) balls to \(S_n\). Hence, \(\bar{Y}_r^{(n)} \leq r^{-1} S_n\) and we obtain \((14)\), by Markov’s inequality. Now \((1)\) follows from \((14)\) and the fact that \(n^{-1} \bar{Y}_{i,r}^{(n)} \to 0\) and \(n^{-1} Y_{i,r}^{(n)} \to y_{i,r}\) in probability as \(n \to +\infty\) for \((i, r) = (0, 0)\) and \(i \geq 0, r \geq 1\). This fact follows from Lemma 2: we have \(n^{-1} E \bar{Y}_r^{(n)} \to 0\), \(n^{-1} E Y_{i,r}^{(n)} \to y_{i,r}\) and \(\text{Var}(n^{-1} Y_{i,r}^{(n)}) \to 0\).

Relation \((5)\) follows from \((25)\): we have \((nk)^{-1} E X_i^{(n)} \to x_i\) and \(\text{Var}((nk)^{-1} X_i^{(n)}) \to 0\).

Let us prove \((4)\). In the proof we use the following facts:

(i) for any \(a, x > 0\) such that \(x > 10^3\) and \(a/x < 10^{-3}\) we have
\[
\ln(\Gamma(a)) - \ln(\Gamma(x + a)) = -a \ln x + Rx^{-1}, \quad \text{where } |R| \leq c(a^2 + 1); \tag{15}
\]
(ii) for any \(t \geq 1\) we have
\[
P(\Lambda > t) = P(e^{\lambda \sqrt{\ln \Lambda}} > e^{\sqrt{\ln t}}) \leq e^{-t \sqrt{\ln t}} \leq c_2 e^{-t \sqrt{\ln t}};
\]
(iii) for \(t \to +\infty\) we have
\[
P(Z > t) \sim P(T_1 > t) e^\Lambda \sim \lambda \Gamma(2 + \alpha) t^{-1-\alpha}.
\]

We remark that \((i)\) can be obtained from Binet’s first formula; \((ii)\) follows by Markov’s inequality; the first relation of \((iii)\) follows from a general result about the tail probability asymptotics of a randomly stopped sum of iid sub-exponential random variables, see, e.g., [9].

Now we show \((4)\). For \(i \geq 1\) we write the expectation in \((1)\) in the form
\[
E X Y \mathbb{I}_{Z \leq i, \Lambda \geq 1} = I_1 + I_2 + I_3, \tag{16}
\]
where \(X = X(i, \Lambda), Y = Y(Z, \Lambda), \)
\[
X(i, L) = \frac{\Gamma(i + 2L)}{\Gamma(i + 2L + \alpha + 2)}, \quad Y(z, L) = \frac{\Gamma(z + 2L + \alpha + 1)}{\Gamma(z + 2L)}
\]
and
\[
I_1 = E X Y \mathbb{I}_{\{ln^2 i < Z \leq i, 1 \leq \Lambda \leq \ln i\}}, \quad I_2 = E X Y \mathbb{I}_{\{Z \leq \ln^2 i, 1 \leq \Lambda \leq \ln i\}}, \quad I_3 = E X Y \mathbb{I}_{\{Z \leq i, \Lambda > \ln i\}}.
\]

We shall show for \(i \to +\infty\) that \(I_2 \sim O(i^{-2-\alpha} \ln i), I_3 \sim O(i^{-1-\alpha})\) and
\[
I_1 \sim \lambda (1 + \alpha) \Gamma(2 + \alpha) i^{-2-\alpha} \ln i. \tag{17}
\]

Firstly, combining the inequality \(X Y \leq 1\) with \((ii)\) we obtain \(I_3 \leq P(\Lambda > \ln i) = O(i^{-1-\alpha})\). Next, we evaluate \(I_1\), \(I_2\) using the approximations
\[
X(i, L) = i^{-2-\alpha} (1 + O(i^{-1} \ln i)), \quad Y(z, L) = z^{1+\alpha} (1 + O(\ln^{-1} i)), \tag{18}
\]
which hold uniformly in \(L \in [1, \ln i]\) and \(z \geq \ln^2 i\). We note that \((18)\) follow from \((i)\).
We estimate $I_2$ using the inequality $Y \leq c_\alpha (1 + (Z + 2\Lambda)^{1 + \alpha})$. For large values of $Z + 2\Lambda$, say, $Z + 2\Lambda > (\alpha + 1)10^3$, this inequality follows from (i). For $Z + 2\Lambda \leq (\alpha + 1)10^3$ the inequality is obvious. Hence, we obtain, using the first relation of (18),

$$I_2 \leq i^{-2-\alpha}(1 + O(i^{-1}\ln i))EY_{|[Z \leq \ln^2i]} \leq c'_\alpha i^{-2-\alpha}\left(1 + E\Lambda^{1+\alpha}_{|[Z \leq \ln^2i]} + E|\Lambda|^{1+\alpha}\right).$$

Now, invoking the bound

$$E\Lambda^{1+\alpha}_{|[Z \leq \ln^2i]} \leq \int_0^{(\ln^2i)^{1+\alpha}} P(Z^{1+\alpha} \geq t)dt + O(1) = O(\ln \ln i), \tag{19}$$

which follows from (iii), and the bound $E|\Lambda|^{1+\alpha} = O(1)$, we obtain $I_2 = O(i^{-2-\alpha}\ln \ln i)$.

It remains to prove (17). We write using (18)

$$I_1 = i^{-2-\alpha}(1 + O(i^{-1}\ln i))E\Lambda^{1+\alpha}_{|[\ln^2i < Z \leq i]}E\Lambda^{1+\alpha}_{|[1 \leq \Lambda \leq \ln i]} = O(i^{-9}). \tag{20}$$

Next we skip the indicator $\Lambda_{|[1 \leq \Lambda \leq \ln i]}$. Indeed, since $Z > \ln^2i$ implies $\Lambda \geq 1$, and (ii) implies $P(\Lambda > \ln i) = O(i^{-10-\alpha})$ we have

$$0 \leq E\Lambda^{1+\alpha}_{|[\ln^2i < Z \leq i]} - E\Lambda^{1+\alpha}_{|[1 \leq \Lambda \leq \ln i]} \leq i^{1+\alpha}P(\Lambda > \ln i) = O(i^{-9}). \tag{21}$$

Furthermore, (19) implies

$$E\Lambda^{1+\alpha}_{|[\ln^2i < Z \leq i]} = E\Lambda^{1+\alpha}_{|[Z \leq i]} - O(\ln i) \tag{22}$$

and (iii) implies

$$E\Lambda^{1+\alpha}_{|[Z \leq i]} = \int_0^{i^{1+\alpha}} (P(Z^{1+\alpha} \geq t) - P(Z \geq i)) dt + O(1) \sim \lambda (1 + \alpha)\Gamma(2 + \alpha) \ln i. \tag{23}$$

Finally, (17) follows from (20)–(23). \hfill \Box

The remaining part of the section contains auxiliary lemmas.

We write for short $p_{n+1,s} = p_s = s\kappa_n$, where

$$\kappa_n = p_{n+1,1} = \frac{1}{n}\frac{1}{1 + \alpha}\left(1 - \frac{1}{n}\frac{\alpha + \beta}{1 + \alpha + n^{-1}\alpha + n^{-1}\beta}\right), \quad \beta := \frac{l}{\lambda}. \tag{24}$$

Denote

$$x^{(n)}_i = (nk)^{-1}E X^{(n)}_i, \quad y^{(n)}_{i,r} = n^{-1}E Y^{(n)}_{i,r}, \quad \tilde{y}^{(n)}_{i,r} = n^{-1}E \tilde{Y}^{(n)}_{i,r},$$

$$h^{(n)}_{i,j} = (nk)^{-1}\left(E X^{(n)}_i X^{(n)}_j - E X^{(n)}_i E X^{(n)}_j\right), \quad g^{(n)}_{i,j;r} = n^{-2}\left(E Y^{(n)}_{i,r} Y^{(n)}_{j,r} - E Y^{(n)}_{i,r} E Y^{(n)}_{j,r}\right).$$

**Lemma 1** For any $i, j \geq 1$ we have as $n \to +\infty$

$$x^{(n)}_i = x_i + O(n^{-1}), \quad (nk)^{-2}E X^{(n)}_i X^{(n)}_j = x_i x_j + O(n^{-1}). \tag{25}$$

Moreover, the finite limits

$$h_{i,j} = \lim_{n} h^{(n)}_{i,j}, \quad i, j \geq 1, \tag{26}$$
exist and can be calculated using the recursive relations

\[ h_{i,i} = \frac{2(i-1)h_{i,i-1} + ix_i + (i-1)x_{i-1}}{i + i + 1 + \alpha}, \]

\[ h_{i,i+1} = \frac{(i-1)h_{i-1,i+1} + ih_{i,i} - i x_i}{i + (i + 1) + 1 + \alpha}, \]

\[ h_{i,r} = \frac{(i-1)h_{i-1,r} + (r-1)h_{i,r-1}}{i + r + 1 + \alpha}, \quad r \geq i + 2. \]  

In particular, we have for every \( i, j \geq 1, \)

\[(nk)^{-2}E x^{(n)}_i x^{(n)}_j = x^{(n)}_i x^{(n)}_j + h_{i,j}(nk)^{-2} + o(n^{-1}).\]  

Here we use notation \( x_0 = 0 \) and \( h_{i,j} = 0, \) for \( \min\{i,j\} = 0. \)

**Proof of Lemma 1** Let us prove the first relation of (25). The identities

\[ E X^{(n+1)}_1 = (1 - p_1)E(X^{(n)}_1 + k), \]

\[ E X^{(n+1)}_2 = (1 - p_2)E X^{(n)}_2 + p_1 E(X^{(n)}_1 + k), \]

\[ E X^{(n+1)}_i = (1 - p_i)E x^{(n)}_i + p_{i-1}E x^{(n)}_{i-1}, \quad i \geq 1, \]

imply

\[ x^{(n+1)}_1 = x^{(n)}_1(1 - n^{-1} - p_1) + n^{-1} + O(n^{-2}), \quad i = 1 \]

\[ x^{(n+1)}_i = x^{(n)}_i(1 - n^{-1} - p_i) + x^{(n)}_{i-1} p_{i-1} + O(n^{-2}), \quad i \geq 1. \]

Relation (31) combined with Lemma 3 implies \( x^{(n)}_1 = x_1 + O(n^{-1}). \) For \( i \geq 2 \) we proceed recursively: using the fact that \( x^{(n)}_{i-1} = x_{i-1} + O(n^{-1}) \) we conclude from (32) by Lemma 3 that \( x^{(n)}_i = x_i + O(n^{-1}). \)

Next, we observe that the second relation of (25) follows from (26), (30). Furthermore, (30) follows from (27)–(29). Hence we only need to prove (27)–(29).

For convenience we write \( h^{(n)}_{i,j} = 0, \) for \( \min\{i,j\} = 0. \) We also put \( x^{(n)}_0 = 0. \) Clearly, \( h^{(n)}_{i,j} = h^{(n)}_{i,j} \) for \( i, j \geq 0. \)

Let us prove (27). A straightforward calculation shows that

\[ h^{(n+1)}_{i,i} \frac{n + 1}{n} = h^{(n)}_{i,i}(1 - p_i)^2 + h^{(n)}_{i,i-1} 2(1 - p_i)p_{i-1} + h^{(n)}_{i-1,i-1} p_{i-1}^2 \]

\[ + x^{(n)}_{i-1}(p_{i-1} - p_{i-1}^2) + x^{(n)}_i(p_i - p_{i-1}) \frac{n}{n + 1} + O(n^{-2}), \]

\[ h^{(n+1)}_{i,i+1} \frac{n + 1}{n} = h^{(n)}_{i,i+1}(1 - p_i)(1 - p_{i+1}) + h^{(n)}_{i-1,i+1} p_{i-1}(1 - p_{i+1}) + h^{(n)}_{i-1,i} p_{i-1}p_i \]

\[ + h^{(n)}_i(p_i - p_{i-1}) - x^{(n)}_i p_i(1 - p_i) + O(n^{-2}), \]

and, for \( r \geq 2 + i, \)

\[ h^{(n+1)}_{i,r} \frac{n + 1}{n} = h^{(n)}_{i,r}(1 - p_r)^2 + h^{(n)}_{i-1,r} p_{i-1}(1 - p_r) + h^{(n)}_{i-1,r-1} p_{i-1}p_{r-1} \]

\[ + h^{(n)}_{i,r-1}(1 - p_r) + O(n^{-2}). \]

We note that (33) and Lemma 3 imply that the sequence \( \{h^{(n)}_{1,1}\}_{n\geq 1} \) converges to \( h_{1,1} \) defined by (27). Furthermore, using the fact that (26) holds for \( i = j = 1 \) we obtain from (34) and
Lemma 3 that \( h_{1,2}^{(n)} \) converges to \( h_{1,2} \) defined by (28). Next, for \( i = 1 \) and \( r = 3, 4, \ldots \), we proceed recursively: using (35) and Lemma 3 we establish (26), with \( h_{ir} \) given by (29). In this way we prove the lemma for \( i = 1 \) and \( r \geq i \).

The case \( i = 2, r \geq i \) is treated similarly. For \( i = r = 2 \) we apply (33) and Lemma 3. For \( i = 2 \) and \( r = 3 \) we apply (34) and Lemma 3. Finally, for \( i = 2 \) and \( r \geq i + 2 \) we apply (29) and Lemma 3.

Next we proceed recursively and prove the lemma for \( \{ (i, r), r = i, r = i + 1, r = i + 2, \ldots \}, i = 3, 4, \ldots \). \( \square \)

**Lemma 2** Let \( i, j = 0, 1, \ldots \) and \( r = 1, 2 \ldots \) We have as \( n \to +\infty \)

\[
y_i^{(n)} \to y_{i,r}, \quad g_i^{(n)} \to 0, \quad \tilde{y}_i^{(n)} \to 0.
\]

**Equation (36) remains valid for** \( i = j = r = 0. \)

**Proof of Lemma 2** Let \( i, j, r \geq 1. \) We show in Lemma 4 below that

\[
y_{i,0}^{(n+1)} = \tilde{y}_{i,1}^{(n+1)} = 0, \tag{37}
\]

\[
y_{i+1,r}^{(n+1)} \leq y_{i+1,r}^{(n)} (1 - n^{-1}) + y_{i,r}^{(n)} s_{ir} x_n + o(n^{-1}), \tag{38}
\]

\[
y_{0,0}^{(n+1)} = (1 - n^{-1}) y_{0,0}^{(n)} + n^{-1} c_{0,0} \tag{39}
\]

\[
y_{0,r}^{(n+1)} = (1 - n^{-1} - s_{0,r} x_n) y_{0,r}^{(n)} + n^{-1} c_{0,r} + o(n^{-1}), \tag{40}
\]

\[
y_{i,r}^{(n+1)} = (1 - n^{-1} - s_{i,r} x_n) y_{i,r}^{(n)} + s_{i-1,r} x_n y_{i-1,r}^{(n)} + n^{-1} c_{i,r} + o(n^{-1}), \tag{41}
\]

and

\[
g_{0,0}^{(n+1)} = (1 - 2n^{-1}) g_{0,0}^{(n)} + o(n^{-1}), \tag{42}
\]

\[
g_{0,r}^{(n+1)} = (1 - 2n^{-1} - 2 s_{0,r} x_n) g_{0,r}^{(n)} + o(n^{-1}), \tag{43}
\]

\[
g_{i,j,r}^{(n+1)} = (1 - 2n^{-1} - s_{i,j} x_n) g_{i,j,r}^{(n)} + s_{i-1,r} x_n g_{i-1,j,r}^{(n)} + s_{j-1,r} x_n g_{i,j-1,r}^{(n)} + o(n^{-1}), \tag{44}
\]

\[
g_{i,j,r}^{(n+1)} = (1 - 2n^{-1} - (s_{i,j} + s_{j,r}) x_n) g_{i,j,r}^{(n)} + s_{i-1,r} x_n g_{i-1,j,r}^{(n)} + s_{j-1,r} x_n g_{i,j-1,r}^{(n)} + o(n^{-1}). \tag{45}
\]

We recall that \( c_{i,r} \) is defined in (9). Here we prove that (37)–(45) imply (36).

Let us prove the third relation of (36). For \( i = 0 \), and for \( r = 0, 1 \) the relation follows from (37). Next, for any fixed \( r \geq 2 \) we proceed recursively: from (38) combined with the fact that \( y_{i,r}^{(n)} \to 0 \) we conclude by Lemma 3 that \( \tilde{y}_{i+1,r}^{(n)} \to 0 \).

Let us prove the first and second relation of (36) for various \( i, j \) and \( r. \) Firstly, combining (39) (respectively (42)) with Lemma 3 we obtain the first (respectively second) relation of (36), for \( i = j = r = 0. \) Secondly, combining (40) (respectively (43)) with Lemma 3 we obtain the first (respectively second) relation of (36), for \( i = j = 0, r \geq 1. \)

Now we prove the first relation of (36) for (the remaining values of indices) \( i \geq 1 \) and \( r \geq 1. \) We fix \( r \) and proceed recursively: from the fact that \( y_{i,r}^{(n)} \to y_{i-1,r} \) and relation (41) we conclude by Lemma 3 that \( y_{i,r}^{(n)} \to y_{i,r}. \)

Next, we prove the second relation of (36) for (the remaining values of indices) \( r \geq 1 \) and \( i + j \geq 1. \) Note that \( g_{i,j,r}^{(n)} = g_{j,i,r}^{(n)}. \) We fix \( r \) and proceed recursively in \( i \) and \( j. \)
For $i = 0$ and $j \geq 1$ we proceed as follows: from the fact that $s_{0,j-1;r}^{(n)} \to 0$ and relation (44) we conclude by Lemma 3 that $s_{1,j;r}^{(n)} \to 0$. In this way we prove the second relation of (36) for $(i, j)$ such that $i = 0$ and $j \geq 1$.

Now, consider indices $i = 1$ and $j \geq 1$. From the fact that $g_{1,j-1;r}^{(n)} \to 0$ and relation (45) we conclude by Lemma 3 that $s_{1,j;r}^{(n)} \to 0$. In this way we prove the second relation of (36) for $(i, j)$ such that $i = 1$ and $j \geq 1$.

Proceeding similarly we establish the second relation of (36) for $\{(i, i), (i, i + 1), (i, i + 2), \ldots\}, i = 2, 3, \ldots$.

\[ \square \]

**Lemma 3** Let $b, h \in \mathbb{R}$. Let $\{b_n\}_{n \geq 1}$ be a real sequence converging to $b$. Let $\{h_n\}_{n \geq 1}$ be a real sequence converging to $h$. Let $\{a_n\}_{n \geq 1}$ be a real sequence satisfying the recurrence relation

\[ a_{n+1} = a_n(1 - n^{-1}b) + n^{-1}h_n, \quad n \geq 1. \]

Assume that the series $\sum_{n \geq 1} n^{-1} |b_n - b|$ converges.

(i) For $b > 0$ we have $a_n \to hb^{-1}$. Suppose, in addition, that $b_n - b = O(n^{-1})$, $h_n - h = O(n^{-1})$. Then for $b \neq 1$ we have $a_n - hb^{-1} = O(n^{-1}b)$, and for $b = 1$ we have $a_n - hb^{-1} = O(n^{-1} \ln n)$.

(ii) For $b = 0$ we have $a_n \ln^{-1} n \to h$.

(iii) For $b < 0$ we have $a_n n^{-b} \to c_0$, for some positive constant $c_0$.

Let $\tilde{b} > 0$. Let $\{\tilde{a}_n\}_{n \geq 1}$, $\{\tilde{b}_n\}_{n \geq 1}$, $\{\tilde{h}_n\}_{n \geq 1}$ be non negative sequences such that $\tilde{b}_n \to \tilde{b}$, $\tilde{h}_n \to 0$, the series $\sum_{n \geq 1} n^{-1} |\tilde{b}_n - b|$ converges, and $\{\tilde{a}_n\}_{n \geq 1}$ satisfies the inequality

\[ \tilde{a}_{n+1} \leq \tilde{a}_n(1 - n^{-1}\tilde{b}_n) + n^{-1}\tilde{h}_n, \quad n \geq 1. \]

Then $\{\tilde{a}_n\}_{n \geq 1}$ converges to 0.

**Proof of Lemma 3** is given in Appendix.

**Acknowledgments** We thank the referee for valuable remarks. M. Bloznelis thanks K. Rybarczyk for discussion. The work of M. Bloznelis was supported by the Research Council of Lithuania Grant MIP-067/2013 and by the SFB 701 Grant at Bielefeld university.

**Appendix**

Here we prove Lemma 3. Then we evaluate the first and second order moments of $X_i^{(n)}$, $Y_i^{(n)}$, and $\tilde{Y}_i^{(n)}$.

**Proof of Lemma 3** We denote $\varepsilon_n = 1 - n^{-1}b_n = 1 - n^{-1}b + n^{-1}\delta_n$ and write, by iteration,

\[ a_{n+1} = a_n \varepsilon_n + \frac{h_n}{n} = a_1 \prod_{j=1}^{n} \varepsilon_j + \sum_{i=1}^{n} \frac{h_i}{i} \prod_{j=i+1}^{n} \varepsilon_j. \]

Let us consider sufficiently large $j_0$, so that for $j > j_0$ we have that $b/j$ and $|\delta_j|$ are less than 0.001. Then we can use the expansion

\[ \ln \varepsilon_j = \ln(1 - j^{-1}b + j^{-1}\delta_j) = -j^{-1}b + j^{-1}\delta_j + j^{-2}\theta_{b,j}, \]

where $\theta_{b,j}$ is a term of order $j^{-2}$. Then we have

\[ a_{n+1} = a_1 \prod_{j=1}^{n} \varepsilon_j + \sum_{i=1}^{n} \frac{h_i}{i} \prod_{j=i+1}^{n} \varepsilon_j. \]
where \( |\theta_{b,j}| \) is bounded uniformly in \( j > j_0 \). Invoking the approximation

\[
\sum_{1 \leq i \leq n} i^{-1} = \ln n + \gamma + r_n,
\]

(49)

with \( \gamma \) denoting Euler’s constant and \( |r_n| \leq 2n^{-1} \), we write, for \( i \geq j_0 \), see (48)

\[
\prod_{j=1}^{n} \varepsilon_j = \exp \left\{ \sum_{j=i+1}^{n} \ln \varepsilon_j \right\} = \left( \frac{i}{n} \right)^b (1 + \tau_i).
\]

(50)

Here \( |\tau_i| \leq c_1 \sum_{j>i} j^{-1} |\delta_j| + c_2 i^{-1} = o(1) \) as \( i \to +\infty \). \( c_1 \) and \( c_2 \) denote absolute constants.

In order to show that \( a_n \to hb^{-1} \) we invoke (50) in (47) and obtain

\[
a_{n+1} = \sum_{i=j_0}^{n} \frac{h_i}{i} \left( \frac{i}{n} \right)^b (1 + \tau_i) + O(n^{-b})
\]

\[
= \sum_{i=j_0}^{n} \frac{h_i}{i} \left( \frac{i}{n} \right)^b + R + O(n^{-b}),
\]

(51)

where \( |R| \leq n^{-b} \sum_{i=j_0}^{n} i^{b-1} (|\tau_i h_i| + |h - h_i|) = o(1) \). Furthermore, we have

\[
\sum_{i=j_0}^{n} \frac{h_i}{i} \left( \frac{i}{n} \right)^b = \sum_{i=1}^{n} \frac{h_i}{i} \left( \frac{i}{n} \right)^b + O(n^{-b} - h b^{-1} + O(n^{-1}) + O(n^{-b}).
\]

Finally, assuming that \( h - h_n = O(n^{-1}) \) and \( b - b_n = O(n^{-1}) \) we obtain \( R = O(n^{-1}) \) for \( b \neq 1 \), and we obtain \( R = O(n^{-1} \ln n) \) for \( b = 1 \).

Proof of (ii). We remark that (48) implies that the numbers \( \eta_j(m) = \prod_{1 \leq i \leq m} \varepsilon_{i+j} \) converge to 1 as \( j, m \to +\infty \). More precisely, we have \( \sup_{m \geq 1} |\eta_j(m) - 1| = O(1) \) as \( j \to +\infty \). Furthermore, there exists a constant \( c^* > 0 \) such that \( |\eta_j(m)| \leq c^* \) for every \( 0 \leq j < n \). Now, using (47) we write

\[
a_{n+1} \ln^{-1} n = \ln^{-1} n \left( \eta_0(n) + \sum_{1 \leq j \leq \ln \ln n} \frac{h_j}{j} \eta_j(n-j) \right) + \ln^{-1} n \sum_{\ln \ln n < j \leq n} \frac{h_j}{j} \eta_j(n-j)
\]

and observe that the first summand in the right is \( o(1) \), and the second summand converges to \( \lim_n \ln^{-1} n \sum_{\ln \ln n < j \leq n} h_j^{-1} = h \).

Proof of (iii). We recall that now \( b < 0 \). Denote \( z_0(n) = n^b \prod_{j=1}^{n} \varepsilon_j \). For \( 0 \leq i < n \) denote \( z_i(n) = (n/i)^b \prod_{j=i+1}^{n} \varepsilon_j \). We first prove the statement (iii′): for every \( i \) the sequence \( \{z_i(n)\}_n \) converges to some limit \( z_i \), and there exists an absolute constant \( c' > 0 \) such that \( |z_i(n)| \leq c' \), for every \( 0 \leq i \leq n \).

Given \( i \geq j_0 \), we write using (49)

\[
z_i(n) = \exp \left\{ \sum_{j=i+1}^{n} \ln \varepsilon_j + b \sum_{j=i+1}^{n} j^{-1} + b(r_n - r_i) \right\}.
\]

(52)

Since \( r_i - r_n \to r_i \) as \( n \to +\infty \), and the series \( \sum_{j} |\ln \varepsilon_j + bj^{-1}| \) converges, see (48), we obtain that \( \{z_i(n)\}_n \) converges. Furthermore, since the sequence \( \{r_n\}_n \) is bounded, we
conclude from (52) that
\[ |z_i(n)| \leq \exp \left\{ 2|b| \max_{m \geq 1} |r_m| + \sum_{j \geq 1} |\ln \varepsilon_j + bj^{-1}| \right\} =: c''. \]

For \( 0 \leq i \leq j_0 \), we write \( z_i(n) = \tilde{z}_i z_{j_0}(n) \), where \( \tilde{z}_i = (j_0/i)^b \prod_{i+1 \leq j \leq j_0} \varepsilon_j \) is a constant. Hence, for every \( i \geq 0 \) the sequence \( \{z_i(n)\}_n \) converges and \( |z_i(n)| \leq (1 + \max_{0 \leq i \leq j_0} |\tilde{z}_i|)c'' =: c' \).

Now, using (47) we obtain
\[ n^b a_{n+1} = a_1 z_0(n) + \sum_{i=1}^{n} h_i i^{b-1} z_i(n) \to a_1 z_0 + \sum_{i \geq 1} h_i i^{b-1} z_i =: c_0 \quad \text{as} \quad n \to +\infty. \]

The last statement of the lemma about the convergence \( \tilde{a}_n \to 0 \) follows from (i).

\[ \Box \]

Calculation of Moments

Lemma 4 For \( n \to +\infty \) relations (37)–(45) hold true.

Before the proof of Lemma 4 we introduce some notation and collect auxiliary results. The proof is postponed up to the end of the section.

Let \( N(v) \) denote the set of bins/attributes related to \( v \). Denote \( \mathcal{X} = \mathcal{X}(n) = \{N(v), \ v \in V_n\} \).

By \( E_\mathcal{X} \) and \( P_\mathcal{X} \) we denote the conditional expectation and probability given \( \mathcal{X} \). Furthermore, for a random vector \( (\xi, \eta) \) we denote the conditional covariance \( \text{Cov}_\mathcal{X}(\xi, \eta) = E_\mathcal{X} \xi \eta - (E_\mathcal{X} \xi)(E_\mathcal{X} \eta) \). Let \( \mathcal{Z}(n) \) denote the number of balls produced by the vertex \( v_n \). For \( i = 0, 1, \ldots \) denote
\[ q_{X,i} = q_{X,i}^{(n)} = P_\mathcal{X}(Z^{(n+1)} = i), \quad q_i^{(n)} = P(Z^{(n+1)} = i), \quad f_i(\lambda) = e^{-\lambda \lambda^i}/i!. \]

Let \( \mathcal{Y}_{i,r}^{(n)} \) (respectively \( \tilde{\mathcal{Y}}_{i,r}^{(n)} \)) denote the set of regular (respectively non regular) vertices of \( G_n \) of degree \( i \) and activity \( r \). Denote events \( \mathcal{V}_{i,r} = \{v_{n+1} = \mathcal{Y}_{i,r}^{(n+1)}\}, \tilde{\mathcal{V}}_{i,r} = \{v_{n+1} = \tilde{\mathcal{Y}}_{i,r}^{(n+1)}\} \),
\[ \mathcal{A}_u = \{u \sim v_{n+1}\}, \quad \mathcal{A}_v^+ = \mathcal{A}_v \cap \{v \in \mathcal{Y}_{i,r}^{(n+1)}\}, \quad \mathcal{A}_v^- = \mathcal{A}_v \cap \{v \in \tilde{\mathcal{Y}}_{i,r}^{(n+1)}\} \]
and their indicators \( I_{i,r} = I_{\mathcal{V}_{i,r}}, \tilde{I}_{i,r} = I_{\tilde{\mathcal{V}}_{i,r}}, I_v = I_{\mathcal{A}_v}, \text{ and } I_v^+ = I_{\mathcal{A}_v^+}, I_v^- = I_{\mathcal{A}_v^-}. \) Here \( u \in V_n \) and \( v \in \mathcal{Y}_{i-1,r}^{(n)} \). Furthermore, denote \( p_{u,\mathcal{X}} = P_\mathcal{X}(\mathcal{A}_u) \) and \( p_{v,\mathcal{X}} = P_\mathcal{X}(\mathcal{A}_v^+) \).

Lemma 5 For every \( n \geq 1 \) we have
\[ E \sum_{i \geq 1} i X_i^{(n)} = l + nk + n\lambda. \tag{53} \]

For \( n \to +\infty \) we have
\[ (nk)^{-1} E \sum_{i \geq 1} i^2 X_i^{(n)} = \begin{cases} \frac{2+\alpha+\alpha^{-1}}{\alpha-1} + o(1), & \text{for } \alpha > 1, \\ \frac{2+\alpha+\alpha^{-1}}{\alpha+1} \ln n + o(\ln n), & \text{for } \alpha = 1, \\ n^{(1-\alpha)/(1+\alpha)}(c + o(1)), & \text{for } \alpha < 1. \end{cases} \tag{54} \]

Here \( c > 0 \) is a constant depending on \( \lambda \) and \( k \).
Proof of Corollary 1

From these identities we obtain the recurrence relation for the number of balls in the network after the $n$-th step. Hence $E S_1 = l + k + \lambda$. For $n \geq 1$ we proceed recursively. The expected number of balls brought to the network by the vertex $v_{n+1}$ equals

$$EZ^{(n+1)} = E E_X Z^{(n+1)} = E\left(x_n(S_n + k)\right) = x_n E(S_n + k) = x_n(l + (n + 1)k + n\lambda) = \lambda.$$ 

Let us prove (54). The identities

$$E_X X_1^{(n+1)} = X_1^{(n)}(1 - p_1) + k(1 - p_1),$$

$$E_X X_2^{(n+1)} = X_2^{(n)}(1 - p_2) + X_1^{(n)} p_1 + kp_1,$$

$$E_X X_i^{(n+1)} = X_i^{(n)}(1 - p_i) + X_{i-1}^{(n)} p_{i-1}, \quad 3 \leq i \leq n + 1,$$

$$E_X X_{n+2}^{(n+1)} = X_{n+1}^{(n)} p_{n+1}$$

imply the first identity below

$$E_X \sum_{i \geq 1} i^2 X_i^{(n+1)} = \sum_{i \geq 1} i^2 X_i^{(n)} + 2 \sum_{i \geq 1} i p_i X_i^{(n)} + \sum_{i \geq 1} X_i^{(n)} p_i + k(1 + 3 p_1)$$

$$= (1 + 2x_n) \sum_{i \geq 1} i^2 X_i^{(n)} + x_n \sum_{i \geq 1} i X_i^{(n)} + k(1 + 3 p_1).$$

From these identities we obtain the recurrence relation for $a_n = (nk)^{-1} E \sum_{i \geq 1} i^2 X_i^{(n)}$

$$a_{n+1} = a_n \left(1 - n^{-1} + 2(1 + \alpha)^{-1} n^{-1} + O(n^{-2})\right) + ((1 + \alpha)^{-1}(1 + \alpha)^{-1} + 1)n^{-1} + O(n^{-2}).$$

Now (54) follows by Lemma 3.

Corollary 1 For any integer $r \geq 0$ we have $(nk)^{-1} E \sum_{i \geq 1} i X_i^{(n)} = \sum_{i \geq 1} i x_i + O(n^{-1}).$

Proof of Corollary 1 From (53) and (7) we obtain

$$(nk)^{-1} E \sum_{i \geq 1} i X_i^{(n)} = 1 + \alpha^{-1} + (nk)^{-1} l = \sum_{i \geq 1} i x_i + (nk)^{-1} l.$$ 

Now the corollary follows from the relation, see the first identity of (25),

$$(nk)^{-1} E \sum_{1 \leq j \leq r} i X_i^{(n)} = \sum_{1 \leq j \leq r} i x_i + O(n^{-1}).$$

Lemma 6 Let $i, r \geq 0$ be integers. We have as $n \to +\infty$

$$E [q_{X,i} - f_i(\lambda)] = o(1), \quad q_i^{(n)} = f_i(\lambda) + o(1),$$

$$E [P_X(V_i,r) - c_{i,r}] = o(1), \quad P(V_i,r) = c_{i,r} + o(1)$$

$$P(\hat{V}_{i,r}) = o(1).$$

For any vertex $v \in \mathcal{V}_{i,r}^{(n)}$ we have

$$p_{v,X} = s_{i,r} x_n + O(n^{-2}), \quad p_{v,X}^- \leq 2(r + 2^{-1} i)^2 x_n^2, \quad p_{v,X}^+ = s_{i,r} x_n + O(n^{-2}),$$

where the remainder $O(n^{-2})$ is uniform in $X_n$ but depends on $i, r, \alpha$. We recall that $s_{i,r} = i + 2r$ is the total number of balls contained in attributes related to $v \in \mathcal{V}_{i,r}^{(n)}$. 

Springer
For any vertex \( u \in \tilde{Y}_{i,r}^{(n)} \) we have
\[
p_{u,\lambda} \leq s_{i,r,r} x_n + O(n^{-2}).
\] (60)

For any \( u, v \in V_n \) we have
\[
E_{\lambda,\mu} \mathbb{I}_{u,v} = p_{u,\lambda} p_{v,\lambda} + \mathbb{I}_{v\sim u} \text{Cov}_{X}(\mathbb{I}_u, \mathbb{I}_v), \quad p_{u,\lambda}(1 - p_{v,\lambda}) \geq \text{Cov}_{X}(\mathbb{I}_u, \mathbb{I}_v) \geq 0.
\] (61)

Here and below \( \mathbb{I}_{v\sim u} \) denotes the indicator of the event that \( u \) and \( v \) are adjacent.

For each \( i \geq 1 \) and \( n > 2i + 4r \), we have, for any \( v \in \tilde{Y}_{i,r}^{(n)} \) and \( u \in V_n \), that
\[
E_{\lambda,\mu} \mathbb{I}_{u,v}^+ = p_{u,\lambda} p_{v,\lambda}^+ + \mathbb{I}_{v\sim u} \text{Cov}_{X}(\mathbb{I}_u, \mathbb{I}_v^+), \quad p_{u,\lambda}(1 - p_{v,\lambda}^+) \geq \text{Cov}_{X}(\mathbb{I}_u, \mathbb{I}_v^+) \geq 0.
\] (62)

For each \( i, j \geq 1 \) and \( n > 2(i + j) + 4(r_1 + r_2) \), we have, for any \( v \in \tilde{Y}_{i,r_1}^{(n)} \) and \( u \in \tilde{Y}_{j,r_2}^{(n)} \), that
\[
E_{\lambda,\mu} \mathbb{I}_{u,v}^{++} = p_{u,\lambda} p_{v,\lambda}^{++} + \mathbb{I}_{v\sim u} \text{Cov}_{X}(\mathbb{I}_u^+, \mathbb{I}_v^+), \quad p_{u,\lambda}(1 - p_{v,\lambda}^{++}) \geq \text{Cov}_{X}(\mathbb{I}_u^+, \mathbb{I}_v^+) \geq 0.
\] (63)

Remark 1 Let \( (A_1, \ldots, A_r) \) be a random vector such that \(|A_i| \leq 1\) almost surely, for \( 1 \leq i \leq r \). Let \((a_1, \ldots, a_r)\) be a non-random vector such that \(|a_i| \leq 1\), \(1 \leq i \leq n\). Then
\[
|E(A_1 \ldots A_r) - a_1 \ldots a_n| \leq E|A_1 - a_1| + \cdots + E|A_r - a_r|.
\]

Proof of Lemma 6 The first relation of (56) implies the second one. Let us prove the first relation. Denote
\[
\lambda_{\lambda} = E(Z^{(n+1)}|X) = \sum_{i \geq 1} X_i^{(n)} p_i^* + kp_1,
\]
\[
\Delta_1 = q_{\lambda,i} - f_i(\lambda_{\lambda}), \quad \Delta_2 = f_i(\lambda_{\lambda}) - f_i(\lambda).
\]

We write \( |q_{\lambda,i} - f_i(\lambda)| \leq E|\Delta_1| + E|\Delta_2| \) and estimate \( E|\Delta_i|, i = 1, 2 \). We apply Le Cam’s inequality (see, e.g., [14]) to \( \Delta_1 \) and obtain
\[
|\Delta_1| \leq 2 \sum_{j \geq 1} p_j^2 X_j^{(n)}. \tag{64}
\]

Now (54) implies
\[
E|\Delta_1| = \begin{cases} O(n^{-1}), & \text{for } \alpha > 1, \\ O(n^{-1} \ln n), & \text{for } \alpha = 1, \\ O(n^{-2\alpha/(1+\alpha)}), & \text{for } \alpha < 1. \end{cases} \tag{65}
\]

Next, we apply the mean value theorem to the function \( t \to f_i(t) \) and write
\[
|\Delta_2| \leq |\lambda_{\lambda} - \lambda|.
\] (66)

Then in (66) we substitute \( \lambda = \frac{\lambda k}{\mu + k} \sum_{j \geq 1} jx_j \), see (7), and \( \lambda_{\lambda} = \lambda_{\lambda}' - R \), see (24), where
\[
\lambda_{\lambda}' = \frac{\lambda k}{\mu} \sum_{j \geq 1} j x_j^{(n)}, \quad R = \frac{1}{n} \frac{\alpha + \beta}{1 + \alpha + (\alpha + \beta)/n} \lambda_{\lambda}' - kp_1.
\]
Invoking the bound $\mathbf{E}|R| = O(n^{-1})$, see (53), we obtain, for any fixed integer $r > 0$,

$$\mathbf{E}|\lambda - \lambda| \leq \mathbf{E}|\lambda' - \lambda| + O(n^{-1})$$

$$\leq \frac{\lambda k}{\lambda + k} \sum_{j > r} j \frac{X_j(n)}{nk} + \frac{\lambda k}{\lambda + k} \sum_{j > r} j x_j + \sum_{1 \leq j \leq r} j \mathbf{E} \left| \frac{X_j(n)}{nk} - x_j \right| + O(n^{-1})$$

$$\leq \frac{2\lambda k}{\lambda + k} \sum_{j > r} j x_j + O(n^{-1/2}).$$

(67)

In the last step we used Corollary 1 and estimated, see (25),

$$\left( \mathbf{E} \left| \frac{X_j(n)}{nk} - x_j \right| ^2 \right) \leq \mathbf{E} \left( \frac{X_j(n)}{nk} - x_j \right)^2 = O(n^{-1}).$$

We remark that the remainder term $O(n^{-1/2})$ in (67) depends on $r$. Choosing $r$ large enough and then letting $n \to +\infty$ we can make the right-hand side of (67) arbitrarily small. Hence $\mathbf{E} |\Delta| = o(1)$ as $n \to +\infty$. The proof of (56) is complete.

Let us prove (57). Since the second relation follows from the first one we only prove the first relation. Introduce events

$$\mathcal{A} = \{v_{n+1}\} \text{ is a regular vertex of } G_{n+1},$$

$$\mathcal{B} = \{v_{n+1} \text{ has activity } r\}, \quad \mathcal{C} = \{v_{n+1} \text{ has degree } i \text{ in } G_{n+1}\},$$

$$\mathcal{D} = \{\text{after } n + 1 \text{ steps the number of balls in bins } w \in W_{n+1} \text{ related to } v_{n+1} \text{ is } i + 2r\}.$$

For an event $\mathcal{H}$, we denote by $\mathcal{H}'$ the complement event, and $\mathcal{I}_{\mathcal{H}}$ will denote the indicator of $\mathcal{H}$. In the proof we use the observation that $\mathcal{I}_{V_{i,r}} = \mathcal{I}_{\mathcal{A}} \mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{C}} = \mathcal{I}_{\mathcal{A}} \mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{D}}$. These identities imply

$$\mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{D}} - \mathcal{I}_{\mathcal{A}} \mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{D}} \leq \mathcal{I}_{\mathcal{A}} \mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{D}} = \mathcal{I}_{V_{i,r}} \leq \mathcal{I}_{\mathcal{B}} \mathcal{I}_{\mathcal{D}},$$

and the later inequalities imply

$$\mathbf{E} |\mathbf{P}_{\mathcal{X}}(V_{i,r}) - c_{i,r}| \leq \mathbf{E} |\mathbf{P}_{\mathcal{X}}(\mathcal{B} \cap \mathcal{D}) - c_{i,r}| + \mathbf{P}(\mathcal{A} \cap \mathcal{B} \cap \mathcal{D}).$$

Next, we show that both summands on the right hand side are $o(1)$. We have, cf. (8),

$$\mathbf{P}(\mathcal{B} \cap \mathcal{D} | \mathcal{X}) = q_{\mathcal{X}, r} \mathbf{P}(\mathcal{D} | \mathcal{X}, \mathcal{X}),$$

$$\mathbf{P}(\mathcal{D} | \mathcal{X}, \mathcal{X}) = \sum_{u_1 + \cdots + u_{i+1} = r, \text{ } 1u_2 + 2u_3 + \cdots + iu_{i+1} = i} \frac{(X_1^{(n)} + k)_{u_1} (X_2^{(n)})_{u_2} \cdots (X_{i+1}^{(n)})_{u_{i+1}} r!}{(n+1)k)_r u_1! \cdots u_{i+1}!}.$$

Here $u_i$ counts balls sent by $v_{n+1}$ to attributes of score $i$. Invoking the bound

$$\mathbf{E} q_{\mathcal{X}, r} \mathbf{P}(\mathcal{D} | \mathcal{Z}^{n+1} = r, \mathcal{X}) - f_r(\lambda) \sum_{u_1 + \cdots + u_{i+1} = r, \text{ } 1u_2 + 2u_3 + \cdots + iu_{i+1} = i} x_1^{u_1} \cdots x_{i+1}^{u_{i+1}} \frac{r!}{u_1! \cdots u_{i+1}!} = o(1),$$

which follows from (25) and (56) by Remark 1, we obtain $\mathbf{E} |\mathbf{P}_{\mathcal{X}}(\mathcal{B} \cap \mathcal{D}) - c_{i,r}| = o(1).$
Let us show that $\mathbf{P}(\tilde{A} \cap \mathcal{B} \cap \mathcal{D}) = o(1)$. Let $w_1', w_2', \ldots, w_r'$ denote an array of bins related to $v_{n+1}$, and let $t_1 = s_n(w_1'), \ldots, t_r = s_n(w_r')$ denote scores of the bins. Let $S_1 = \{v'_1, \ldots, v'_{t_1-1}\}, \ldots, S_r = \{v'_{t_1 + \cdots + t_{r-1} - r + 1}, \ldots, v'_{t_1 + \cdots + t_r - r}\}$ denote the sets of vertices from $V_n$ related to bins $w_1', \ldots, w_r'$ respectively. Note that every $t_j$ does not exceed $i$, since the degree of $v_{n+1}$ in $G_{n+1}$ is $i$. We may assume, without loss of generality, that the first ball of $v_{n+1}$ was sent to $w_1'$, the second ball was sent to $w_2'$, etc. After sending the first ball to $w_1'$ we read the names of vertices from $S_1$. They are all different. Then we send the second ball and read the names of vertices from $S_2$. It may happen that some of these names coincide with those from $S_1$. We continue the process and count all the coincidences one after another. The sum obtained is denoted $H = \sum_{1 \leq j_1 < j_2 \leq r} \sum_{u \in S_{j_1}} \sum_{v \in S_{j_2}} \mathbb{P}_{u \rightarrow v}$.

Here we write $\mathbb{P}_{u \rightarrow v} = 1$ whenever $v = u$. Clearly, the fact that $v_{n+1}$ is a non regular vertex of $G_{n+1}$ will be witnessed by at least one coincidence. Hence, $\mathbf{P}(\tilde{A} \cap \mathcal{B} \cap \mathcal{D}) \leq \mathbf{P}(H \geq 1) \leq \mathbf{E}H$. Now we show that $\mathbf{E}H = o(1)$. We firstly note that the event $\mathcal{I}_m = \{W_m \cap \{w_1', \ldots, w_r'\} \neq \emptyset\}$ has probability $o(1)$ for $m = m_n = o(n)$. Indeed, let us choose $m' = m'_n$ such that $m'm = o(n)$ and $m'_n \rightarrow +\infty$ as $n \rightarrow +\infty$. We have $\mathbf{P}(\mathcal{I}_m) \leq \sum_{j'} \mathbf{P}(w_{j'} \in W_m) = \sum_{j'} \mathbf{P}\left(\sum_{j' \in W_m} X^{(n)}_{t_{j'}} \geq m'mx_{t_{j'}}\right) - o(1),$

because $\mathbf{P}(X^{(n)}_{t_{j'}} < m'mx_{t_{j'}}) = o(1)$, by Lemma 1. Furthermore, since $|W_m| = O(m)$, we obtain $\mathbf{P}\left(w_{j'} \in W_m, X^{(n)}_{t_{j'}} \geq m'mx_{t_{j'}}\right) \leq \mathbf{P}\left(w_{j'} \in W_m, X^{(n)}_{t_{j'}} \geq m'mx_{t_{j'}}\right) = O(m/(m'm)) = o(1)$.

We secondly evaluate expected values of summands of $H$. Fix $j_1 < j_2$ and $u \in S_{j_1}$ and $v \in S_{j_2}$. The summand $\mathbb{I}_{u \rightarrow v}^{(n)}$ attains value 1 whenever the vertex $u$ (its name is already known from the set $S_{j_1}$) has sent a ball to $w_{j_2}$. The probability of such an event is at most $p_{i, n_\mu}$, where $n_\mu$ is the number of $u$ (customer $u$ entered the network on the $n_\mu$th step). Now we choose $m \approx \ln n$ and obtain $\mathbf{E}H = \mathbf{E}H \mathbb{I}_{\tilde{I}_m} + \mathbf{E}H \mathbb{I}_{\tilde{I}_m} = O(\ln^{-1} n) + o(1)$.

Here we use the fact that on the event $\tilde{I}_m$ we have $p_{i, n_\mu} \leq c \ln n$ for every $u \in S_{j_1} \cup S_{j_2}$, and also use the simple deterministic upper bound that holds uniformly in $n$

$$H \leq \sum_{1 \leq j_1 < j_2 \leq r} \sum_{u \in S_{j_1}} \sum_{v \in S_{j_2}} 1 \leq \binom{r}{2} i^2.$$
Here \( \sum_{w \in N(v)} p_s(w) = s_{i,r}x_n = (i + 2r)x_n \) and
\[
0 \leq R_v \leq \sum_{\{w, w'\} \subset N(v)} p_s(w)p_s(w') \leq x_n^2 \sum_{\{w, w'\} \subset N(v)} s(w)s(w') \leq x_n^2(2r + i/2)^2. \tag{69}
\]
In the last inequality we used the fact that \( s(w) \geq 2 \), for \( w \in N(v) \), and \( \sum_{w \in N(v)} s(w) = i + 2r \). The first relation of (59) follows from (68), (69). The second relation follows from the inequalities
\[
p_v^- \leq \sum_{\{w, w'\} \subset N(v)} P_\xi(v_{n+1} \text{ is related to } w \\ \text{and } w') = \sum_{\{w, w'\} \subset N(v)} p_s(w)p_s(w')
\]
and (69). The last relation of (59) follows from the first two, since \( p_v^+ + p_v^- = p_v. \)

Let us prove (60). We observe that every bin \( w \) related to \( u \) may have at most \( i + 2 \) balls. Hence,
\[
\sum_{w \in N(u)} p_s(w) \leq r(i + 2)x_n = s_{i,r}x_n, \quad \sum_{\{w, w'\} \subset N(u)} p_s(w)p_s(w') \leq \left(\frac{r}{2}\right)(i + 2)^2x_n^2.
\]
Next, we proceed as in (68), (69) and obtain (60).

Let us prove (61). The first identity follows from the fact that for non-adjacent vertices \( u \) and \( v \) the random variables \( I_u \) and \( I_v \) are conditionally independent, given \( \xi \). The remaining inequalities of (61) follows from the inequalities
\[
p_{v,\xi}p_{u,\xi} \leq E_{\xi}I_uI_v \leq E_{\xi}I_vI_u = p_u,\xi.
\]
We only prove the first inequality. Denote \( A, B, C \subset W_n \) the distinct sets of attributes such that all attributes from \( A \cup B \) are related to \( v \) and all attributes from \( B \cup C \) are related to \( u \). Let \( a, b, c \) denote the numbers of balls sent by \( v_{n+1} \) to attributes from \( A, B \) and \( C \) respectively. Then \( I_u = 1 \) whenever \( a + b \geq 1 \) and \( I_v = 1 \) whenever \( b + c \geq 1 \). We have
\[
E_{\xi}I_uI_v = P_\xi(a + b \geq 1, b + c \geq 1)
= P_\xi(a \geq 1, a + b \geq 1, b + c \geq 1) + P_\xi(a = 0, a + b \geq 1, b + c \geq 1)
= P_\xi(a \geq 1, b + c \geq 1) + P_\xi(a = 0, b \geq 1)
= P_\xi(a \geq 1)P_\xi(b + c \geq 1) + P_\xi(a = 0)P_\xi(b \geq 1).
\]
Now, we substitute \( P_\xi(a \geq 1) = p_{v,\xi} - P_\xi(a = 0)P_\xi(b \geq 1) \) and \( P_\xi(b + c \geq 1) = p_{u,\xi} \) and obtain
\[
E_{\xi}I_uI_v = (p_{v,\xi} - P_\xi(a = 0)P_\xi(b \geq 1))p_{u,\xi} + P_\xi(a = 0)P_\xi(b \geq 1) \geq p_{v,\xi}p_{u,\xi}.
\]
Let us prove (62). The proof is similar to that of (61). We only show that \( p_{v,\xi}p_{u,\xi} \leq E_{\xi}I_uI_v \). We note that since \( v \) is regular, the set \( B \) consists of a single bin and we have \( A_v^+ = \{a + b = 1\} \) and \( A_u = \{b + c \geq 1\} \). Hence, \( E_{\xi}I_vI_u = P_\xi(a + b = 1, b + c \geq 1) \). The identities
\[
P_\xi(a + b = 1, b + c \geq 1) = P_\xi(a + b = 1) - P_\xi(a = 1)P_\xi(b + c = 0)
= P_\xi(a + b = 1)P_\xi(b + c \geq 1) + \delta P_\xi(b + c = 0),
\]
where \( \delta = P_\xi(a + b = 1) - P_\xi(a = 1) \), combined with the inequality \( \delta \geq 0 \) complete the proof. We note that the inequality \( \delta \geq 0 \) is trivial for \( A = \emptyset \). For \( A \neq \emptyset \) this inequality follows from the identity
\[
\delta = P_\xi(a = 0)P_\xi(b = 1) + P_\xi(a = 1)P_\xi(b = 0) - P_\xi(a = 1),
\]
and the inequalities $\mathbf{P}_X(a = 0) \geq 2^{-1} \geq \mathbf{P}_X(a = 1)$. It suffices to verify the first of these inequalities. We have

$$
\mathbf{P}_X(a = 0) = \prod_{u \in A} (1 - p_{X(u)}) \geq 1 - \sum_{u \in A} p_{X(u)} \geq 1 - s_{i,r} \kappa_n > 2^{-1}.
$$

(70)

The latter inequality holds for sufficiently large $n$, e.g., for $n \geq 2s_{i,r}$.

Let us prove (63). The proof is similar to that of (61), (62). We only show that $p^+_v \mathbf{P}^+_u \leq \mathbf{E}_X^+ \mathbf{I}^+ u$. We note that since $u$ and $v$ are regular, the set $B$ consists of a single bin and we have $A^+_v = \{a + b = 1\}$ and $A^+_u = \{b + c = 1\}$. Hence, $\mathbf{E}_X^+ \mathbf{I}^+ u = \mathbf{P}_X(a + b = 1, b + c = 1)$. Using the notation $x_\varepsilon = \mathbf{P}_X(x = \varepsilon)$, for $x = a, b, c$ and $\varepsilon = 0, 1$, we write

$$
\mathbf{E}_X^+ \mathbf{I}^+ u - \mathbf{E}_X^+ \mathbf{I}^+ v \mathbf{E}^+ u = (a_0 b_0 c_0 + a_0 b_0 c_1) - (a_0 b_1 + a_1 b_0) (b_1 c_0 + b_1 c_1) = ((a_0 b_1 c_0 - a_0 b_1 c_0) - a_0 b_0 b_1 c_1 + (a_1 b_0 c_1 - a_1 b_0 c_1) - a_1 b_0 b_1 c_0) = (a_0 b_0 b_1 c_0 - a_0 b_0 b_1 c_1 + (a_1 b_0 b_1 c_1 - a_1 b_0 b_1 c_1) = a_0 b_0 b_1 c_0 + a_1 b_0 b_1 c_1 - c_0) = b_0 b_1 (a_0 - a_1) (c_0 - c_1).
$$

Here we used the observation that $b_0 + b_1 = 1$. Now, the inequalities $a_0 \geq 2^{-1} \geq a_1$ and $c_0 \geq 2^{-1} \geq 1$ which hold for $n \geq 2i + 4r_1$ and $n \geq 2j + 4r_2$, see (70), imply $\mathbf{E}_X^+ \mathbf{I}^+ u - \mathbf{E}_X^+ \mathbf{I}^+ v \mathbf{E}^+ u \geq 0$. □

**Proof of Lemma 4**

Before the proof we introduce some notation. Given $r \geq 1$ denote

$$
\tilde{g}^{(n)}_{i,j,r} = n^{-2} \mathbf{E} Y_{i,r}^{(n)} Y_{j,r}^{(n)},
$$

$$
A_j = Y_{i,r}^{(n+1)} = \sum_{v \in \mathcal{Y}^{(n)}_{i,r}} (1 - \mathbb{I}_v) + \sum_{u \in \mathcal{Y}^{(n)}_{i-1,r}} \mathbb{I}^+ u + \mathbb{I}_{j,r},
$$

$$
A_0 = Y_{0,r}^{(n+1)} = \sum_{v' \in \mathcal{Y}^{(n)}_{0,r}} (1 - \mathbb{I}_v) + \mathbb{I}_{0,r} = A_{0,1} + A_{0,2}.
$$

(71)

(72)

Proof of (39), (42). These relations follow from (57) and the identity

$$
Y_{0,0}^{(n+1)} = Y_{0,0}^{(n)} + \mathbb{I}_{0,0}.
$$

Proof of (37), (38). (37) follows from the obvious identities $\tilde{Y}_{0,r}^{(n)} = 0$ and $\tilde{Y}_{1,r}^{(n)} = 0$. Let us prove (38). We observe that the new vertex $v_{n+1}$ moves its neighbours from $\tilde{Y}_{i-1,r}^{(n)}$ to the set $\tilde{Y}_{i+1,r}^{(n+1)}$, the neighbours from $\tilde{Y}_{i,r}^{(n)}$ are moved to $\tilde{Y}_{i+1,r}^{(n+1)}$. Furthermore, the neighbours of $v_{n+1}$ from $\tilde{Y}_{i,r}^{(n)}$ that share with $v_{n+1}$ at least two bins are moved to $\tilde{Y}_{i+1,r}^{(n+1)}$. Finally, the vertex $v_{n+1}$ itself may become irregular upon arrival. Hence we have

$$
\tilde{Y}_{i+1,r}^{(n+1)} = \sum_{v \in \tilde{Y}_{i+1,r}^{(n)}} (1 - \mathbb{I}_v) + \sum_{u \in \tilde{Y}_{i,r}^{(n)}} \mathbb{I}_u + \sum_{v' \in \tilde{Y}_{i,r}^{(n)}} \mathbb{I}_{v'} + \mathbb{I}_{i+1,r}.
$$

Using the inequality $1 - \mathbb{I}_v \leq 1$ and (58)–(60) we obtain

$$
\mathbf{E}_X \tilde{Y}_{i+1,r}^{(n)} \leq \tilde{Y}_{i+1,r}^{(n)} + (s_{i,r} + O(n^{-2}) \kappa_n \tilde{Y}_{i,r}^{(n)} + 2(r + 2^{-1} i) \kappa_n 2^{(n)} + o(1).
$$

The latter inequality implies (38).
Proof of (40), (41). It follows from (71) that
\[ E_{X}Y^{(n+1)}_{j, r} = (1 - p_{v, r}) Y^{(n)}_{j, r} + p_{u, r} Y^{(n)}_{j-1, r} + P_{X}(V_{j, r}). \]

Invoking (59) and (57) we obtain (41). Relation (40) is derived from (72) in much the same way.

Proof of (43)–(45). In the proof we use the approximations of moments $E_{A_{i, s} A_{j, t}}$ shown below. For $i, j \geq 1$ we have
\[ E_{A_{i, 1} A_{j, 1}} = (1 - s_{i, r} \kappa_n - s_{j, r} \kappa_n) n^2 \bar{g}^{(n)}_{i, j, r} + O(1), \quad (73) \]
\[ E_{A_{i, 1} A_{j, 2}} = s_{j-1, r} \kappa_n n^2 \bar{g}^{(n)}_{i, j-1, r} + O(1), \quad (74) \]
\[ E_{A_{i, 1} A_{j, 3}} = n \left( y^{(n)}_{i, r} c_{j, r} + o(1) \right), \quad (75) \]
\[ E_{A_{i, s} A_{j, t}} = O(1), \quad 2 \leq s, t \leq 3. \quad (76) \]

For $i \geq 1$ we have
\[ E_{A_{i, 1} A_{0, 1}} = (1 - s_{i, r} \kappa_n - s_{0, r} \kappa_n) n^2 \bar{g}^{(n)}_{i, 0, r} + O(1), \quad (77) \]
\[ E_{A_{i, 1} A_{0, 2}} = n \left( y^{(n)}_{i, r} c_{0, r} + o(1) \right), \quad (78) \]
\[ E_{A_{i, 2} A_{0, 1}} = s_{i-1, r} \kappa_n n^2 \bar{g}^{(n)}_{i, 1, 0, r} + O(1), \quad (79) \]
\[ E_{A_{i, 3} A_{0, 1}} = n \left( y^{(n)}_{0, r} c_{i, r} + o(1) \right), \quad (80) \]
\[ E_{A_{i, s} A_{0, 2}} = O(1), \quad s = 2, 3. \quad (81) \]

Furthermore, we have
\[ E_{A_{0, 1} A_{0, 1}} = (1 - 2s_{0, r} \kappa_n \kappa_n) n^2 \bar{g}^{(n)}_{0, 0, r} + O(1), \quad (82) \]
\[ E_{A_{0, 1} A_{0, 2}} = n \left( y^{(n)}_{0, r} c_{0, r} + o(1) \right), \quad (83) \]
\[ E_{A_{0, 2} A_{0, 2}} = O(1). \quad (84) \]

Proof of (43)–(45). To show (45) we substitute (73)–(76) in the identity
\[ \bar{g}^{(n+1)}_{i, j, r} = n^{-2} \sum_{1 \leq s, t \leq 3} E_{A_{i, s} A_{j, t}} \]

We obtain
\[ \bar{g}^{(n)}_{i, j, r} = \left( 1 - 2n - s_{i, r} \kappa_n - s_{j, r} \kappa_n \right) \bar{g}^{(n)}_{i, j, r} + (s_{i-1, r} \kappa_n \bar{g}^{(n)}_{i-1, j, r} + s_{j-1, r} \kappa_n \bar{g}^{(n)}_{i, j-1, r}) \kappa_n + n^{-1} \left( y^{(n)}_{i, r} c_{i, r} + y^{(n)}_{j, r} c_{j, r} \right) + o(n^{-1}). \]

Combining this relation with (41) we obtain (45). We similarly derive (44) from (77)–(81) and (40), (41). Furthermore, we derive (43) from (82)–(84) and (40).

In the rest of the proof we show (73)–(84). Let us prove (73)–(76).

Proof of (73). For $i = j$ we write
\[ E_{X} A_{i, 1} A_{i, 1} = \sum_{v \in Y^{(n)}_{i, r}} \sum_{v' \in Y^{(n)}_{i, r} \setminus \{ v \}} E_{X} (1 - \mathbb{I}_{v}) (1 - \mathbb{I}_{v'}) + \sum_{v' \in Y^{(n)}_{i, r}} E_{X} (1 - \mathbb{I}_{v'}). \]

Invoking the identity
\[ E_{X} (1 - \mathbb{I}_{v}) (1 - \mathbb{I}_{v'}) = (1 - p_{v, r}) (1 - p_{v', r}) + \mathbb{I}_{v \sim v'} \text{Cov}_{X} (\mathbb{I}_{v}, \mathbb{I}_{v'}), \quad (85) \]
we obtain
\[
\mathbb{E}_X A_{i,1}A_{i,1} = (1 - p_v, x)^2(Y_{i,r}^{(n)})^2 + ((1 - p_v, x) - (1 - p_v, x)^2)\, Y_{i,r}^{(n)} + R_{i,i},
\]
\[
R_{i,i} = \sum_{v \in \gamma_{i,r}^{(n)}} \sum_{v' \in \gamma_{i,r}^{(n)} \setminus \{v\}} \mathbb{I}_{v \sim v'} \mathbb{Cov}_X(\mathbb{I}_v, \mathbb{I}_{v'}). \tag{86}
\]

Next, we estimate \( R_{i,i} \) using the fact that \( \sum_{v' \in \gamma_{i,r}^{(n)} \setminus \{v\}} \mathbb{I}_{v \sim v'} \) does not exceed the degree of \( v \), which equals \( i \), and using inequalities (61). We obtain \( 0 \leq R_{i,i} \leq p_v, x_i Y_{i,r}^{(n)} \). Now, from (86), (59) we obtain (73).

For \( i \neq j \) the proof is almost the same. We write, see (85),
\[
\mathbb{E}_X A_{i,1}A_{j,1} = \sum_{v \in \gamma_{i,r}^{(n)}} \sum_{u \in \gamma_{j,r}^{(n)}} \mathbb{E}_X(1 - \mathbb{I}_v)(1 - \mathbb{I}_u)
\]
\[
= (1 - p_v, x)(1 - p_u, x) Y_{i,r}^{(n)} Y_{j,r}^{(n)} + R_{i,j},
\]
\[
R_{i,j} = \sum_{v \in \gamma_{i,r}^{(n)}} \sum_{v' \in \gamma_{j,r}^{(n)} \setminus \{v\}} \mathbb{I}_{v \sim v'} \mathbb{Cov}_X(\mathbb{I}_v, \mathbb{I}_{v'}). \tag{87}
\]

and estimate \( R_{i,j} \) using the fact that \( \sum_{u \in \gamma_{j,r}^{(n)}} \mathbb{I}_{v \sim v'} \) does not exceed the degree of \( v \), which equals \( i \), and using inequalities (61). We obtain \( 0 \leq R_{i,j} \leq p_v, x_i Y_{i,r}^{(n)} \). Finally, from (87), (59) we obtain (73).

Proof of (74). For \( i = j - 1 \) we use the identity \( (1 - \mathbb{I}_v)^+ = 0 \) and write
\[
\mathbb{E}_X A_{i,1}A_{j,1} = \sum_{v \in \gamma_{i,r}^{(n)}} \sum_{v' \in \gamma_{j,r}^{(n)} \setminus \{v\}} (1 - \mathbb{I}_v)^+ Y_{i,r}^{(n)} Y_{j,r}^{(n)} - 1 - R_{i,i}^+. \tag{88}
\]

Invoking the inequalities \( 0 \leq R_{i,i}^+ \leq p_v, x_i Y_{i,r}^{(n)} \), which are derived from (62) in the same way as those for \( R_{i,i} \) in (86), and using (59), we obtain (74) from (88). For \( i \neq j - 1 \) the proof is almost the same. We write
\[
\mathbb{E}_X A_{i,1}A_{j,2} = \mathbb{E}_X \sum_{v \in \gamma_{i,r}^{(n)}} \sum_{u \in \gamma_{j-1,r}^{(n)}} (1 - \mathbb{I}_v)^+ = (1 - p_v, x) p_{u, x}^+ Y_{i,r}^{(n)} Y_{j-1,r}^{(n)} - R_{i,j-1}^+, \tag{89}
\]

Invoking inequalities \( 0 \leq R_{i,j-1}^+ \leq p_v, x_i Y_{i,r}^{(n)} \), which are derived from (62) in the same way as those for \( R_{i,j-1} \) in (87), and using (59), we obtain (74) from (89).

**Proof of (75).** The relations
\[
\mathbb{E} A_{i,1}A_{j,3} = \mathbb{E} Y_{i,r}^{(n)} \mathbb{I}_{j,r} + O(1), \quad \mathbb{E} Y_{i,r}^{(n)} \mathbb{I}_{j,r} = \mathbb{E} Y_{i,r}^{(n)} \mathbb{P}_X(\mathcal{V}_{j,r})
\]
combined with (57) imply (75).
Proof of (76). We only consider the case \( s = t = 2 \). Other cases are obvious. For \( i = j \) we write
\[
E_h A_{i,2} A_{i,2} = \sum_{v \in \mathcal{Y}_i^{(n)}} \sum_{v' \in \mathcal{Y}_i^{(n)}} E_h \mathbb{I}_{v} \mathbb{I}_{v'}^+ \mathbb{I}_{v'}^+ + \sum_{v \in \mathcal{Y}_i^{(n)}} E_h \mathbb{I}_{v}^+
\]
\[
= \left( p_{v,i}^+ \right)^2 \mathcal{Y}_i^{(n)} (\mathcal{Y}_i^{(n)} - 1) + R_{i,i}^{++} + p_{v,i}^+ \mathcal{Y}_i^{(n)}.
\]
\[
R_{i,i}^{++} = \sum_{v \in \mathcal{Y}_i^{(n)}} \sum_{v' \in \mathcal{Y}_i^{(n)}} \mathbb{I}_{v} \mathbb{I}_{v'} \text{Cov}_h (\mathbb{I}_{v}^+, \mathbb{I}_{v'}^+). \tag{90}
\]
Invoking inequalities \( 0 \leq R_{i,i}^{++} \leq p_{v,i}^+ \mathcal{Y}_i^{(n)} \), see (63), and using (59) we derive (76) from (90). For \( i \neq j \) the proof is much the same. We write
\[
E_h A_{i,2} A_{j,2} = \sum_{v \in \mathcal{Y}_i^{(n)}} \sum_{u \in \mathcal{Y}_j^{(n)}} E_h \mathbb{I}_{v} \mathbb{I}_{u}^+ \mathbb{I}_{u}^+ = p_{v,i}^+ p_{u,j}^+ \mathcal{Y}_i^{(n)} \mathcal{Y}_j^{(n)} + R_{i,j}^{++},
\]
\[
R_{i,j}^{++} = \sum_{v \in \mathcal{Y}_i^{(n)}} \sum_{u \in \mathcal{Y}_j^{(n)}} \mathbb{I}_{v} \mathbb{I}_{u} \text{Cov}_h (\mathbb{I}_{v}^+, \mathbb{I}_{u}^+).
\]
and use inequalities \( 0 \leq R_{i,j}^{++} \leq p_{v,i}^+ \mathcal{Y}_i^{(n)} \) and (59).

The proof of (77)–(81) and (82)–(84) is similar to that of (73)–(76), but simpler because \( \text{Cov}_h (\mathbb{I}_v, \mathbb{I}_{v'}) = 0 \) for \( v' \in \mathcal{Y}_0^{(n)} \).

References

1. Aiello, W., Bonato, A., Cooper, C., Janssen, J., Prałat, P.: A spatial web graph model with local influence regions. Internet Math. 5, 173–193 (2009)
2. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Tenth Printing. National Bureau of Standards. Applied Mathematics Series, vol. 55. U.S. Government Printing Office, Washington (1972)
3. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
4. Bloznelis, M., Damarackas, J.: Degree distribution of an inhomogeneous random intersection graph. Electron. J. Comb. 20(3), R3 (2013)
5. Bloznelis, M., Kariofiški, M.: Random intersection graph process. In: Bonato, A., Mitzenmacher, M., Prałat, P. (eds.) WAW 2013. LNCS, vol. 8305, pp. 93–105. Springer, Berlin (2013)
6. Bollobás, B., Riordan, O., Spencer, J., Tusnády, G.: The degree sequence of a scale-free random graph process. Random Struct. Algorithms 18, 279–290 (2001)
7. Dereich, S., Mörters, P.: Random networks with sublinear preferential attachment: degree evolutions. Electron. J. Probab. 13, 1222–1267 (2009)
8. Guillaume, J.L., Latapy, M.: Bipartite structure of all complex networks. Inf. Process. Lett. 90, 215–221 (2004)
9. Foss, S., Korshunov, D., Zachary, S.: An Introduction to Heavy-Tailed and Subexponential Distributions. ACM, New York (2011)
10. Móri, T.F.: On random trees. Stud. Sci. Math. Hung. 39, 143–155 (2002)
11. Móri, T.F.: On a 2-parameter class of scale free random graphs. Acta Math. Hung. 114, 37–48 (2007)
12. Martin, T., Ball, B., Karrer, B., Newman, M.E.J.: Coauthorship and citation patterns in the Physical Review. Phys. Rev. E 88, 012814 (2013)
13. Newman, M.E.J., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. Proc. Natl. Acad. Sci. USA 99(Suppl. 1), 2566–2572 (2002)
14. Steele, J.M.: Le Cam’s inequality and Poisson approximations. Am. Math. Mon. 101, 48–54 (1994)