LOCAL PROPERTIES OF RANDOM MAPPINGS WITH EXCHANGEABLE IN-DEGREES

JENNIE C. HANSEN,* *Heriot-Watt University*

JERZY JAWORSKI,** *Adam Mickiewicz University*

Abstract

In this paper we investigate the 'local' properties of a random mapping model, $T_n^{\hat{D}}$, which maps the set $\{1, 2, \ldots, n\}$ into itself. The random mapping $T_n^{\hat{D}}$, which was introduced in a companion paper (Hansen and Jaworski (2008)), is constructed using a collection of exchangeable random variables $\hat{D}_1, \ldots, \hat{D}_n$ which satisfy $\sum_{i=1}^n \hat{D}_i = n$. In the random digraph, $G_n^{\hat{D}}$, which represents the mapping $T_n^{\hat{D}}$, the in-degree sequence for the vertices is given by the variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$, and, in some sense, $G_n^{\hat{D}}$ can be viewed as an analogue of the general independent degree models from random graph theory. By local properties we mean the distributions of random mapping characteristics related to a given vertex v of $G_n^{\hat{D}}$—for example, the numbers of predecessors and successors of v in $G_n^{\hat{D}}$. We show that the distribution of several variables associated with the local structure of $G_n^{\hat{D}}$ can be expressed in terms of expectations of simple functions of $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. We also consider two special examples of $T_n^{\hat{D}}$ which correspond to random mappings with preferential and anti-preferential attachment, and determine, for these examples, exact and asymptotic distributions for the local structure variables considered in this paper. These distributions are also of independent interest.

Keywords: Random mapping; local structure; exchangeable in-degrees

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60F05; 05C80

1. Introduction

The study of random mapping models was initiated independently by several authors in the 1950s (see the companion paper [9] and the references therein) and the properties of these models have received much attention in the literature. In particular, these models have been useful as models for epidemic processes, and have natural applications in cryptology (see, for example, [3], [4], [7], [8], [12], [13], [18], [20], and [21]). To date, the most widely studied models have been special cases of a general model, denoted by $T_{\boldsymbol{p}(n)}$, which can be defined as follows. Let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$, and let $\mathcal{M}_n$ denote the set of all mappings from $[n]$ into $[n]$. For each $n \geq 1$, let $\boldsymbol{p}(n) = \{p_{ij}(n)\}_{n \times n}$ be an array such that $p_{ij}(n) \geq 0$ for $1 \leq i, j \leq n$ and $\sum_{j=1}^n p_{ij}(n) = 1$ for every $1 \leq i \leq n$, and let $X_1^n, X_2^n, \ldots, X_n^n$ be independent random variables such that $\Pr\{X_i^n = j\} = p_{ij}(n)$ for all $1 \leq i, j \leq n$. Then the random mapping $T_{\boldsymbol{p}(n)} \colon [n] \to [n]$ is defined (in terms of the variables $X_1^n, X_2^n, \ldots, X_n^n$) by

$$T_{\boldsymbol{p}(n)}(i) = j \quad \text{if and only if} \quad X_i^n = j \tag{1.1}$$

183

for all $1 \le i, j \le n$. It follows from (1.1) that the distribution of $T_{\boldsymbol{p}(n)}$ is given by

$$\Pr\{T_{\boldsymbol{p}(n)} = f\} = \prod_{i=1}^{n} p_{if(i)}(n)$$

for each $f \in \mathcal{M}_n$. Any mapping $f \in \mathcal{M}_n$ can be represented as a directed graph $G(f)$ on a set of vertices labelled $1, 2, \ldots, n$, such that there is a directed edge from vertex i to vertex j in $G(f)$ if and only if $f(i) = j$. So $G_{\boldsymbol{p}(n)} \equiv G(T_{\boldsymbol{p}(n)})$ is a random directed graph on a set of vertices labelled $1, 2, \ldots, n$ which represents the action of the random mapping $T_{\boldsymbol{p}(n)}$ on $[n]$. We note that, since each vertex in $G_{\boldsymbol{p}(n)}$ has out-degree 1, the components of $G_{\boldsymbol{p}(n)}$ consist of directed cycles with directed trees attached. Also, it follows from the definition of $T_{\boldsymbol{p}(n)}$ that the variables $X_1^n, X_2^n, \ldots, X_n^n$ can be interpreted as the independent 'choices' of the vertices $1, 2, \ldots, n$ in the random digraph $G_{\boldsymbol{p}(n)}$ (see, in addition, [11] and [17]). The special case of $T_{\boldsymbol{p}(n)}$ which is best understood is the uniform random mapping, $T_n \equiv T_{\boldsymbol{p}(n)}$, where $p_{ij}(n) = 1/n$ for all $1 \le i, j \le n$ (see, for example, [2] and the monograph [15], and the references therein).

The key feature in the model discussed above is that each vertex in $G_{\boldsymbol{p}(n)}$ 'chooses' the vertex that it is mapped to independently of the 'choices' made by all other vertices. In this paper we consider the properties of a new random mapping model, $T_n^{\hat{D}}$, which was introduced in the companion paper [9], and which has the property that the vertex 'choices' are not necessarily independent. The model is introduced in Section 2, and is constructed by first specifying the in-degrees $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ of the vertices labelled $1, 2, \ldots, n$, and then selecting a random mapping uniformly from all mappings with the given in-degree sequence $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. In the context of applications, the exact and asymptotic distributions for local structure characteristics such as the number of predecessors and the number of successors of a vertex in $G_{\boldsymbol{p}(n)}$ have received much attention in the literature for many special cases of the model $T_{\boldsymbol{p}(n)}$ (see [3]–[7], [12], [13], [16], and [18]). In Section 3 we show that the distributions and first moments of these and other random mapping statistics for the local structure of $G_n^{\hat{D}}$, the random digraph which represents $T_n^{\hat{D}}$, can be computed as expectations of functions of the (random) in-degree sequence $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. In Section 4 we apply these results to two special examples—the preferential and anti-preferential attachment models—which turn out to be equivalent to special cases of $T_n^{\hat{D}}$.

2. The model

In order to define the model $T_n^{\hat{D}}$, we adopt the following notation. For $n \ge 1$, suppose that $f \in \mathcal{M}_n$. Then, for $1 \le i \le n$, we let $d_i(f)$ denote the in-degree of vertex i in the digraph $G(f)$ which represents the mapping f, and define $\vec{d}(f) \equiv (d_1(f), \ldots, d_n(f))$. Also, given a vector $\vec{d} \equiv (d_1, d_2, \ldots, d_n)$ of nonnegative integers such that $\sum_{i=1}^{n} d_i = n$, define

$$\mathcal{M}_n(\vec{d}) \equiv \{f \in \mathcal{M}_n \colon \vec{d}(f) = \vec{d}\}$$

to be the set of all mappings $f \in \mathcal{M}_n$ with in-degree sequence $\vec{d}$.

To define $T_n^{\hat{D}}$, we start with a collection of nonnegative integer-valued, exchangeable random variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ such that $\hat{D}_1 + \hat{D}_2 + \cdots + \hat{D}_n = n$. Given the event $\{\hat{D}_i = d_i, i = 1, 2, \ldots, n\}$ (with $\Pr\{\hat{D}_i = d_i, i = 1, 2, \ldots, n\} > 0$), we define the conditional distribution of

$T_n^{\hat{D}}$ by

$$\Pr\{T_n^{\hat{D}} = f \mid \hat{D}_i = d_i,\ i = 1, 2, \ldots, n\} = \begin{cases} \dfrac{\prod_{i=1}^n d_i!}{n!} & \text{if } d_i(f) = d_i,\ i = 1, 2, \ldots, n, \\[3mm] 0 & \text{otherwise.} \end{cases}$$

(2.1)

In other words, *given* $(\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n) = (d_1, d_2, \ldots, d_n) = \vec{d}$, $T_n^{\hat{D}}$ is uniformly distributed over $\mathcal{M}_n(\vec{d})$. It follows from (2.1) that, for any $f \in \mathcal{M}_n$,

$$\Pr\{T_n^{\hat{D}} = f\} = \frac{\prod_{i=1}^n (d_i(f))!}{n!} \Pr\{\hat{D}_i = d_i(f),\ 1 \le i \le n\}.$$

Given the random mapping $T_n^{\hat{D}}$, let $G_n^{\hat{D}} \equiv G(T_n^{\hat{D}})$ denote the random digraph on n labelled vertices that represents $T_n^{\hat{D}}$. We note that it follows, from the exchangeability of the variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ and (2.1), that for any permutation $\sigma : [n] \to [n]$, we have

$$\sigma \circ T_n^{\hat{D}} \overset{\mathrm{d}}{\sim} T_n^{\hat{D}} \circ \sigma \overset{\mathrm{d}}{\sim} T_n^{\hat{D}}.$$

In other words, the distribution of the corresponding digraph $G_n^{\hat{D}}$ is invariant under relabelling of the vertices of the graph. An important class of examples can be constructed as follows. Suppose that $D_1, D_2, \ldots, D_n$ are independent and identically distributed (i.i.d.), nonnegative integer-valued random variables with $\Pr\{\sum_{i=1}^n D_i = n\} > 0$, and let $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ be a sequence of random variables with joint distribution given by

$$\Pr\{\hat{D}_i = d_i,\ 1 \le i \le n\} = \Pr\left\{D_i = d_i,\ 1 \le i \le n \ \middle| \ \sum_{i=1}^n D_i = n\right\}.$$

Clearly, the variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ are exchangeable with $\sum_{i=1}^n \hat{D}_i = n$, so we can use $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ to construct $T_n^{\hat{D}}$ and $G_n^{\hat{D}}$. We note that it is easy to check that if $D_1, D_2, \ldots, D_n$ are i.i.d. Poisson variables then $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ have a multinomial distribution with parameters n and $(1/n, 1/n, \ldots, 1/n)$, and the corresponding random mapping $T_n^{\hat{D}}$ is just the usual uniform random mapping. There are interesting interpretations of $T_n^{\hat{D}}$ in the cases where the underlying i.i.d. variables $D_1, D_2, \ldots, D_n$ have (i) a generalised negative binomial distribution, and (ii) a binomial $\text{Bin}(m, p)$ distribution. In particular, case (i) corresponds to a random mapping with 'preferential attachment' and case (ii) corresponds to a random mapping with 'anti-preferential attachment'.

In this paper we consider some local properties of the random digraph $G_n^{\hat{D}}$, which represents the random mapping $T_n^{\hat{D}}$. By local properties we mean the distributions of random mapping characteristics related to a given vertex v—for example, the numbers of predecessors and successors of v in $G_n^{\hat{D}}$. To investigate such variables, we introduce some further notation and definitions. For any $f \in \mathcal{M}_n$ and any positive integer k, let $f^{(k)}$ denote the kth iterate of f and, for every $i \in [n]$, define $f^{(0)}(i) \equiv i$. We say that $i \in [n]$ is a *cyclic* vertex of f if, for some $k > 0$, $f^{(k)}(i) = i$. In particular, if i is a cyclic vertex of f then vertex i lies on a cycle in the digraph $G(f)$. We also note that every component of $G(f)$ consists of a directed cycle with trees, directed towards the cycle, attached to it. For any $f \in \mathcal{M}_n$, let $\mathcal{L}(f)$ denote the set of cyclic vertices in the component of $G(f)$ which contains the vertex 1. Define $\ell(f) = |\mathcal{L}(f)|$, and define $h(f)$, the height of vertex 1 in $G(f)$, by

$$h(f) = \min\{k \ge 0 : f^{(k)}(1) \in \mathcal{L}(f)\}.$$

Next, let

$$\mathscr{P}(f) \equiv \{j \in [n]: f^{(k)}(j) = 1 \text{ for some } k \geq 0\}$$

denote the predecessors of vertex 1 under f, and let

$$\mathscr{S}(f) \equiv \{j \in [n]: f^{(k)}(1) = j \text{ for some } k \geq 0\}$$

denote the successors of vertex 1 under f. We define $p(f) = |\mathscr{P}(f)|$ and $s(f) = |\mathscr{S}(f)|$. In this paper we are interested in the local properties of $G_n^{\hat{D}}$, which are described by the random variables $\ell_n^{\hat{D}} \equiv \ell(T_n^{\hat{D}})$, $h_n^{\hat{D}} \equiv h(T_n^{\hat{D}})$, $p_n^{\hat{D}} \equiv p(T_n^{\hat{D}})$, and $s_n^{\hat{D}} \equiv s(T_n^{\hat{D}})$. We mention here that the distributions for the number of cyclic vertices, the number of components, and the size of a typical component of $G_n^{\hat{D}}$ have been determined in the companion paper [9].

3. Results

In this section we derive general formulae for the distributions and first moments of the variables $\ell_n^{\hat{D}}$, $h_n^{\hat{D}}$, $p_n^{\hat{D}}$, and $s_n^{\hat{D}}$ described above. The joint distribution of $h_n^{\hat{D}}$ and $\ell_n^{\hat{D}}$ is given by Theorem 3.1, below, and the distributions for $s_n^{\hat{D}}$, $h_n^{\hat{D}}$, and $\ell_n^{\hat{D}}$ are respectively obtained as Corollaries 3.1, 3.2, and 3.3, below.

Theorem 3.1. *For $0 \leq x, y \leq n - 1$,*

$$\Pr\{h_n^{\hat{D}} = x, \ell_n^{\hat{D}} = y + 1\} = \frac{1}{n} \operatorname{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_{x+y}] \mathbf{1}_{\{x \neq 0,\, x+y+1 \leq n\}}$$

$$+ \frac{1}{n} \operatorname{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_{y+1}] \mathbf{1}_{\{x=0\}}.$$

Proof. First suppose that $1 \leq x \leq n - 1$ and $0 \leq y \leq n - x - 1$. Let $k = x + y$, then we have

$$\Pr\{h_n^{\hat{D}} = x, \ell_n^{\hat{D}} = y + 1\} = \sum_{\substack{C \subseteq [n]\setminus\{1\} \\ |C| = k}} \Pr\{\mathscr{S}_n^{\hat{D}} = C \cup \{1\}, h_n^{\hat{D}} = x, \ell_n^{\hat{D}} = y + 1\}$$

$$= \binom{n-1}{k} \Pr\{\mathscr{S}_n^{\hat{D}} = C' \cup \{1\}, h_n^{\hat{D}} = x, \ell_n^{\hat{D}} = y + 1\}, \quad (3.1)$$

where $C' = \{2, 3, \ldots, k + 1\}$ and $\mathscr{S}_n^{\hat{D}} \equiv \mathscr{S}(T_n^{\hat{D}})$. We note that the second equality holds since the distribution of $G_n^{\hat{D}}$ is invariant under the relabelling of its vertices. Next, let $\Sigma(k)$ denote the set of all permutations, σ, of $\{2, 3, \ldots, k + 1\}$ and, for any $\sigma \in \Sigma(k)$, let

$$\mathscr{A}_\sigma = \{(T_n^{\hat{D}})^{(m)}(1) = \sigma(m + 1) \text{ for } 1 \leq m \leq k \text{ and } (T_n^{\hat{D}})^{(k+1)}(1) = \sigma(x + 1)\}.$$

Then we have

$$\Pr\{\mathscr{S}_n^{\hat{D}} = C' \cup \{1\}, h_n^{\hat{D}} = x, \ell_n^{\hat{D}} = y + 1\} = \sum_{\sigma \in \Sigma(k)} \Pr\{\mathscr{A}_\sigma\} = k! \Pr\{\mathscr{A}_{id}\}. \quad (3.2)$$

Again, the second equality follows from the invariance of the distribution of $G_n^{\hat{D}}$ under the relabelling of its vertices. To determine $\Pr\{\mathscr{A}_{id}\}$, we write

$$\Pr\{\mathscr{A}_{id}\} = \sum_{\substack{d_i \geq 0 \\ \sum d_i = n}} \Pr\{\mathscr{A}_{id} \mid \hat{D}_i = d_i,\, 1 \leq i \leq n\} \Pr\{\hat{D}_i = d_i,\, 1 \leq i \leq n\}. \quad (3.3)$$

We note that

$$\Pr\{\mathcal{A}_{id} \mid \hat{D}_i = d_i, \ 1 \le i \le n\} \ne 0$$

if and only if $d_i \ge 1$ for $2 \le i \le k+1$ and, in addition, $d_{x+1} \ge 2$. In this case it follows, from (2.1) and straightforward counting arguments, that

$$\Pr\{\mathcal{A}_{id} \mid \hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$= \frac{(n-k-1)!}{d_1! \, (d_2 - 1)! \cdots (d_{x+1} - 2)! \cdots (d_{k+1} - 1)! \, d_{k+2}! \cdots d_n!} \frac{d_1! \cdots d_n!}{n!}$$
$$= \frac{d_2 d_3 \cdots d_{x+1}(d_{x+1} - 1) \cdots d_{k+1}}{n(n-1) \cdots (n-k)}. \tag{3.4}$$

Observe that (3.4) is still valid when $d_i = 0$ for some $2 \le i \le k+1$ or $d_{x+1} = 1$. So it follows, from (3.3) and (3.4), that

$$\Pr\{\mathcal{A}_{id}\} = \frac{\mathrm{E}[\hat{D}_2 \cdots \hat{D}_{x+1}(\hat{D}_{x+1} - 1) \cdots \hat{D}_{k+1}]}{n(n-1) \cdots (n-k)} = \frac{\mathrm{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_k]}{n(n-1) \cdots (n-k)}. \tag{3.5}$$

The last equality holds by the exchangeability of $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. So, combining (3.1), (3.2), and (3.5), we obtain

$$\Pr\{h_n^{\hat{D}} = x, \ \ell_n^{\hat{D}} = y + 1\} = \frac{1}{n} \, \mathrm{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_{x+y}],$$

since $k = x + y$. Next, suppose that $x = 0$ and $1 \le y \le n - 1$. Then, as above, we have

$$\Pr\{h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = y + 1\} = \sum_{\substack{C \subseteq [n]\setminus\{1\} \\ |C| = y}} \Pr\{\mathcal{S}_n^{\hat{D}} = C \cup \{1\}, \ h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = y + 1\}$$
$$= \binom{n-1}{y} \Pr\{\mathcal{S}_n^{\hat{D}} = C' \cup \{1\}, \ h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = y + 1\}, \tag{3.6}$$

where $C' = \{2, 3, \ldots, y + 1\}$. Again, let $\Sigma(y)$ denote the set of all permutations, σ, of C' and, for any $\sigma \in \Sigma(y)$, let

$$\mathcal{B}_\sigma = \{(T_n^{\hat{D}})^{(m)}(1) = \sigma(m+1) \text{ for } 1 \le m \le y \text{ and } (T_n^{\hat{D}})^{(y+1)}(1) = 1\}.$$

Then, as above, we have

$$\Pr\{\mathcal{S}_n^{\hat{D}} = C' \cup \{1\}, \ h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = y + 1\} = \sum_{\sigma \in \Sigma(y)} \Pr\{\mathcal{B}_\sigma\} = k! \Pr\{\mathcal{B}_{id}\}. \tag{3.7}$$

We write

$$\Pr\{\mathcal{B}_{id}\} = \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} \Pr\{\mathcal{B}_{id} \mid \hat{D}_i = d_i, \ 1 \le i \le n\} \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\},$$

and we note that

$$\Pr\{\mathcal{B}_{id} \mid \hat{D}_i = d_i, \ 1 \le i \le n\} \ne 0$$

if and only if $d_i \geq 1$ for $1 \leq i \leq y + 1$. In this case it follows, from (2.1) and counting arguments, that

$$
\begin{aligned}
&\Pr\{\mathcal{B}_{id} \mid \hat{D}_i = d_i, \ 1 \leq i \leq n\} \\
&\quad = \frac{(n - y - 1)!}{(d_1 - 1)! \, (d_2 - 1)! \cdots (d_{y+1} - 1)! \, d_{y+2}! \cdots d_n!} \frac{d_1! \cdots d_n!}{n!} \\
&\quad = \frac{d_1 d_2 \cdots d_{y+1}}{n(n-1) \cdots (n-y)}.
\end{aligned}
\tag{3.8}
$$

Observe that (3.8) is still valid when $d_i = 0$ for some $1 \leq i \leq y + 1$. So, combining (3.6), (3.7), and (3.8), we obtain

$$
\Pr\{h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = y + 1\} = \frac{1}{n} \, \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_{y+1}].
\tag{3.9}
$$

Finally, for $x = y = 0$, we have

$$
\begin{aligned}
\Pr\{h_n^{\hat{D}} &= 0, \ \ell_n^{\hat{D}} = 1\} \\
&= \Pr\{s_n^{\hat{D}} = 1\} \\
&= \Pr\{T_n^{\hat{D}}(1) = 1\} \\
&= \sum_{\substack{d_i \geq 0 \\ \sum d_i = n}} \Pr\{T_n^{\hat{D}}(1) = 1 \mid \hat{D}_i = d_i, \ 1 \leq i \leq n\} \Pr\{\hat{D}_i = d_i, \ 1 \leq i \leq n\}.
\end{aligned}
\tag{3.10}
$$

We note that $\Pr\{T_n^{\hat{D}}(1) = 1 \mid \hat{D}_i = d_i, \ 1 \leq i \leq n\} \neq 0$ if and only if $d_1 \geq 1$, and in this case we have

$$
\Pr\{T_n^{\hat{D}}(1) = 1 \mid \hat{D}_i = d_i, \ 1 \leq i \leq n\} = \frac{(n-1)! \, d_1! \cdots d_n!}{(d_1 - 1)! \, (d_2)! \cdots d_n! \, n!} = \frac{d_1}{n}.
\tag{3.11}
$$

Equation (3.11) remains valid when $d_1 = 0$. So it follows, from (3.10) and (3.11), that

$$
\Pr\{h_n^{\hat{D}} = 0, \ \ell_n^{\hat{D}} = 1\} = \frac{\mathrm{E}[\hat{D}_1]}{n},
$$

and (3.9) holds in the case in which $y = 0$.

Corollary 3.1. *For $0 \leq k \leq n - 1$,*

$$
\Pr\{s_n^{\hat{D}} = k + 1\} = \frac{k}{n} \, \mathrm{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_k] + \frac{1}{n} \, \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_{k+1}].
$$

Proof. The corollary follows from Theorem 3.1 and the observation that, for $0 \leq k \leq n - 1$,

$$
\Pr\{s_n^{\hat{D}} = k + 1\} = \sum_{x=0}^{k} \Pr\{h_n^{\hat{D}} = x, \ \ell_n^{\hat{D}} = k - x + 1\}.
$$

We also immediately obtain the following results from Theorem 3.1.

Corollary 3.2. *For $1 \leq x \leq n - 1$,*

$$\Pr\{h_n^{\hat{D}} = x\} = \frac{1}{n} \sum_{y=0}^{n-x-1} \mathrm{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_{x+y}]$$

and

$$\Pr\{h_n^{\hat{D}} = 0\} = \frac{1}{n} \sum_{y=0}^{n-1} \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_{y+1}].$$

Corollary 3.3. *For $0 \leq y \leq n - 1$,*

$$\Pr\{\ell_n^{\hat{D}} = y + 1\} = \frac{1}{n} \sum_{x=1}^{n-y-1} \mathrm{E}[\hat{D}_1(\hat{D}_1 - 1)\hat{D}_2 \cdots \hat{D}_{x+y}] + \frac{1}{n} \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_{y+1}],$$

where the sum above is interpreted as 0 if $y = n - 1$.

Next, we consider the distribution of $p_n^{\hat{D}}$, the number of predecessors of vertex 1 in $G_n^{\hat{D}}$. In order to determine this distribution, we need to count the number of directed trees, rooted at 1, with a specified degree sequence. To state the required tree counting lemma, we adopt some notation. First, for any $k \geq 2$, let $\mathcal{T}_k$ denote the set of all labelled trees on the vertices $1, 2, \ldots, k$ such that each tree $\tau \in \mathcal{T}_k$ is rooted at vertex 1 and the edges of τ are oriented so that the (shortest) path from any vertex v to the root 1 is directed towards 1. For any $\tau \in \mathcal{T}_k$ and any vertex $v \in \tau$, let $d_v(\tau)$ denote the in-degree of v in τ. Note that we must have $d_1(\tau) \geq 1$ and $\sum_{i=1}^{k} d_i(\tau) = k - 1$. Finally, for any nonnegative integers $d_1, d_2, \ldots, d_k$ such that $d_1 \geq 1$ and $\sum_{i=1}^{k} d_i = k - 1$, let

$$\mathcal{T}_k(d_1, d_2, \ldots, d_k) = \{\tau \in \mathcal{T}_k : d_i(\tau) = d_i, \ 1 \leq i \leq k\},$$

and let $\tau_k(d_1, d_2, \ldots, d_k) = |\mathcal{T}_k(d_1, d_2, \ldots, d_k)|$. Then we have the following result.

Lemma 3.1. *Suppose that $k \geq 2$ and that $d_1, d_2, \ldots, d_k$ are nonnegative integers such that $d_1 \geq 1$ and $\sum_{i=1}^{k} d_i = k - 1$. Then*

$$\tau_k(d_1, d_2, \ldots, d_k) = \frac{(k - 2)!}{(d_1 - 1)! \, d_2! \cdots d_k!}.$$

Proof. Suppose that $k \geq 2$ and $d_1, d_2, \ldots, d_k$ are nonnegative integers such that $d_1 \geq 1$ and $\sum_{i=1}^{k} d_i = k - 1$, and let $\mathcal{S}(d_1, d_2, \ldots, d_k)$ denote the set of all sequences of length $k - 1$ such that, for any $s \in \mathcal{S}(d_1, d_2, \ldots, d_k)$ and for any $1 \leq i \leq k$, the integer i appears d_i times in s and the sequence s ends with the integer 1. To prove the equation in Lemma 3.1, we use a Prüfer tree code [19] to construct a bijection between $\mathcal{S}(d_1, d_2, \ldots, d_k)$ and $\mathcal{T}_k(d_1, d_2, \ldots, d_k)$. Specifically, the Prüfer encoding constructs a sequence $s \in \mathcal{S}(d_1, d_2, \ldots, d_k)$ for every $\tau \in \mathcal{T}_k(d_1, d_2, \ldots, d_k)$ as follows.

1. Suppose that τ is a labelled, directed tree rooted at 1. Suppose that v is the smallest vertex in τ such that the in-degree $d_v(\tau) = 0$, and suppose that $v \to w$ is a directed edge in τ. Delete vertex v from τ and add w to the sequence s.

2. If τ is the trivial tree consisting only of vertex 1, *stop*. Otherwise, go to step 1 and repeat.

The Prüfer decoding constructs a tree $\tau \in \mathcal{T}_k(d_1, d_2, \ldots, d_k)$ for every $s \in \mathcal{S}(d_1, d_2, \ldots, d_k)$ as follows.

1. Start with k labelled isolated vertices, and let $K = (1, 2, \ldots, k)$ denote the ordered list of numbers $1, 2, \ldots, k$.

2. Suppose that i is the smallest number in list K which does not appear in sequence s, and suppose that j is the first number in sequence s. Add the directed edge $i \rightarrow j$ to the graph and remove i from list K and j from sequence s.

3. If $K = (1)$, *stop*. Otherwise, go to step 2 and repeat.

The Prüfer encoding and decoding are inverse operations (see [19]), so

$$|\mathcal{T}_k(d_1, d_2, \ldots, d_k)| = |\mathcal{S}(d_1, d_2, \ldots, d_k)| = \frac{(k-2)!}{(d_1 - 1)! \, d_2! \cdots d_k!},$$

as desired.

Given Lemma 3.1, we can prove the following result.

Theorem 3.2. *For* $0 \le k \le n - 1$,

$$\Pr\{p_n^{\hat{D}} = k + 1\} = \frac{n - k}{n(k + 1)} \Pr\left\{\sum_{i=1}^{k+1} \hat{D}_i = k\right\} + \frac{1}{n} \Pr\left\{\sum_{i=1}^{k+1} \hat{D}_i = k + 1\right\}.$$

Proof. First, suppose that $1 \le k \le n - 1$. Then we have

$$\Pr\{p_n^{\hat{D}} = k + 1\} = \sum_{\substack{C \subseteq [n] \setminus \{1\} \\ |C| = k}} \Pr\{\mathcal{P}_n^{\hat{D}} = C \cup \{1\}\} = \binom{n-1}{k} \Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}\}, \quad (3.12)$$

where $C' = \{2, 3, \ldots, k + 1\}$ and $\mathcal{P}_n^{\hat{D}} \equiv \mathcal{P}(T_n^{\hat{D}})$. Next, observe that

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}\}$$
$$= \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} \Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\} \mid \hat{D}_i = d_i, \ 1 \le i \le n\} \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$= \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} \sum_{\ell=1}^{n} \Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, \ T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$\times \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}. \quad (3.13)$$

Now suppose that $\vec{d} = (d_1, d_2, \ldots, d_n)$ such that $\sum_{i=1}^{n} d_i = n$ and, for $1 \le \ell \le n$, let

$$\mathcal{M}_n(\vec{d}, C', \ell) = \{f \in \mathcal{M}_n : \vec{d}(f) = \vec{d}, \ \mathcal{P}(f) = C' \cup \{1\}, \ f(1) = \ell\}.$$

Then it follows from (2.1) that, for $1 \le \ell \le n$,

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, \ T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, \ 1 \le i \le n\} = \frac{|\mathcal{M}_n(\vec{d}, C', \ell)|}{n! \left(\prod_{i=1}^{n} d_i!\right)^{-1}}. \quad (3.14)$$

So, to determine $\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, 1 \leq i \leq n\}$, we need to count the set $\mathcal{M}_n(\vec{d}, C', \ell)$. There are three cases to consider.

First, suppose that $1 < \ell \leq k + 1$. If $\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}$ and $T_n^{\hat{D}}(1) = \ell$ then 1 is a cyclic vertex of $T_n^{\hat{D}}$ and the vertex set of the connected component in $G_n^{\hat{D}}$ which contains 1 is just $\{1, 2, \ldots, k + 1\}$. Hence, we have

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, 1 \leq i \leq n\} \neq 0$$

if and only if $\sum_{i=1}^{k+1} d_i = k + 1$, $d_\ell \geq 1$, and $d_1 \geq 1$. So, suppose that $\sum_{i=1}^{k+1} d_i = k + 1$, $d_\ell \geq 1$, and $d_1 \geq 1$, and let $f \in \mathcal{M}_n(\vec{d}, C', \ell)$. If the directed edge from 1 to ℓ in $G(f)$ is deleted, we obtain a directed tree on the vertices $1, 2, \ldots, k + 1$ with in-degree sequence $d_1, \ldots, d_\ell - 1, \ldots, d_{k+1}$ and root at vertex 1, and a random mapping graph on the vertices $k + 2, \ldots, n$ with in-degree sequence $d_{k+2}, \ldots, d_n$. It follows that to count $\mathcal{M}_n(\vec{d}, C', \ell)$ it is enough to count certain trees and certain random mapping graphs, i.e.

$$|\mathcal{M}_n(\vec{d}, C', \ell)| = \tau_{k+1}(d_1, \ldots, d_\ell - 1, \ldots, d_{k+1}) \frac{(n - k - 1)!}{d_{k+2}! \cdots d_n!}. \tag{3.15}$$

Substituting (3.15) into (3.14) and appealing to Lemma 3.1, in this case we obtain

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, 1 \leq i \leq n\}$$
$$= \frac{(k - 1)! \, (n - k - 1)!}{n!} d_1 d_\ell \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k+1\}}. \tag{3.16}$$

We note that (3.16) remains valid when $d_1 = 0$ or $d_\ell = 0$, and so it holds for *all* in-degree sequences $d_1, d_2, \ldots, d_n$ such that $\sum_{i=1}^n d_i = n$. Similarly, in the case in which $\ell = 1$, we have

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, T_n^{\hat{D}}(1) = 1 \mid \hat{D}_i = d_i, 1 \leq i \leq n\} \neq 0$$

if and only if $\sum_{i=1}^{k+1} d_i = k + 1$ and $d_1 \geq 2$. Now, provided that $\sum_{i=1}^{k+1} d_i = k + 1$ and $d_1 \geq 2$, we have, by the same argument as given above,

$$|\mathcal{M}_n(\vec{d}, C', 1)| = \tau_{k+1}(d_1 - 1, d_2, \ldots, d_n) \frac{(n - k - 1)!}{d_{k+2}! \cdots d_n!}. \tag{3.17}$$

Again, substituting (3.17) into (3.14) and appealing to Lemma 3.1, we obtain

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, T_n^{\hat{D}}(1) = 1 \mid \hat{D}_i = d_i, 1 \leq i \leq n\}$$
$$= \frac{(k - 1)! \, (n - k - 1)!}{n!} d_1 (d_1 - 1) \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k+1\}}. \tag{3.18}$$

Again, we note that (3.18) remains valid when $d_1 = 0$ or $d_1 = 1$, and so it holds for *all* in-degree sequences $d_1, d_2, \ldots, d_n$ such that $\sum_{i=1}^n d_i = n$.

Finally, suppose that $\ell > k + 1$. If $\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}$ and $T_n^{\hat{D}}(1) = \ell$ then 1 is not a cyclic vertex of $T_n^{\hat{D}}$ and the graph induced by $T_n^{\hat{D}}$ on the vertex set $C' \cup \{1\}$ is a directed tree, rooted at 1. Hence, we have

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C', T_n^{\hat{D}} = \ell \mid \hat{D}_i = d_i, 1 \leq i \leq n\} \neq 0$$

if and only if $\sum_{i=1}^{k+1} d_i = k$, $d_\ell \geq 1$, and $d_1 \geq 1$. So, suppose that $\sum_{i=1}^{k+1} d_i = k$, $d_\ell \geq 1$, $d_1 \geq 1$, and $f \in \mathcal{M}_n(\vec{d}, C', \ell)$. If the directed edge from 1 to ℓ in $G(f)$ is deleted, we obtain a

directed tree on the vertices $1, 2, \ldots, k+1$ with in-degree sequence $d_1, d_2, \ldots, d_{k+1}$ and root at vertex 1, and a random mapping graph on the vertices $k+2, \ldots, n$ with in-degree sequence $d_{k+2}, \ldots, d_\ell - 1, \ldots, d_n$. So, in this case we have

$$|\mathcal{M}_n(\vec{d}, C', \ell)| = \tau_{k+1}(d_1, d_2, \ldots, d_{k+1}) \frac{(n-k-1)!}{d_{k+2}! \cdots (d_\ell - 1)! \cdots d_n!}. \tag{3.19}$$

Substituting (3.19) into (3.14) and appealing to Lemma 3.1, we obtain

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, \ T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$= \frac{(k-1)! \, (n-k-1)!}{n!} d_1 d_\ell \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k\}}. \tag{3.20}$$

We note that (3.20) remains valid when $d_1 = 0$ or $d_\ell = 0$, and so it holds for *all* in-degree sequences $d_1, d_2, \ldots, d_n$ such that $\sum_{i=1}^n d_i = n$.

It follows, from (3.13), (3.16), (3.18), (3.20), and the identities

$$\left(d_1(d_1 - 1) + \sum_{\ell=2}^{k+1} d_1 d_\ell \right) \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k+1\}} = k d_1 \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k+1\}}$$

and

$$\sum_{\ell=k+2}^{n} d_1 d_\ell \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k\}} = (n-k) d_1 \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k\}},$$

that

$$\Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}\}$$
$$= \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} \sum_{\ell=1}^{k+1} \Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, \ T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$\times \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$+ \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} \sum_{\ell=k+2}^{n} \Pr\{\mathcal{P}_n^{\hat{D}} = C' \cup \{1\}, \ T_n^{\hat{D}}(1) = \ell \mid \hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$\times \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$= \frac{k! \, (n-k-1)!}{n!} \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} d_1 \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k+1\}} \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}$$
$$+ \frac{(k-1)! \, (n-k)!}{n!} \sum_{\substack{d_i \ge 0 \\ \sum d_i = n}} d_1 \, \mathbf{1}_{\{\sum_{i=1}^{k+1} d_i = k\}} \Pr\{\hat{D}_i = d_i, \ 1 \le i \le n\}$$

$$
= \frac{k!\,(n-k-1)!}{n!}\,\mathrm{E}\!\left[\hat{D}_1 \;\middle|\; \sum_{i=1}^{k+1}\hat{D}_i = k+1\right]\Pr\!\left\{\sum_{i=1}^{k+1}\hat{D}_i = k+1\right\}
$$

$$
+ \frac{(k-1)!\,(n-k)!}{n!}\,\mathrm{E}\!\left[\hat{D}_1 \;\middle|\; \sum_{i=1}^{k+1}\hat{D}_i = k\right]\Pr\!\left\{\sum_{i=1}^{k+1}\hat{D}_i = k\right\}
$$

$$
= \frac{k!\,(n-k-1)!}{n!}\Pr\!\left\{\sum_{i=1}^{k+1}\hat{D}_i = k+1\right\} + \frac{k!\,(n-k)!}{n!\,(k+1)}\Pr\!\left\{\sum_{i=1}^{k+1}\hat{D}_i = k\right\}. \tag{3.21}
$$

The last equality follows since, by the exchangeability of $\hat{D}_1, \ldots, \hat{D}_{k+1}$,

$$
\mathrm{E}\!\left[\hat{D}_1 \;\middle|\; \sum_{i=1}^{k+1}\hat{D}_i = k+1\right] = 1 \quad \text{and} \quad \mathrm{E}\!\left[\hat{D}_1 \;\middle|\; \sum_{i=1}^{k+1}\hat{D}_i = k\right] = \frac{k}{k+1}.
$$

So, in the case in which $1 \le k \le n-1$, the result follows from (3.12) and (3.21). Finally, for the case in which $k = 0$, we have

$$
\begin{aligned}
\Pr\{p_n^{\hat{D}} = 1\} &= \Pr\{\hat{D}_1 = 0\} + \Pr\{\hat{D}_1 = 1,\ T_n^{\hat{D}}(1) = 1\} \\
&= \Pr\{\hat{D}_1 = 0\} + \Pr\{T_n^{\hat{D}}(1) = 1 \mid \hat{D}_1 = 1\}\Pr\{\hat{D}_1 = 1\} \\
&= \Pr\{\hat{D}_1 = 0\} + \frac{1}{n}\Pr\{\hat{D}_1 = 1\},
\end{aligned}
$$

as required.

In Proposition 3.1 and Corollaries 3.4 and 3.5, below, we give formulae for the expected values of the random variables considered above.

Proposition 3.1. *We have*

$$
\mathrm{E}[p_n^{\hat{D}}] = \mathrm{E}[s_n^{\hat{D}}] = \sum_{k=1}^{n}\left(1 - \frac{k}{n}\right)\mathrm{E}[\hat{D}_1\hat{D}_2\cdots\hat{D}_k] + 1.
$$

Proof. Let $i, j \in [n]$, $i \ne j$, $1 \le k \le n-1$, and let $\mathcal{A}_{i,j}(k,n)$ be the event that 'there exists a path of length k from i to j in $G_n^{\hat{D}}$'. Next we note that

$$
\Pr\{\mathcal{A}_{1,2}(k,n)\} = \binom{n-2}{k-1}(k-1)!\,\Pr\{\tilde{\mathcal{A}}_{1,2}(k,n)\},
$$

where $\tilde{\mathcal{A}}_{1,2}(k,n)$ is the event that 'the path $1 \to 3 \to 4 \to \cdots \to k+1 \to 2$ exists in $G_n^{\hat{D}}$'. Using the same reasoning as in the proof of Theorem 3.1, we can check that, for all sequences $(d_1, d_2, \ldots, d_n)$ such that $d_1 + d_2 + \cdots + d_n = n$ and $d_i \ge 1$ for $i = 2, 3, \ldots, k+1$,

$$
\begin{aligned}
\Pr\{\tilde{\mathcal{A}}_{1,2}(k,n) &\mid \hat{D}_i = d_i,\ 1 \le i \le n\} \\
&= \frac{(n-k)!}{d_1!\,(d_2-1)!\cdots(d_{k+1}-1)!\,d_{k+2}!\cdots d_n!}\,\frac{d_1!\cdots d_n!}{n!} \\
&= \frac{(n-k)!}{n!}\,d_2 d_3 \cdots d_{k+1}. \tag{3.22}
\end{aligned}
$$

Since (3.22) holds even when $d_i = 0$ for some $i = 2, 3, \ldots, k+1$, we have

$$
\begin{aligned}
&\Pr\{\mathcal{A}_{i,j}(k, n)\} \\
&\quad = \Pr\{\mathcal{A}_{1,2}(k, n)\} \\
&\quad = \frac{(n-2)!}{(n-k-1)!} \sum_{\substack{d_i \geq 0 \\ \sum d_i = n}} \Pr\{\tilde{\mathcal{A}}_{1,2}(k, n) \mid \hat{D}_i = d_i,\, 1 \leq i \leq n\} \Pr\{\hat{D}_i = d_i,\, 1 \leq i \leq n\} \\
&\quad = \frac{n-k}{n(n-1)} \,\mathrm{E}[\hat{D}_2 \hat{D}_3 \cdots \hat{D}_{k+1}] \\
&\quad = \frac{1}{n-1}\left(1 - \frac{k}{n}\right) \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_k].
\end{aligned}
\tag{3.23}
$$

Let $I_{i,j}(n)$ be the indicator variable corresponding to the event $\mathcal{A}_{i,j}(n) = \bigcup_{k=1}^{n-1} \mathcal{A}_{i,j}(k, n)$ (i.e. the event that 'there exists a path from i to j in $G_n^{\hat{D}}$'). Note that

$$
\Pr\{I_{1,j}(n) = 1\} = \Pr\{\mathcal{A}_{1,j}(n)\} = \Pr\{\mathcal{A}_{j,1}(n)\} = \Pr\{I_{j,1}(n) = 1\},
$$

and, therefore,

$$
\mathrm{E}[p_n^{\hat{D}}] = \mathrm{E}\left[\sum_{j=2}^{n} I_{j,1}(n) + 1\right] = \mathrm{E}\left[\sum_{j=2}^{n} I_{1,j}(n) + 1\right] = \mathrm{E}[s_n^{\hat{D}}].
\tag{3.24}
$$

The result now follows from (3.23) and (3.24).

Corollary 3.4. *We have*

$$
\mathrm{E}[\ell_n^{\hat{D}}] = \frac{1}{2} + \frac{1}{2}\sum_{j=1}^{n} \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j].
$$

Proof. It follows, from Corollary 3.2 and Corollary 3.3, that

$$
\mathrm{E}[\ell_n^{\hat{D}}] - \mathrm{E}[h_n^{\hat{D}}] = \frac{1}{n}\sum_{j=1}^{n} j\,\mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j].
\tag{3.25}
$$

Conversely, the definition of $s_n^{\hat{D}}$ and Proposition 3.1 imply that

$$
\mathrm{E}[\ell_n^{\hat{D}}] + \mathrm{E}[h_n^{\hat{D}}] = \mathrm{E}[s_n^{\hat{D}}] = 1 + \sum_{j=1}^{n} \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j] - \sum_{j=1}^{n} \frac{j}{n}\,\mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j].
\tag{3.26}
$$

Equations (3.25) and (3.26) immediately give the formula for the expected value of $\ell_n^{\hat{D}}$.

Finally, Corollary 3.4 and (3.25) imply the formula for the expected value of $h_n^{\hat{D}}$.

Corollary 3.5. *We have*

$$
\mathrm{E}[h_n^{\hat{D}}] = \frac{1}{2} + \frac{1}{2}\sum_{j=1}^{n} \mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j] - \sum_{j=1}^{n} \frac{j}{n}\,\mathrm{E}[\hat{D}_1 \hat{D}_2 \cdots \hat{D}_j].
$$

4. Examples

In this section we consider the preferential and anti-preferential attachment models which are defined below. In a companion paper [9] it was shown that these models are equivalent to special cases of $T_n^{\hat{D}}$. Using this equivalence, we apply the results obtained in Section 3 to investigate the local properties of these models.

4.1. A preferential attachment model

We begin by defining $T_n^\rho : [n] \to [n]$, a random mapping with 'preferential attachment', where $\rho > 0$ is a fixed parameter. For $1 \le k \le n$, we define $T_n^\rho(k) = \xi_k^{(\rho,n)}$, where $\xi_1^{(\rho,n)}, \xi_2^{(\rho,n)}, \ldots, \xi_n^{(\rho,n)}$ is a sequence of random variables whose distributions depend on the evolution of an urn scheme. The distribution of each $\xi_k^{(\rho,n)}$ is determined by a (random) n-tuple of nonnegative weights $\vec{a}(k) = (a_1(k), a_2(k), \ldots, a_n(k))$, where, for $1 \le j \le n$, $a_j(k)$ is the 'weight' of the jth urn at the *start* of the kth round of the urn scheme. Specifically, given $\vec{a}(k) = \vec{a} = (a_1, \ldots, a_n)$, we define

$$\Pr\{\xi_k^{(\rho,n)} = j \mid \vec{a}(k) = \vec{a}\} = \frac{a_j}{\sum_{i=1}^n a_i}.$$

The random weight vectors $\vec{a}(1), \vec{a}(2), \ldots, \vec{a}(n)$ associated with the urn scheme are determined recursively. For $k = 1$, we set $a_1(1) = a_2(1) = \cdots = a_n(1) = \rho > 0$. For $k > 1$, $\vec{a}(k)$ depends on both $\vec{a}(k-1)$ and the value of $\xi_{k-1}^{(\rho,n)}$ as follows. Given that $\xi_{k-1}^{(\rho,n)} = j$, we set $a_j(k) = a_j(k-1) + 1$ and, for all other $i \ne j$, we set $a_i(k) = a_i(k-1)$ (i.e. if $\xi_{k-1}^{(\rho,n)} = j$ then a 'ball' with weight 1 is added to the jth urn).

The random mapping T_n^ρ as defined above is a preferential attachment model in the following sense. Since, for $1 \le k \le n$, we have $T_n^\rho(k) = \xi_k^{(\rho,n)}$, and since the (conditional) distribution of $\xi_k^{(\rho,n)}$ depends on the state of the urn scheme at the start of round k, it is clear that vertex k is more likely to be mapped to vertex j if the weight $a_j(k)$ is (relatively) large, i.e. if several of the vertices $1, 2, \ldots, k-1$ have already been mapped to vertex j. Also, the smaller the value of ρ, the stronger the preferential attachment effect. On the other hand, for values of ρ much larger than n, the preferential attachment effect is negligible and this model is essentially equivalent to the uniform random mapping model. The distribution of T_n^ρ is given by Theorem 4.1, below, and was obtained in [9].

Theorem 4.1. *Suppose that $D_1^\rho, D_2^\rho, \ldots$ are i.i.d. random variables with a generalised negative binomial distribution given by*

$$\Pr\{D_1^\rho = k\} = \frac{\Gamma(k+\rho)}{k!\,\Gamma(\rho)} \left(\frac{\rho}{1+\rho}\right)^\rho \left(\frac{1}{1+\rho}\right)^k \quad for\ k = 0, 1, \ldots,$$

where ρ is a positive parameter. For $n \ge 1$, let $\hat{D}(\rho, n) = (\hat{D}_{1,n}^\rho, \hat{D}_{2,n}^\rho, \ldots, \hat{D}_{n,n}^\rho)$ be a sequence of variables with joint distribution given by

$$\Pr\{\hat{D}_{i,n}^\rho = d_i,\ 1 \le i \le n\} = \Pr\left\{D_i^\rho = d_i,\ 1 \le i \le n \;\middle|\; \sum_{i=1}^n D_i^\rho = n\right\}.$$

Then, for every $n \ge 1$, the random mappings $T_n^\rho : [n] \to [n]$ and $T_n^{\hat{D}(\rho,n)} : [n] \to [n]$ have the same distribution.

Since $T_n^\rho \overset{\mathrm{d}}{\sim} T_n^{\hat{D}(\rho,n)}$, it follows that the random digraphs $G_n^\rho \equiv G(T_n^\rho)$ and $G_n^{\hat{D}(\rho,n)}$ have the same distribution. So, we can investigate the local properties of G_n^ρ by applying Theorem 3.1

(and its corollaries) and Theorem 3.2 to $G_n^{\hat{D}(\rho,n)}$. In the calculations that follow we adopt the following notation: if $g(s) = \sum_{k=0}^{\infty} a_k s^k$ then $[s^n]g(s) = a_n$, the coefficient of s^n in the power series expansion of $g(s)$. We also use the fact that the probability generating function for D_1^ρ is given by

$$\mathrm{E}[s^{D_1^\rho}] = \left(\frac{\rho}{1+\rho-s}\right)^\rho. \tag{4.1}$$

We begin by noting that, for any integer $1 \leq x \leq n$, we have

$$\mathrm{E}[\hat{D}_{1,n}^\rho \hat{D}_{2,n}^\rho \cdots \hat{D}_{x,n}^\rho] = \mathrm{E}\left[D_1^\rho D_2^\rho \cdots D_x^\rho \,\Bigg|\, \sum_{i=1}^n D_i^\rho = n \right]$$

$$= \frac{[s^n]\,\mathrm{E}[D_1^\rho s^{D_1^\rho} D_2^\rho s^{D_2^\rho} \cdots D_x^\rho s^{D_x^\rho} s^{D_{x+1}^\rho} \cdots s^{D_n^\rho}]}{[s^n]\,\mathrm{E}[s^{D_1^\rho} \cdots s^{D_n^\rho}]}$$

$$= \frac{[s^n](\mathrm{E}[D_1^\rho s^{D_1^\rho}])^x (\mathrm{E}[s^{D_1^\rho}])^{n-x}}{[s^n](\mathrm{E}[s^{D_1^\rho}])^n}. \tag{4.2}$$

The last equality holds since the variables $D_1^\rho, D_2^\rho, \ldots, D_n^\rho$ are i.i.d. Now, since

$$\mathrm{E}[D_1^\rho s^{D_1^\rho}] = s\left(\frac{\rho}{1+\rho-s}\right)^{\rho+1}, \tag{4.3}$$

it follows, from (4.1)–(4.3) and routine calculations, that

$$\mathrm{E}[\hat{D}_{1,n}^\rho \hat{D}_{2,n}^\rho \cdots \hat{D}_{x,n}^\rho] = \frac{\rho^x (n)_x}{(n\rho + x - 1)_x}, \tag{4.4}$$

where $(n)_x \equiv n(n-1)\cdots(n-x+1)$. Similarly, for $1 \leq x \leq n-1$, we have

$$\mathrm{E}[\hat{D}_{1,n}^\rho (\hat{D}_{1,n}^\rho - 1) \hat{D}_{2,n}^\rho \cdots \hat{D}_{x,n}^\rho]$$

$$= \mathrm{E}\left[D_1^\rho (D_1^\rho - 1) D_2^\rho \cdots D_x^\rho \,\Bigg|\, \sum_{i=1}^n D_i^\rho = n \right]$$

$$= \frac{[s^n]\,\mathrm{E}[D_1^\rho (D_1^\rho - 1) s^{D_1^\rho} D_2^\rho s^{D_2^\rho} \cdots D_x^\rho s^{D_x^\rho} s^{D_{x+1}^\rho} \cdots s^{D_n^\rho}]}{[s^n]\,\mathrm{E}[s^{D_1^\rho} \cdots s^{D_n^\rho}]}$$

$$= \frac{[s^n]\,\mathrm{E}[D_1^\rho (D_1^\rho - 1) s^{D_1^\rho}](\mathrm{E}[D_1^\rho s^{D_1^\rho}])^{x-1}(\mathrm{E}[s^{D_1^\rho}])^{n-x}}{[s^n](\mathrm{E}[s^{D_1^\rho}])^n}. \tag{4.5}$$

Since

$$\mathrm{E}[D_1^\rho (D_1^\rho - 1) s^{D_1^\rho}] = \left(\frac{1+\rho}{\rho}\right) s^2 \left(\frac{\rho}{1+\rho-s}\right)^{\rho+2},$$

it follows, from (4.5) and routine calculations, that

$$\mathrm{E}[\hat{D}_{1,n}^\rho (\hat{D}_{1,n}^\rho - 1) \hat{D}_{2,n}^\rho \cdots \hat{D}_{x,n}^\rho] = \rho^x (1+\rho) \frac{(n)_{x+1}}{(n\rho + x)_{x+1}}. \tag{4.6}$$

Proposition 4.1, below, now follows.

Proposition 4.1. *Suppose that $\rho > 0$. Then, for $1 \leq k \leq n - 1$,*

$$\Pr\{h(T_n^\rho) = k\} = \frac{1}{n} \sum_{x=k}^{n-1} \frac{\rho^x (1 + \rho)(n)_{x+1}}{(n\rho + x)_{x+1}}$$

and

$$\Pr\{h(T_n^\rho) = 0\} = \frac{1}{n} \sum_{x=1}^{n} \frac{\rho^x (n)_x}{(n\rho + x - 1)_x}.$$

Moreover,

$$\mathrm{E}[h(T_n^\rho)] = \frac{1}{2} + \sum_{j=1}^{n} \frac{n - 2j}{2n} \frac{\rho^j (n)_j}{(n\rho + j - 1)_j}.$$

Proof. It follows, from Corollary 3.2, Theorem 4.1, (4.4), and (4.6), that, for $\rho > 0$ and $1 \leq k \leq n - 1$,

$$\begin{aligned}
\Pr\{h(T_n^\rho) = k\} &= \Pr\{h_n^{\hat{D}(\rho,n)} = k\} \\
&= \frac{1}{n} \sum_{y=0}^{n-k-1} \mathrm{E}[\hat{D}_{1,n}^\rho (\hat{D}_{1,n}^\rho - 1)\hat{D}^\rho{}_{2,n} \cdots \hat{D}_{k+y,n}^\rho] \\
&= \frac{1}{n} \sum_{x=k}^{n-1} \frac{\rho^x (1 + \rho)(n)_{x+1}}{(n\rho + x)_{x+1}}.
\end{aligned}$$

Similarly,

$$\Pr\{h(T_n^\rho) = 0\} = \frac{1}{n} \sum_{x=1}^{n} \mathrm{E}[\hat{D}_{1,n}^\rho \hat{D}_{2,n}^\rho \cdots \hat{D}_{x,n}^\rho] = \sum_{x=1}^{n} \frac{\rho^x (n)_x}{(n\rho + x - 1)_x}.$$

Finally, the formula for the mean follows immediately from Corollary 3.5 and (4.4).

From Proposition 4.1, by the usual asymptotic calculations, we obtain the following result.

Corollary 4.1. *For fixed $\rho > 0$ and $k = \lfloor x\sqrt{n} \rfloor$ for some fixed $0 < x < \infty$,*

$$\Pr\{h(T_n^\rho) = k\} \sim \frac{1}{\sqrt{n}} \int_x^\infty \left(\frac{1 + \rho}{\rho} \right) \exp\left(\frac{-(1 + \rho)u^2}{2\rho} \right) du.$$

Moreover,

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{h(T_n^\rho)}{\sqrt{n}} \right] = \frac{1}{2} \sqrt{\frac{\pi}{2} \frac{\rho}{\rho + 1}}.$$

As in the proof of Proposition 4.1, Proposition 4.2, below, follows from Corollary 3.3, Corollary 3.4, Theorem 4.1, (4.4), and (4.6).

Proposition 4.2. *For $\rho > 0$ and for $0 \leq k \leq n - 1$,*

$$\Pr\{\ell(T_n^\rho) = k + 1\} = \frac{1}{n} \sum_{y=k+1}^{n-1} \frac{\rho^y (1 + \rho)(n)_{y+1}}{(n\rho + y)_{y+1}} + \frac{1}{n} \frac{\rho^{k+1}(n)_{k+1}}{(n\rho + k)_{k+1}}.$$

Moreover,

$$\mathrm{E}[\ell(T_n^\rho)] = \frac{1}{2} + \frac{1}{2} \sum_{j=1}^n \frac{\rho^j(n)_j}{(n\rho + j - 1)_j}.$$

From Proposition 4.2 we obtain the following result.

Corollary 4.2. *For fixed $\rho > 0$ and $k = \lfloor y\sqrt{n} \rfloor$ for some fixed $0 < y < \infty$,*

$$\Pr\{\ell(T_n^\rho) = k + 1\} \sim \frac{1}{\sqrt{n}} \int_y^\infty \left(\frac{1+\rho}{\rho} \right) \exp\left(\frac{-(1+\rho)u^2}{2\rho} \right) \mathrm{d}u.$$

Moreover,

$$\lim_{n\to\infty} \mathrm{E}\left[\frac{\ell(T_n^\rho)}{\sqrt{n}} \right] = \frac{1}{2}\sqrt{\frac{\pi}{2} \frac{\rho}{\rho+1}}.$$

We note that it follows, from Corollaries 4.1 and 4.2, that the variables $h(T_n^\rho)/\sqrt{n}$ and $\ell(T_n^\rho)/\sqrt{n}$ converge in distribution to a variable Y_α with density

$$f_\alpha(y) = \int_y^\infty \alpha \, \exp\left(-\frac{\alpha u^2}{2} \right) \mathrm{d}u \quad \text{for } y \in (0, \infty)$$

and parameter $\alpha = (\rho + 1)/\rho$. The density of Y_α can be rewritten as

$$f_\alpha(y) = \sqrt{2\pi\alpha}(1 - \Phi(\sqrt{\alpha}y)),$$

where $\Phi(\cdot)$ is the standard normal distribution function and $\mathrm{E}[Y_\alpha] = \frac{1}{2}\sqrt{\pi/2\alpha}$. So, the first moments of $h(T_n^\rho)/\sqrt{n}$ and $\ell(T_n^\rho)/\sqrt{n}$ converge to the first moment of Y_α with $\alpha = (\rho + 1)/\rho$. We also note that f_1 is the density for the limiting distributions of $h(T_n)/\sqrt{n}$ and $\ell(T_n)/\sqrt{n}$, where T_n is the uniform random mapping on $[n]$. So, for 'large' ρ (and α close to 1), we obtain asymptotic distributions which are close to the corresponding distributions obtained for T_n. Clearly, the smaller ρ, the smaller the mean length (height), but as in the uniform case, the mean is still of the order $\sqrt{n}$ when ρ is fixed as n tends to ∞. In a companion paper [10] we consider the asymptotic structure of $G_n^{D(\rho,n)}$ when $\rho = \rho(n) \to 0$. Finally, it is not very surprising that the random variables $h(T_n^\rho)$ and $\ell(T_n^\rho)$ behave asymptotically in the same way. In particular, it follows, from Theorems 3.1 and 4.1, that given $s(T_n^\rho) = m$, where $1 < m \le n$, then, for all $1 \le k, j \le m - 1$, $\Pr\{h(T_n^\rho) = k \mid s(T_n^\rho) = m\} = \Pr\{\ell(T_n^\rho) = j \mid s(T_n^\rho) = m\}$ (i.e. the conditional distributions of $h(T_n^\rho)$ and $\ell(T_n^\rho)$ are equal and uniform, except for one value).

Next, we obtain Proposition 4.3, below, from Corollary 3.1, Proposition 3.1, Theorem 4.1, (4.4), and (4.6).

Proposition 4.3. *For $\rho > 0$ and $0 \le k \le n - 1$,*

$$\Pr\{s(T_n^\rho) = k + 1\} = (k(1+\rho) + \rho)\frac{(n-k)\rho^k(n)_k}{n(n\rho + k)_{k+1}}$$

$$= \frac{\rho^k(n-1)_k}{(n\rho + k - 1)_k} - \frac{\rho^{k+1}(n-1)_{k+1}}{(n\rho + k)_{k+1}}.$$

Moreover,

$$\mathrm{E}[s(T_n^\rho)] = \sum_{j=1}^{n-1} \frac{\rho^j(n-1)_j}{(n\rho + j - 1)_j} + 1.$$

From Proposition 4.3 we obtain the following result.

Corollary 4.3. *For fixed $\rho > 0$ and $k = \lfloor x\sqrt{n} \rfloor$ for some fixed $0 < x < \infty$,*

$$\Pr\{s(T_n^\rho) = k + 1\} \sim \frac{1}{\sqrt{n}} \frac{(\rho + 1)x}{\rho} \exp\left(-\left(\frac{\rho + 1}{\rho}\right)\frac{x^2}{2}\right).$$

Moreover,

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{s(T_n^\rho)}{\sqrt{n}}\right] = \sqrt{\frac{\pi}{2} \frac{\rho}{\rho + 1}}.$$

The distribution obtained in Corollary 4.3 is the well-known Rayleigh distribution with parameter $\sigma = \sqrt{\rho/(\rho + 1)}$ and mean equal to $\sigma\sqrt{\pi/2}$.

Finally, for $0 \le k \le n - 1$, it follows, from Theorem 3.2 and Theorem 4.1, that

$$\Pr\{p(T_n^\rho) = k + 1\} = \frac{n - k}{n(k + 1)} \Pr\left\{\sum_{i=1}^{k+1} \hat{D}_{i,n}^\rho = k\right\} + \frac{1}{n} \Pr\left\{\sum_{i=1}^{k+1} \hat{D}_{i,n}^\rho = k + 1\right\}$$

$$= \frac{n - k}{n(k + 1)} \Pr\left\{\sum_{i=1}^{k+1} D_i^\rho = k \,\Big|\, \sum_{i=1}^{n} D_i^\rho = n\right\}$$

$$+ \frac{1}{n} \Pr\left\{\sum_{i=1}^{k+1} D_i^\rho = k + 1 \,\Big|\, \sum_{i=1}^{n} D_i^\rho = n\right\}.$$

Using routine generating function calculations to compute the conditional probabilities above and applying Proposition 3.1, we obtain the following result.

Proposition 4.4. *Suppose that $\rho > 0$. Then, for $0 \le k \le n - 2$,*

$$\Pr\{p(T_n^\rho) = k + 1\} = \frac{\rho n + n - 1}{n} \binom{n}{k + 1} \frac{\Gamma(\rho(k + 1) + k)\Gamma((\rho + 1)(n - k - 1))\Gamma(\rho n)}{\Gamma(\rho(k + 1))\Gamma(\rho(n - k - 1))\Gamma(\rho n + n)}$$

$$= \frac{1}{n}\binom{n - 1}{k} \frac{\rho}{\rho + 1} \frac{((\rho + 1)(k + 1) - 1)_k((\rho + 1)(n - k - 1))_{n-k-1}}{((\rho + 1)(k + 1) - 1)((\rho + 1)n - 2)_{n-2}},$$

while

$$\Pr\{p(T_n^\rho) = n\} = \frac{1}{n}.$$

Moreover,

$$\mathrm{E}[p(T_n^\rho)] = \sum_{j=1}^{n-1} \frac{\rho^j(n - 1)_j}{(n\rho + j - 1)_j} + 1.$$

We note that the distribution above is strongly related to the *quasi-hypergeometric distribution I* (see [14, Equation (2.122)]). We also obtain from Proposition 4.4, by the usual asymptotic calculations, the following result.

Corollary 4.4. *For fixed $\rho > 0$ and $k = 0, 1, \ldots,$*

$$\lim_{n \to \infty} \Pr\{p(T_n^\rho) = k + 1\}$$

$$= \frac{1}{\rho^k} \binom{(\rho + 1)(k + 1)}{k} \frac{1}{k + 1} \left(\frac{(\rho + 1)(k + 1) - k}{(\rho + 1)(k + 1) - 1}\right) \left(\frac{\rho}{\rho + 1}\right)^{(\rho + 1)(k + 1)}$$

$$= \frac{\rho}{(\rho + 1)k + \rho} \binom{(\rho + 1)k + \rho}{k} \left(\frac{1}{\rho + 1}\right)^k \left(\frac{\rho}{\rho + 1}\right)^{(\rho + 1)k + \rho - k}.$$

Moreover,

$$\mathrm{E}[p(T_n^\rho)] \sim \sqrt{\frac{\pi}{2} \frac{\rho}{\rho + 1}} n.$$

We note that in this case the limiting distribution given above is related to the *generalised binomial distribution* (see [14, Equation (2.121)]) and has infinite mean.

4.2. An anti-preferential attachment model

In this subsection we consider

$$T_n^m : [n] \to [n],$$

a random mapping with 'anti-preferential attachment', where $m \geq 1$ is a fixed integer parameter. For $1 \leq k \leq n$, we define $T_n^m(k) = \eta_k^{(m,n)}$, where, as in the definition of T_n^ρ, the variables $\eta_1^{(m,n)}, \eta_2^{(m,n)}, \ldots, \eta_n^{(m,n)}$ depend on the evolution of an urn scheme. The distribution of each variable $\eta_k^{(m,n)}$ is determined by a (random) n-tuple of nonnegative weights $\vec{b}(k) = (b_1(k), b_2(k), \ldots, b_n(k))$, where, for $1 \leq j \leq n$, $b_j(k)$ is the number of balls in the jth urn at the start of the kth round of the urn scheme. Specifically, given $\vec{b}(k) = \vec{b} = (b_1, \ldots, b_n)$, we define

$$\Pr\{\eta_k^{(m,n)} = j \mid \vec{b}(k) = \vec{b}\} = \frac{b_j}{\sum_{i=1}^n b_i}.$$

The random weight vectors $\vec{b}(1), \vec{b}(2), \ldots, \vec{b}(n)$ associated with the urn scheme are determined recursively. For $k = 1$, we set $b_1(1) = b_2(1) = \cdots = b_n(1) = m$. For $k > 1$, $\vec{b}(k)$ depends on both $\vec{b}(k - 1)$ and the value of $\eta_{k-1}^{(m,n)}$ as follows. Given that $\eta_{k-1}^{(m,n)} = j$, we set $b_j(k) = b_j(k - 1) - 1$ and, for all other $i \neq j$, we set $b_i(k) = b_i(k - 1)$ (i.e. if $\eta_{k-1}^{(m,n)} = j$ then a ball is removed from the jth urn).

The random mapping T_n^m as defined above is an anti-preferential attachment model in the following sense. Since, for $1 \leq k \leq n$, we have $T_n^m(k) = \eta_k^{(m,n)}$ and since the (conditional) distribution of $\eta_k^{(m,n)}$ depends on the state of the urn scheme at the start of round k, it is clear that vertex k is less likely to 'choose' vertex j if the weight $b_j(k)$ is (relatively) small, i.e. if several of the vertices $1, 2, \ldots, k - 1$ have already been mapped to vertex j. More precisely, the smaller the value of m, the stronger the anti-preferential effect. For values of m much larger than n, the anti-preferential effect is negligible and this model is essentially equivalent to the uniform random mapping model. It is also clear from the definition of T_n^m that the in-degree of any vertex in the random digraph $G_n^m \equiv G(T_n^m)$ is at most m and, for the case in which $m = 1$, T_n^1 is a (uniform) random permutation. The distribution of T_n^m is given by Theorem 4.2, below, and was obtained in [9].

Theorem 4.2. *Suppose that $D_1^m, D_2^m, \ldots$ are i.i.d. $\mathrm{Bin}(m, p)$ variables, where m is a positive integer parameter. Let $\hat{D}(m, n) = (\hat{D}_{1,n}^m, \hat{D}_{2,n}^m, \ldots, \hat{D}_{n,n}^m)$ be a sequence of variables with joint*

distribution given by

$$\Pr\{\hat{D}_{i,n}^m = d_i,\ 1 \le i \le n\} = \Pr\left\{D_i^m = d_i,\ 1 \le i \le n \ \bigg|\ \sum_{i=1}^n D_i^m = n\right\}.$$

Then the random mappings T_n^m and $T_n^{\hat{D}(m,n)}$ have the same distribution.

Since $T_n^m \overset{d}{\sim} T_n^{\hat{D}(m,n)}$, it follows that the random digraphs $G_n^m \equiv G(T_n^m)$ and $G_n^{\hat{D}(n,m)}$ have the same distribution. So, as in the case of the preferential attachment model, we can investigate the local properties of G_n^m by considering the local properties of $G_n^{\hat{D}(n,m)}$. We begin by noting that, for any integer $1 \le x \le n - 1$, we have

$$\mathrm{E}[\hat{D}_{1,n}^m(\hat{D}_{1,n}^m - 1)\hat{D}_{2,n}^m \cdots \hat{D}_{x,n}^m]$$

$$= \mathrm{E}\left[D_{1,n}^m(D_{1,n}^m - 1)D_{2,n}^m \cdots D_{x,n}^m \ \bigg|\ \sum_{i=1}^n D_{i,n}^m = n\right]$$

$$= \sum_{\sum_{i=1}^n d_i = n} (d_1 - 1)d_1 d_2 \cdots d_x \frac{\binom{m}{d_1} \cdots \binom{m}{d_n}}{\binom{nm}{n}}$$

$$= \sum_{t=x+1}^{\min(n,xm)} \sum_{\substack{\sum_{i=1}^x d_i = t \\ \text{and } \sum_{i=1}^n d_i = n}} (d_1 - 1)d_1 d_2 \cdots d_x \frac{\binom{m}{d_1} \cdots \binom{m}{d_n}}{\binom{nm}{n}}$$

$$= \sum_{t=x+1}^{\min(n,xm)} \sum_{\sum_{i=1}^x d_i = t} (d_1 - 1)d_1 d_2 \cdots d_x \frac{\binom{m}{d_1} \cdots \binom{m}{d_x}\binom{nm-xm}{n-t}}{\binom{nm}{n}}$$

$$= m^x(m-1) \sum_{t=x+1}^{\min(n,xm)} \sum_{\sum_{i=1}^x d_i = t} \frac{\binom{m-2}{d_1-2}\binom{m-1}{d_2-1} \cdots \binom{m-1}{d_x-1}\binom{nm-xm}{n-t}}{\binom{nm}{n}}$$

$$= m^x(m-1) \sum_{t=x+1}^{\min(n,xm)} \frac{\binom{xm-x-1}{t-x-1}\binom{nm-xm}{n-t}}{\binom{nm}{n}}$$

$$= m^x(m-1) \frac{\binom{nm-x-1}{n-x-1}}{\binom{nm}{n}}$$

$$= m^x(m-1) \frac{(n)_{x+1}}{(nm)_{x+1}}. \tag{4.7}$$

In the summations above the sum is always taken over those degree sequences $\vec{d}$ for which the binomial coefficients are defined. We also adopt the formal convention that $\binom{0}{0} = 1$. By similar calculations we also obtain, for $1 \le x \le n$ (and $m \ge 2$),

$$\mathrm{E}[\hat{D}_{1,n}^m \hat{D}_{2,n}^m \cdots \hat{D}_{x,n}^m] = m^x \frac{\binom{nm-x}{n-x}}{\binom{nm}{n}} = m^x \frac{(n)_x}{(nm)_x}. \tag{4.8}$$

As in the case of the preferential attachment model, we obtain Proposition 4.5, below, from Corollary 3.2, Corollary 3.5, Theorem 4.2, (4.7), and (4.8).

Proposition 4.5. *For $m \geq 2$ and $1 \leq k \leq n-1$,*

$$\Pr\{h(T_n^m) = k\} = \frac{1}{n} \sum_{t=k}^{n-1} \frac{m^t (m-1)(n)_{t+1}}{(nm)_{t+1}} \quad and \quad \Pr\{h(T_n^m) = 0\} = \frac{1}{n} \sum_{t=1}^{n} m^t \frac{(n)_t}{(nm)_t}.$$

Moreover,

$$\mathrm{E}[h(T_n^m)] = \frac{1}{2} + \sum_{j=1}^{n} \frac{n - 2j}{2n} \frac{m^j (n)_j}{(nm)_j}.$$

We obtain Corollary 4.5, below, from Proposition 4.5, by routine asymptotic calculations.

Corollary 4.5. *For fixed $m \geq 2$ and $k = \lfloor x\sqrt{n} \rfloor$ for some fixed $0 < x < \infty$,*

$$\Pr\{h(T_n^m) = k\} \sim \frac{1}{\sqrt{n}} \int_x^{\infty} \left(\frac{m-1}{m}\right) \exp\left(-\left(\frac{m-1}{m}\right) \frac{u^2}{2}\right) du.$$

Moreover,

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{h(T_n^m)}{\sqrt{n}}\right] = \frac{1}{2} \sqrt{\frac{\pi}{2} \frac{m}{m-1}}.$$

Likewise, we obtain Proposition 4.6 and Corollary 4.6, below, from Corollary 3.3, Corollary 3.4, Theorem 4.2, (4.7), and (4.8).

Proposition 4.6. *For $m \geq 2$ and $0 \leq k \leq n-1$,*

$$\Pr\{\ell(T_n^m) = k+1\} = \frac{1}{n} \sum_{t=k+1}^{n-1} \frac{m^t (m-1)(n)_{t+1}}{(nm)_{t+1}} + \frac{1}{n} m^{k+1} \frac{(n)_{k+1}}{(nm)_{k+1}}.$$

Moreover,

$$\mathrm{E}[\ell(T_n^m)] = \frac{1}{2} + \frac{1}{2} \sum_{j=1}^{n} \frac{m^j (n)_j}{(nm)_j}.$$

Corollary 4.6. *For fixed $m \geq 2$ and $k = \lfloor y\sqrt{n} \rfloor$ for some fixed $0 < y < \infty$,*

$$\Pr\{\ell(T_n^m) = k+1\} \sim \frac{1}{\sqrt{n}} \int_y^{\infty} \left(\frac{m-1}{m}\right) \exp\left(-\left(\frac{m-1}{m}\right) \frac{u^2}{2}\right) du.$$

Moreover,

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{\ell(T_n^m)}{\sqrt{n}}\right] = \frac{1}{2} \sqrt{\frac{\pi}{2} \frac{m}{m-1}}.$$

It follows, from Corollary 4.5 and Corollary 4.6, that, as in the case of the preferential model, the variables $h(T_n^m)/\sqrt{n}$ and $\ell(T_n^m)/\sqrt{n}$ converge in distribution to Y_α, but with $\alpha = (m-1)/m$ in this case, and their expected values converge to $\mathrm{E}[Y_\alpha]$. Again, for 'large' m, we obtain limiting distributions which are close to those obtained for uniform random mappings. The smaller m, the bigger the mean length (height), but since we consider only integers $m \geq 2$, this model is not radically different from the uniform random mapping model.

From Corollary 3.1, Proposition 3.1, Theorem 4.2, (4.7), and (4.8) we obtain the following result.

Proposition 4.7. *For $m \geq 2$ and $0 \leq k \leq n - 1$,*

$$\Pr\{s(T_n^m) = k + 1\} = \frac{m^k}{n}(k(m-1) + m)\frac{(n)_{k+1}}{(nm)_{k+1}}$$

$$= m^k \frac{(n-1)_k}{(nm)_k} - m^{k+1}\frac{(n-1)_{k+1}}{(nm)_{k+1}}.$$

Moreover,

$$\mathrm{E}[s(T_n^m)] = \sum_{j=1}^{n-1} \frac{m^j(n-1)_j}{(nm)_j} + 1.$$

Proposition 4.7 immediately implies the following result.

Corollary 4.7. *For fixed $m \geq 2$ and $k = \lfloor x\sqrt{n}\rfloor$ for some fixed $0 < x < \infty$,*

$$\Pr\{s(T_n^m) = k + 1\} \sim \frac{1}{\sqrt{n}}\frac{(m-1)x}{m}\exp\left(-\left(\frac{m-1}{m}\right)\frac{x^2}{2}\right).$$

Moreover,

$$\lim_{n \to \infty} \mathrm{E}\left[\frac{s(T_n^m)}{\sqrt{n}}\right] = \sqrt{\frac{\pi}{2}\frac{m}{m-1}}.$$

As in the preferential case, the distribution obtained in Corollary 4.7 is the Rayleigh distribution, now with parameter $\sigma = \sqrt{m/(m-1)}$.

Finally, we have the following result.

Proposition 4.8. *For $m \geq 2$ and $0 \leq k \leq n - 1$,*

$$\Pr\{p(T_n^m) = k + 1\} = \binom{n-1}{k}\frac{m}{m(k+1)}\frac{(m(k+1))_k(m(n-k-1))_{n-k-1}}{(mn)_{n-1}}.$$

Moreover,

$$\mathrm{E}[p(T_n^m)] = \sum_{j=1}^{n-1} \frac{m^j(n-1)_j}{(nm)_j} + 1.$$

Proof. It follows, from Theorem 3.2 and Theorem 4.2, that, for $m \geq 2$ and $0 \leq k \leq n - 1$,

$$\Pr\{p(T_n^m) = k + 1\}$$

$$= \frac{n-k}{n(k+1)}\Pr\left\{\sum_{i=1}^{k+1}\hat{D}_{i,n}^m = k\right\} + \frac{1}{n}\Pr\left\{\sum_{i=1}^{k+1}\hat{D}_{i,n}^m = k+1\right\}$$

$$= \frac{n-k}{n(k+1)}\Pr\left\{\sum_{i=1}^{k+1}D_i^m = k \;\Big|\; \sum_{i=1}^n D_i^m = n\right\} + \frac{1}{n}\Pr\left\{\sum_{i=1}^{k+1}D_i^m = k+1 \;\Big|\; \sum_{i=1}^n D_i^m = n\right\}$$

$$= \frac{n-k}{n(k+1)}\frac{\binom{m(k+1)}{k}\binom{m(n-k-1)}{n-k}}{\binom{mn}{n}} + \frac{1}{n}\frac{\binom{m(k+1)}{k+1}\binom{m(n-k-1)}{n-k-1}}{\binom{mn}{n}}$$

$$= \frac{(mn-n+1)\binom{m(k+1)}{k+1}\binom{m(n-k-1)}{n-k-1}}{n(mk+m-k)\binom{mn}{n}}.$$

The formula for the mean follows directly from Proposition 3.1 and (4.8).

We note that the above distribution is also related to the *quasi-hypergeometric distribution I* (see [14, Equation (2.122)]). Straightforward asymptotic calculations establish the following result.

Corollary 4.8. *For $m \geq 2$ and $k \in \{0, 1, 2, \ldots\}$,*

$$\lim_{n \to \infty} \Pr\{p(T_n^m) = k+1\} = \frac{1}{(m-1)^k} \binom{mk+m}{k} \frac{1}{k+1} \left(1 - \frac{1}{m}\right)^{mk+m}.$$

Moreover,

$$\mathrm{E}[p(T_n^m)] \sim \sqrt{\frac{\pi}{2} \frac{m}{m-1}} \, n.$$

We note that the asymptotic distribution for $p(T_n^m)$ is the Consul distribution (see [14, p. 98]) with parameters m and $\theta = 1/m$, and infinite mean.

5. Final remarks

One of the main advantages of the random mapping model $T_n^{\hat{D}}$ is that we have a calculus for this model which allows us to determine the distributions of several variables associated with the structure of $G_n^{\hat{D}}$ in terms of expectations of simple functions of $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$. As we have seen above, in the special case where the variables $\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_n$ have the same distribution as a collection of i.i.d. variables $D_1, D_2, \ldots, D_n$ *conditioned* on $\sum_{i=1}^n D_i = n$, it is straightforward to use this calculus to obtain exact and asymptotic distributions for various important random variables associated with the local structure of $G_n^{\hat{D}}$. The calculus for $T_n^{\hat{D}}$ also illustrates the fundamental importance of the distribution of the underlying degree sequence $\hat{D}_1, \ldots, \hat{D}_n$ to the structure of the random mapping digraph. This suggests that in various modelling applications the key to fitting a random mapping model is to fit the joint distribution of the vertex in-degree data. As an example, we mention the work of [1] on fitting random mappings with constraints on coalescence to shift register data. In the companion paper [9] we discuss how $T_n^{\hat{D}}$ provides an easier and more natural model for such data.

In this paper we have also considered the local structure of two special examples, T_n^{ρ} and T_n^m, for the case where the parameters ρ and m, respectively, are fixed. In another companion paper [10] we investigate the structure T_n^{ρ} when ρ is a function of n, and show that the asymptotic structure of T_n^{ρ} depends on whether $\rho n \to \infty$ or $\rho n \to \beta > 0$, or $\rho n \to 0$ as $n \to \infty$. In the model T_n^m the parameter must be a positive integer. Most of our results are given for $m \geq 2$ since the case $m = 1$ corresponds to a uniform random permutation. For the model T_n^m, the case in which $m = 2$ corresponds to the greatest anti-preferential effect. It would be interesting to consider an extension, $\tilde{T}_n^m$, where the parameter m is *any* real number greater than 1 and is such that $\tilde{T}_n^m = T_n^m$ when m is an integer. In particular, we would like to study how the structure of $\tilde{T}_n^m$ 'evolves' from a random mapping to a (uniform) random permutation as m decreases to 1.

Acknowledgement

References

[1] ARNEY, J. AND BENDER, E. A. (1982). Random mappings with constraints on coalescence and number of origins. *Pacific J. Math.* **103,** 269–294.

[2] ARRATIA, R., STARK, D. AND TAVARÉ, S. (1995). Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies. *Ann. Prob.* **23,** 1347–1388.

[3] BERG, S. (1981). On snowball sampling, random mappings and related problems. *J. Appl. Prob.* **18,** 283–290.

[4] BERG, S. (1983). Random contact processes, snowball sampling and factorial series distributions. *J. Appl. Prob.* **20,** 31–46.

[5] BERG, S. AND JAWORSKI, J. (1992). Probability distributions related to the local structure of a random mapping. In *Random Graphs*, eds A. Frieze and T. Łuczak, John Wiley, New York, pp. 1–21.

[6] BERG, S. AND MUTAFCHIEV, L. (1990). Random mappings with an attracting center: Lagrangian distributions and a regression function. *J. Appl. Prob.* **27,** 622–636.

[7] BURTIN, Y. D. (1980). On a simple formula for random mappings and its applications. *J. Appl. Prob.* **17,** 403–414.

[8] FLAJOLET, P. AND ODLYZKO, A. M. (1990). Random mapping statistics. In *Advances in Cryptology— EUROCRYPT'89* (Lecture Notes Comput. Sci. **434**), Springer, Berlin, pp. 329–354.

[9] HANSEN, J. C. AND JAWORSKI, J. (2008). Random mappings with exchangeable in-degrees. To appear in *Random Structures Algorithms.*

[10] HANSEN, J. C. AND JAWORSKI, J. (2008). A random mapping with preferential attachment. Submitted.

[11] JAWORSKI, J. (1990). Random mappings with independent choices of the images. In *Random Graphs*, Vol. 1, John Wiley, New York, pp. 89–101.

[12] JAWORSKI, J. (1998). Predecessors in a random mapping. *Random Structures Algorithms* **13,** 501–519.

[13] JAWORSKI, J. (1999). Epidemic processes on digraphs of random mappings. *J. Appl. Prob.* **36,** 780–798.

[14] JOHNSON, N. L., KOTZ, S. AND KEMP, A. W. (1992). *Univariate Discrete Distributions*, 2nd edn. John Wiley, New York.

[15] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, New York.

[16] MUTAFČIEV, L. (1982). A limit distribution related to random mappings and its application to an epidemic process. *Serdica* **8,** 197–203.

[17] MUTAFČIEV, L. (1984). On some stochastic problems of discrete mathematics. In *Proc. Math. Math. Education* (Sunny Beach, 1984), Akad. Nauk, Sofia, pp. 57–80.

[18] PITTEL, B. (1983). On distributions related to transitive closures of random finite mappings. *Ann. Prob.* **11,** 428–441.

[19] PRÜFER, H. (1918). Neuer Beweis eines Satzes uber Permutationen. *Archiv Math. Phys.* **27,** 142–144.

[20] QUISQUATER, J.-J. AND DELESCAILLE, J.-P. (1990). How easy is collision search? Application to DES. In *Advances in Cryptology—Eurocrypt'89* (Lecture Notes Comput. Sci. **434**), Springer, Berlin, pp. 429–434.

[21] VAN OORSCHOT, P. C. AND WIENER, M. J. (1994). Parallel collision search with applications to hash functions and discrete logarithms. In *Proc. 2nd ACM Conf. Comput. Commun. Security*, ACM, New York, pp. 210–218.