Random mappings with a given number of cyclical points

Jennie C. Hansen[*] and Jerzy Jaworski[†][‡]

Abstract

In this paper we consider a random mapping, $\hat{T}_n$, of the finite set $\{1, 2, ..., n\}$ into itself for which the digraph representation $\hat{G}_n$ is constructed by: (1) selecting a random number, $\hat{L}_n$, of cyclic vertices, (2) constructing a uniform random forest of size n with the selected cyclic vertices as roots, and (3) forming 'cycles' of trees by applying a random permutation to the selected cyclic vertices. We investigate $\hat{k}_n$, the size of a 'typical' component of $\hat{G}_n$, and, under the assumption that the random permutation on the cyclical vertices is uniform, we obtain the asymptotic distribution of $\hat{k}_n$ conditioned on $\hat{L}_n = m(n)$. As an application of our results, we show in Section 3 that provided $\hat{L}_n$ is of order much larger than $\sqrt{n}$, then the joint distribution of the normalized order statistics of the component sizes of $\hat{G}_n$ converges to the Poisson-Dirichlet(1) distribution as $n \to \infty$. Other applications and generalizations are also discussed in Section 3.

1 Introduction

In this paper we consider the distribution of the size of a 'typical' component in a random mapping from the set $V_n = \{1, 2, ..., n\}$ into V_n *conditioned* on the total number of cyclical vertices in the mapping. Before discussing the motivation for this investigation, we introduce some notation and review some well-known results for the component structure of the uniform random mapping.

[*]Actuarial Mathematics and Statistics and the Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK. E-mail address: J.Hansen@ma.hw.ac.uk

[†]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Umultowska 87, 61-614 Poznań, Poland. E-mail address: jaworski@amu.edu.pl

Throughout this paper, for $n \geq 1$, let $\mathcal{M}_n$ denote the set of mappings $f : V_n \to V_n$. For any $f \in \mathcal{M}_n$, we can represent f by a directed graph $G(f)$ on n labelled vertices such that a directed edge from vertex i to vertex j exists in $G(f)$ if and only if $f(i) = j$. We say that vertex i in $G(f)$ is a cyclic vertex if the m-fold composition $f^{(m)}(i) = i$ for some $m \geq 1$ and we let $L(f)$ denote the number of cyclic vertices in $G(f)$. We also note that since each vertex in $G(f)$ has out-degree 1, the components of $G(f)$ consist of directed cycles with directed trees attached. Finally, let T_n denote the uniform random mapping of V_n into V_n with distribution given by

$$\Pr\left\{T_n = f\right\} = \frac{1}{n^n}$$

for each $f \in \mathcal{M}_n$ and let $G_n \equiv G(T_n)$ denote the corresponding random directed graph which represents the random mapping T_n.

Much is known about the component structure of the random digraph G_n which represents T_n (see for example the monograph by Kolchin [13], survey by Mutafchiev [14], Aldous [1], Arratia et.al. [2], [3], Hansen [8]). In particular, it is known that, given $\mathcal{L}_n$, the set of cyclic vertices of T_n, the random mapping T_n *restricted* to $\mathcal{L}_n$ is a uniformly distributed random permutation on the vertices in $\mathcal{L}_n$. This observation motivates the following, alternative construction of the random digraph G_n as a special case of the general random mapping model described below.

A random mapping model
Suppose that $\hat{L}_n$ is a discrete random variable such that $1 \leq \hat{L}_n \leq n$. Also, for $1 \leq m$, let $\hat{\sigma}_m$ be a random, but not necessarily uniform, permutation of the set $\{1, 2, ..., m\}$ such that for such that for any subsets $B, C \subseteq \{1, 2, ..., m\}$ with $|B| = |C|$, we have

$$\Pr\{\hat{\sigma}_m \text{ restricted to B is a cyclic permutation }\}$$
$$= \Pr\{\hat{\sigma}_m \text{ restricted to C is a cyclic permutation }\}$$

For any subset $A \subseteq V_n$ such that $|A| = m$, we use the natural ordering of the set A to identify the elements of A with the elements of $\{1, 2, ..., m\}$, and we define $\hat{\sigma}_A$ to be the random permutation of A which is induced by the action $\hat{\sigma}_m$ via the identification of A with $\{1, 2, .., m\}$. Now the random mapping $\hat{T}_n : V_n \to V_n$ is constructed using $\hat{L}_n$ and the random permutations $\hat{\sigma}_m$ as follows: *Given* $\hat{L}_n = m$, let $\mathcal{A}_m$ denote a uniform random subset of size m from the vertices V_n (i.e. all subsets of size m are equally likely). *Given* $\mathcal{A}_m = A \subseteq V_n$, let $\mathcal{F}_n(A)$ denote the uniform random rooted forest on the vertices V_n, where A is the set of roots, and the edges in the trees of $\mathcal{F}_n(A)$ are directed such that any path from a vertex to a root is directed towards the root. Finally, suppose that $\hat{\sigma}_A$ is a

random permutation of A (as described above) which is also *independent* of the random forest $\mathcal{F}_n(A)$. We form the directed graph $\hat{G}_n$ from the rooted forest $\mathcal{F}_n(A)$ by adding a directed edge from $i \in A$ to $j \in A$ if $\hat{\sigma}_A(i) = j$, and we let $\hat{T}_n$ denote the random mapping which is represented by $\hat{G}_n$.

The construction described above gives us a class of random mapping models which are determined by the distribution chosen for $\hat{L}_n$, the number of cyclic vertices, and by the distributions of the random permutations $\hat{\sigma}_m$. We note that if $\hat{L}_n$ has the same distribution as L_n (denoted $\hat{L}_n \overset{d}{\sim} L_n$), where $L_n \equiv L(T_n)$ is the number of cyclic vertices in the uniform random mapping T_n, and if, for every $A \subseteq V_n$, $\hat{\sigma}_A = \sigma_A$, the uniform random permutation of A, then $\hat{G}_n \overset{d}{\sim} G_n$ and $\hat{T}_n \overset{d}{\sim} T_n$. On the other hand, if $\hat{L}_n \equiv n$ (i.e. $\hat{L}_n$ is degenerate), then $\hat{T}_n$ is just the random permutation $\hat{\sigma}_n$.

In general, it is of interest to determine the structure of $\hat{G}_n$ for various choices for the distributions of $\hat{L}_n$ and of the random permutations $\hat{\sigma}_m$. In particular, questions concerning the structure of $\hat{G}_n$ arise in the analysis of cryptographic systems (e.g. DES see [16]), in applications of Pollard's algorithm (see [15], [18]), in simulations of shift register sequences, and in computational number theory and random number generation. We note that since both uniform random mappings and uniform random permutations are special cases of our general model, we can view these models as part of a 'continuum' of random mapping models. Specifically, we note that for uniform random mappings, it is known (see [11]) that $L_n = O(\sqrt{n})$, and more precisely, the normalized variable $\frac{L_n}{\sqrt{n}}$ converges in distribution to L as $n \to \infty$, where L is a continuous random variable with density given by $f_L(x) = xe^{-x^2/2}$, $x \geq 0$. So, if $\hat{L}_n$ is of order much greater than $\sqrt{n}$ (but not necessarily of order n) and if, for $m \geq 1$, $\hat{\sigma}_m$, is a uniform permutation, then we obtain a model $\hat{G}_n$ which is 'sandwiched' (in some sense) between the uniform random mapping model and the uniform random permutation on V_n. One goal of this paper is to investigate the component structure of such models.

We begin our investigation by noting that if $\hat{T}_n$ is a random mapping as described above, and if $\phi : \mathcal{M}_n \to Z$ is a discrete functional, then the distribution of $\phi(\hat{T}_n)$ is determined by the distribution of $\hat{L}_n$ and by the conditional probabilities

$$\Pr\{\phi(\hat{T}_n) = k \,|\, \hat{L}_n = m\}. \tag{1.1}$$

Hence the conditional probabilities in (1.1) are fundamental to any investigation of $\hat{T}_n$ and of the corresponding random digraph $\hat{G}_n$. We note that sometimes it can be very easy to compute the conditional probabilities in (1.1). For example, suppose that, for $f \in \mathcal{M}_n$, $\phi(f)$ equals the number of components in $G(f)$, then the conditional distribution of $\phi(\hat{T}_n)$ *given*

that $\hat{L}_n = m$ is the same as the distribution of the number of cycles in the random permutation $\hat{\sigma}_m$. In the case where $\hat{\sigma}_m = \sigma_m$, the uniform random permutation of $\{1, 2, .., m\}$, the distribution of the number of cycles in σ_m is well-known (see Feller [7], p.258). However, for many functionals it can be much more complicated to compute the conditional probabilities in (1.1).

In this paper we investigate the conditional distribution of $\hat{k}_n \equiv k_n(\hat{T}_n)$ given $\hat{L}_n$, where, for $f \in \mathcal{M}_n$, $k_n(f)$ is defined to be the size of the component in $G(f)$ which contains the vertex 1. Since vertex 1 is 'arbitrary', we can say that $\hat{k}_n$ is the size of a 'typical' component in $\hat{G}_n$. We also note that by selecting a component in $\hat{G}_n$ which contains the vertex 1, we are selecting a component using 'size-biased' sampling. Size-biased sampling has been studied and used in the context of both uniform random mappings and uniform random permutations (see Aldous [1] and Vershik and Schmidt [17]). In Section 2, we determine a general formula (see Fact 1) which can be used to compute the conditional probabilities

$$\Pr\{\hat{k}_n = k \mid \hat{L}_n = m\}. \tag{1.2}$$

In the special case where, for $m \geq 1$, $\hat{\sigma}_m = \sigma_m$, the uniform random permutation on $\{1, 2, ..., m\}$, but $\hat{L}_n$ is arbitrary we obtain exact formulae for the conditional probabilities (1.2). A key observation in this case is that, as a consequence of the construction of $\hat{T}_n$, we have

$$\Pr\{\hat{k}_n = k \mid \hat{L}_n = m\} = \Pr\{k_n = k \mid L_n = m\}. \tag{1.3}$$

where $k_n \equiv k(T_n)$. So, in this case, it is enough to consider the conditional distribution of k_n given L_n. In Section 2, we also determine the asymptotic distribution of k_n conditioned on $L_n = m(n)$ under four distinct regimes: (i) $m(n) = O(1)$, (ii) $m(n) = o(\sqrt{n})$ but $m(n) \to \infty$ as $n \to \infty$, (iii) $m(n) = \alpha\sqrt{n}$, and (iv) $\sqrt{n} = o(m(n))$. We remark here that it is well-known for the uniform random mapping T_n that

$$\lim_{n\to\infty} \Pr\left\{a < \frac{k_n}{n} < b\right\} = \int_a^b \frac{1}{2\sqrt{1-u}} du \tag{1.4}$$

for any fixed $0 < a < b < 1$. This result can be obtained 'directly' , i.e. without using the the conditional distribution of k_n given L_n. However, in case of uniform random mappings, it follows from our results for regime (iii) described above that there is an interesting and complicated interaction, concealed in the asymptotic result (1.4), between the size of a typical component in G_n and the number of cyclic vertices in G_n.

In Section 3 we use our results for regime (iv) described above to establish Theorem 5, which is the main result of this paper. We show that provided $\sqrt{n} = o(\hat{L}_n)$ (in a sense which we make precise in the statement

of Theorem 5) and $\hat{\sigma}_m = \sigma_m$ for $m \geq 1$, then the joint distribution of the normalized order statistics of the component sizes of $\hat{G}_n$ converges, as $n \to \infty$, to the *Poisson − Dirichlet*(θ) distribution with parameter $\theta = 1$, which we denote by $\mathcal{PD}(1)$, on the simplex

$$\nabla = \left\{ \{x_i\} : \sum x_i \leq 1, x_i \geq x_{i+1} \geq 0 \text{ for every } i \geq 1 \right\}.$$

The key point of this theorem is that we always obtain the *same* limiting distribution provided *only* that, with high probability, the number of cyclic vertices, $\hat{L}_n$, is much larger than $\sqrt{n}$. In contrast, we note that in the uniform case the joint distribution of the normalized order statistics for the component sizes of G_n converges to the $\mathcal{PD}(1/2)$ distribution on ∇ as $n \to \infty$ ([1]). So Theorem 5 shows that there is a sharp qualitative difference between the component structure of $\hat{G}_n$ and of G_n as soon as $\sqrt{n} = o(\hat{L}_n)$.

Finally, throughout this paper we adopt the following abuse of notation: Suppose that $0 < x < \infty$ is fixed and $n \in Z^+$, then by 'integer $m = xf(n)$', where f is a function of Z^+, we mean $m = \lfloor xf(n) \rfloor$. Likewise, if X is an integer-valued random variable, by '$X = xf(n)$' we mean $X = \lfloor xf(n) \rfloor$.

2 The size of a connected component

We begin by giving a general probability formula which holds for any random mapping digraph $\hat{G}_n$ which is constructed as described in Section 1.

Fact 1. *Let $\hat{l}_n$ be the length of the cycle in the connected component in $\hat{G}_n$ to which the vertex 1 belongs, $\hat{k}_n$ be the size of this component and let $\hat{L}_n$ be the total number of cyclical vertices of $\hat{G}_n$. Then, for $m = 2, 3, \ldots, n$, $k = 1, 2, \ldots, n-2$ and $j = \max\{1, m-n+k+1\}, \ldots, \min\{k+1, m-1\}$, we have*

$$\Pr\{\hat{k}_n = k+1, \hat{l}_n = j \mid \hat{L}_n = m\} =$$

$$\binom{n-m}{k-j+1}\binom{m}{j} \frac{j\,(m-j)\,p_{m,j}}{nm} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2},$$

where $p_{m,j}$ denotes the probability that in the random permutation $\hat{\sigma}_m$ of m elements a given j-element set forms a cycle; and for $m = 1, \ldots, n$

$$\Pr\{\hat{k}_n = n, \hat{l}_n = m \mid \hat{L}_n = m\} = p_{m,m}$$

and if $j \neq m$

$$\Pr\{\hat{k}_n = n, \hat{l}_n = j \mid \hat{L}_n = m\} = 0.$$

Proof. Assume that $k \neq n - 1$ and consider the probability

$$\Pr\{\hat{k}_n = k + 1, \, \hat{l}_n = j, \, \hat{L}_n = m\} \, .$$

There are $\binom{n-1}{k}$ ways to choose k vertices which together with the vertex 1 will form the connected component, $\binom{k+1}{j}$ ways to choose the j vertices of the cycle in this component and finally there are $\binom{n-1-k}{m-j}$ ways to choose the remaining cyclical vertices. It follows from the definition of a random mapping model $\hat{T}_n$ that all such choices are equally likely. Moreover, a given m-element set will become a set of cyclic vertices of $\hat{T}_n$ with the probability $\Pr\{\hat{L}_n = m\}/\binom{n}{m}$; assuming this, a given j-element subset forms a cycle with probability $p_{m,j}$.

The arcs from the remaining $k + 1 - j$ vertices of the connected component generate a forest on $k + 1$ vertices with j trees rooted at the vertices of the cycle. By Cayley's formula there are $j(k + 1)^{k-j+1-1}$ such forests. Similarly the arcs from non-cyclic vertices outside the connected component form a forest on $n - k - 1$ vertices with $m - j$ trees rooted at the cyclical vertices; we have $(m - j)(n - k - 1)^{n-m-k+j-2}$ such forests. Hence there are $(m - j)j(k + 1)^{k-j+1-1}(n - k - 1)^{n-m-k+j-2}$ forests which satisfy the constraints, each appearing with the probability $\left(m \, n^{n-m-1}\right)^{-1}$. Since we are assuming that the forest is independent of the permutation $\hat{\sigma}_m$ of m cyclical vertices it follows that for $k \neq n - 1$

$$\Pr\{k_n(\hat{T}_n) = k + 1, \, l_n(\hat{T}_n) = j, \, \hat{L}_n = m\}$$

$$= \binom{n-1}{k}\binom{k+1}{j}\binom{n-1-k}{m-j}\frac{\Pr\{\hat{L}_n = m\}}{\binom{n}{m}}p_{m,j}$$

$$\times \frac{j(m - j)(k + 1)^{k-j}(n - k - 1)^{n-m-k+j-2}}{mn^{n-m-1}}$$

$$= \binom{n-m}{k-j+1}\binom{m}{j}\frac{j(m - j)p_{m,j}}{mn}\left(\frac{k+1}{n}\right)^{k-j+1}$$

$$\times \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2}\Pr\{\hat{L}_n = m\}.$$

In the same manner one can show the second assertion of Fact 1.

□

We now consider the special case where, for $m \geq 1$, $\hat{\sigma}_m = \sigma_m$, the uniform permutation on $\{1, 2, ..., m\}$. In light of (1.3), it is enough to consider the structure of the uniform random mapping digraph G_n given the number of cyclic vertices in G_n. In this case, a straightforward application of Fact 1 in the uniform case yields the following useful Corollaries.

346

Corollary 1. *Let l_n be the length of the cycle in the connected component in G_n to which the vertex 1 belongs, k_n be the size of this component and let L_n be the total number of cyclical vertices of G_n. Then, for $m = 2, \ldots, n$, $k = 1, \ldots, n-2$ and $j = \max\{1, m-n+k+1\}, \ldots, \min\{k+1, m-1\}$, we have*

$$\Pr\{k_n = k+1, \ l_n = j \mid L_n = m\}$$

$$= \binom{n-m}{k-j+1} \frac{m-j}{nm} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2} \tag{2.1}$$

and for $m = 1, \ldots, n$

$$\Pr\{k_n = n, \ l_n = m \mid L_n = m\} = \frac{1}{m}$$

and if $j \neq m$

$$\Pr\{k_n = n, \ l_n = j \mid L_n = m\} = 0.$$

Proof. Note that if $\hat{\sigma}_m$ is the uniform random permutation, then for $j = 1, 2, \ldots, m$

$$p_{m,j} = \frac{(j-1)!(m-j)!}{m!} = \frac{1}{j\binom{m}{j}}.$$

Hence the assertion follows directly from the Fact 1. $\square$

Summing the probabilities given by (2.1) over j leads us immediately to the following result.

Corollary 2. *Let k_n be the size of the connected component to which the vertex 1 belongs and L_n be the total number of cyclical vertices of G_n. Then, for $k = 0, \ldots, n-2$; $m = 2, \ldots, n$ we have*

$$\Pr\{k_n = k+1 \mid L_n = m\}$$

$$= \sum_{j=j^*}^{\min\{k+1,m-1\}} \frac{m-j}{nm} \binom{n-m}{k-j+1} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2}$$

$$= \sum_{t=\max\{0,k+2-m\}}^{\min\{k,n-m\}} \frac{m-k-1+t}{nm} \binom{n-m}{t} \left(\frac{k+1}{n}\right)^{t} \left(1 - \frac{k+1}{n}\right)^{n-m-t-1}$$

where $j^* = \max\{1, m-n+k+1\}$, and for $m = 1, \ldots, n$

$$\Pr\{k_n = n \mid L_n = m\} = \frac{1}{m}.$$

$\square$

Before proceeding to prove the main results of this section we make a few remarks. First, for the applications considered in Section 3 below, we are interested in the case where the number of cyclic vertices in G_n is at least $O(\sqrt{n})$. However, for completeness, in Theorems 1 and 2 we also consider the case where the number of cyclic vertices in G_n is $o(\sqrt{n})$. We note that in this case, the size of the typical component is $n - o(n)$, and so in Theorems 1 and 2 the variable of interest is $n - k_n$, i.e. the number of vertices *not* contained in the component which contains vertex 1. Theorems 1 and 2 may also be of independent interest.

Theorem 1. *Suppose that $m = O(1)$ as $n \to \infty$. Then the number of vertices outside the typical connected component, given that $L_n = m$, has asymptotically the following discrete distribution:*

$$\Pr\left\{n - k_n = \ell \,\middle|\, L_n = m\right\} \sim \begin{cases} \dfrac{1}{m} & \text{if } \ell = 0 \\[2ex] \dfrac{1}{m} \dfrac{\ell^\ell}{\ell!} e^{-\ell} & \text{if } \ell = 1, ..., m-1 \\[2ex] \dfrac{1}{m}\left(\dfrac{\ell^\ell}{\ell!} - \dfrac{\ell^{\ell-m}}{(\ell-m)!}\right) e^{-\ell} & \text{if } \ell = m, m+1, ... \end{cases}$$

Proof. Let us assume that $m = O(1)$ as $n \to \infty$. From the Corollary 1 we have
$$\Pr\{n - k_n = 0 \,|\, L_n = m\} = \frac{1}{m}$$
and for $\ell > 0$ (by Corollary 2)

$$\Pr\{n - k_n = \ell \,\big|\, L_n = m\} = \Pr\{k_n = n - \ell \,|\, L_n = m\}$$

$$= \sum_{j=\max\{1, m-\ell\}}^{\min\{n-\ell, m\}} \frac{m-j}{\ell m}\binom{n-m}{\ell-m+j}\left(\frac{\ell}{n}\right)^{\ell-m+j}\left(1 - \frac{\ell}{n}\right)^{n-\ell-j} \tag{2.2}$$

Therefore for $\ell = 1, 2, \dots, m-1$ we obtain

$$\Pr\{n - k_n = \ell \,|\, L_n = m\}$$

$$\sim \sum_{j=m-\ell}^{m} \frac{m-j}{m\ell} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!} = \sum_{j=m-\ell}^{m} \frac{\ell - (\ell - m + j)}{m\ell} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!}$$

$$= \frac{1}{m} \sum_{j=m-\ell}^{m} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!} - \frac{1}{m} \sum_{j=m-\ell+1}^{m} \frac{\ell^{\ell-m+j-1} e^{-\ell}}{(\ell-m+j-1)!} = \frac{1}{m} \frac{\ell^{\ell} e^{-\ell}}{\ell!}.$$

Similarly for any bounded $\ell = m, m+1, \ldots$ we have

$$\Pr\{n - k_n = \ell \,|\, L_n = m\}$$

$$\sim \sum_{j=1}^{m} \frac{m-j}{m\ell} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!} = \sum_{j=1}^{m} \frac{\ell - (\ell-m+j)}{m\ell} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!}$$

$$= \frac{1}{m} \sum_{j=1}^{m} \frac{\ell^{\ell-m+j} e^{-\ell}}{(\ell-m+j)!} - \frac{1}{m} \sum_{j=1}^{m} \frac{\ell^{\ell-m+j-1} e^{-\ell}}{(\ell-m+j-1)!}$$

$$= \frac{1}{m} \frac{\ell^{\ell} e^{-\ell}}{\ell!} - \frac{1}{m} \frac{\ell^{\ell-m} e^{-\ell}}{(\ell-m)!}.$$

To show that this is a proper probability distribution one can use another discrete distribution (see [12]), namely

$$p_k = a \frac{(a+k)^{k-1}}{k!} e^{-a-k}, \quad k = 0, 1, \ldots, \quad \text{where } a \text{ is a positive integer}.$$

$\square$

Theorem 2. *Suppose that $m = o(\sqrt{n})$ but $m \to \infty$ as $n \to \infty$, and suppose y is fixed, $0 < y < \infty$. Then*

$$\Pr\left\{\frac{n - k_n}{m^2} = y \,\middle|\, L_n = m\right\} \sim \frac{1}{m^2} \frac{1}{\sqrt{2\pi y}} \left(1 - \exp\left\{-\frac{1}{2y}\right\}\right).$$

Proof. Let us assume that $m = o(\sqrt{n})$ but $m \to \infty$ as $n \to \infty$, $j = mx$, where x is fixed, $0 < x < 1$ and $n - k_n = \ell = ym^2$, where $0 < y < \infty$ is fixed. Then

$$\frac{\ell - m + j - (n-m)\frac{\ell}{n}}{\sqrt{(n-m)\frac{\ell}{n}\left(1 - \frac{\ell}{n}\right)}} = \frac{-m + xm + \frac{m\ell}{n}}{\sqrt{\ell}\sqrt{\left(1 - \frac{m}{n}\right)\left(1 - \frac{\ell}{n}\right)}} \sim -\frac{1}{\sqrt{y}} + \frac{x}{\sqrt{y}},$$

since $\ell = o(n)$ and $j = o(\sqrt{n})$. Therefore the approximation given by the local de Moivre-Laplace theorem is applicable for the binomial probability (see Feller [7])

$$\binom{n-m}{\ell-m+j} \left(\frac{\ell}{n}\right)^{\ell-m+j} \left(1 - \frac{\ell}{n}\right)^{n-\ell-j}$$

and together with (2.2) it implies that

349

$$\Pr\left\{\frac{n-k_n}{m^2} = y \,\middle|\, L_n = m\right\} \sim \frac{1}{m^2} \int_0^1 \frac{1-x}{y} \frac{1}{\sqrt{y}\sqrt{2\pi}} \exp\left\{-\frac{(1-x)^2}{2y}\right\} dx$$

Taking $u = \dfrac{(1-x)^2}{2y}$, we have $du = -\dfrac{1-x}{y}\,dx$ and

$$\Pr\left\{\frac{n-k_n}{m^2} = y \,\middle|\, L_n = m\right\} \sim \frac{1}{m^2} \frac{1}{\sqrt{2\pi y}} \int_0^{\frac{1}{2y}} \exp\left\{-u\right\} du,$$

which immediately leads to the assertion.

$\square$

The main application of the following result is given is a companion paper which investigates a cutting process for random mappings (see [10]). In that application we require error bounds for the local limit, so they are included in the statement of the Theorem 3.

Theorem 3. *Suppose that $\alpha > 0$, $0 < \delta < 1/2$, $0 < \delta < x < 1 - \delta < 1$ and n is such that $\delta^5 \sqrt{n} > \alpha^3 \vee 1$. Then*

$$\Pr\left\{\frac{k_n}{n} = x \,\middle|\, \frac{L_n}{\sqrt{n}} = \alpha\right\}$$

$$= \frac{1}{n} \frac{\alpha}{\sqrt{2\pi x(1-x)^3}} \int_0^1 (1-y) \exp\left\{-\frac{(x-y)^2 \alpha^2}{2x(1-x)}\right\} dy(1+\varepsilon(n,\alpha,x)) \quad (2.3)$$

where $|\varepsilon(n,\alpha,x)| \leq C\left(\dfrac{\alpha^3 \vee 1}{\min(\delta^5,\alpha)}\right) \dfrac{\exp(\alpha^2/\delta^2)}{\sqrt{n}}$ and C is a constant which does not depend on n, α, or δ.

Proof. Let $m = \alpha\sqrt{n}$ and $k+1 = xn$, then it follows from Corollary 2 and a careful application of the usual deMoivre-Laplace local limit theorem for the binomial distribution (see Feller [7]), that

$$\Pr\{k_n = k+1 \,|\, L_n = m\}$$

$$= \sum_{j=1}^m \frac{m-j}{nm} \binom{n-m}{k-j+1} \left(\frac{k+1}{n}\right)^{k-j+1} \left(1-\frac{k+1}{n}\right)^{n-m-k+j-2} \quad (2.4)$$

$$= \frac{1}{n} \frac{\alpha}{\sqrt{2\pi x(1-x)^3}} \sum_{j=1}^m \frac{1}{m}\left(1-\frac{j}{m}\right) \exp\left\{-\frac{(x-\frac{j}{m})^2 \alpha^2}{2x(1-x)}\right\} (1+\Delta(n,\alpha,x)),$$

where $|\Delta(n, \alpha, x)| \le \tilde{C} \left(\frac{\alpha^3 \vee 1}{\min(\delta^5, \alpha)} \right) \frac{1}{\sqrt{n}}$ and $\tilde{C}$ is a constant which does not depend on n, α, or δ. Next, since

$$\left| \sum_{j=1}^m \frac{1}{m} (1 - \frac{j}{m}) \exp \left\{ -\frac{(x - \frac{j}{m})^2 \alpha^2}{2x(1-x)} \right\} - \int_0^1 (1 - y) \exp \left\{ -\frac{(x-y)^2 \alpha^2}{2x(1-x)} \right\} dy \right|$$

$$\le \frac{4}{m} \tag{2.5}$$

and, for $0 < \delta < x < 1 - \delta < 1$,

$$\int_0^1 (1 - y) \exp \left\{ -\frac{(x-y)^2 \alpha^2}{2x(1-x)} \right\} dy \ge \frac{1}{2} \exp(-\alpha^2/\delta^2), \tag{2.6}$$

we can obtain (2.3) from (2.4)-(2.6).

We note that one can show that

$$\int_0^1 \frac{\alpha}{\sqrt{2\pi x(1-x)^3}} \int_0^1 (1 - y) \exp \left\{ -\frac{(x-y)^2 \alpha^2}{2x(1-x)} \right\} dy dx = 1$$

(i.e. that the local limit is a the proper density function) by using the substitution

$$z = \frac{(x-y)\alpha}{\sqrt{x(1-x)}},$$

changing the order of integration and some "routine" calculations. □

We note that in Theorem 3 above, the value $\frac{L_n}{\sqrt{n}} = \alpha$ is a *parameter* in the limiting conditional distribution of $\frac{k_n}{n}$ given $\frac{L_n}{\sqrt{n}}$. Recall that as $n \to \infty$, $\frac{L_n}{\sqrt{n}}$ converges in distribution to a variable L with density $f_L(\alpha) = \alpha e^{-\alpha^2/2}, \alpha > 0$. It can be checked that the integral over $(0, \infty)$ of the local limit in Theorem 3 with respect to the density $f_L(\alpha)$ yields, in light of (1.4), the expected result, i.e.

$$\frac{1}{n} \int_0^\infty \frac{\alpha^2 e^{\alpha^2/2}}{\sqrt{2\pi x(1-x)^3}} \int_0^1 (1 - y) \exp \left\{ -\frac{(x-y)^2 \alpha^2}{2x(1-x)} \right\} dy d\alpha = \frac{1}{n} \frac{1}{2\sqrt{1-x}}.$$

Finally, in Theorem 4 below, we consider the case where $\sqrt{n} = o(m)$. In order to apply this result to prove our main result in the next section, we have given the error bounds for the local limit in the statement of the theorem.

Theorem 4. *Suppose that $\sqrt{n} = o(m)$, a is fixed, $0 < a < \frac{1}{2}$, and let fix x, $0 < a < x < 1 - a < 1$. Then, as $n \to \infty$,*

$$\Pr\left\{\frac{k_n}{n} = x \,\Big|\, L_n = m\right\} = \frac{1}{n}\left(1 + \varepsilon(n, m, x)\right),$$

where $|\varepsilon(n, m, x)| \leq \frac{(n-m)^{1/4}}{\sqrt{a\,m}}$.

Proof. First, we suppose that $n - m \to \infty$ as $n \to \infty$. We also fix $0 < a < \frac{1}{2}$ and suppose that $k + 1 = n\,x$, where x is fixed, $0 < a < x < 1 - a < 1$. Let $\Delta(j) \equiv j - mx$ and let $\delta(n, m, a) \equiv \sqrt{a\,m}(n - m)^{1/4}$. Then by Corollary 2, we have

$$\Pr\{k_n = k + 1 \,|\, L_n = m\} = S_1 + S_2 + S_3 = \qquad (2.7)$$

$$\sum_{j=j^*}^{\min\{k+1,m\}} \frac{1_{\{|\Delta(j)|\leq\delta(n,m,a)\}}}{n} \binom{n - m}{xn - j} x^{xn-j}(1 - x)^{n-m-xn+j}$$

$$- \sum_{j=j^*}^{\min\{k+1,m\}} \frac{1_{\{|\Delta(j)|\leq\delta(n,m,a)\}}}{n(1 - x)} \frac{\Delta(j)}{m} \binom{n - m}{xn - j} x^{xn-j}(1 - x)^{n-m-xn+j}$$

$$+ \sum_{j=j^*}^{\min\{k+1,m\}} \frac{1_{\{|\Delta(j)|>\delta(n,m,a)\}}}{n(1 - x)} \frac{m - j}{m} \binom{n - m}{xn - j} x^{xn-j}(1 - x)^{n-m-xn+j}$$

where $j^* = \max\{1, m - n + k + 1\}$.
We consider the three sums in (2.7). Let X be a Binomial$(n-m, x)$ variable and let $Y = (X - (n - m)x)/\sqrt{(n - m)x(1 - x)}$. Then we have

$$S_3 \leq \sum_{j=j^*}^{\min\{k+1,m\}} \frac{1_{\{|\Delta(j)|>\delta(n,m,a)\}}}{n(1 - x)} \binom{n - m}{xn - j} x^{xn-j}(1 - x)^{n-m-xn+j}$$

$$\leq \frac{1}{n(1 - x)}\Pr\left\{|Y| > \frac{\delta(n, m, a)}{\sqrt{(n - m)x(1 - x)}}\right\} \leq \frac{1}{n}\left(\frac{\sqrt{n - m}}{a\,m}\right)$$

where $j^* = \max\{1, m - n + k + 1\}$ and the last inequality follows from Chebyshev's inequality. Next, observe that if $|\Delta(j)| \leq \delta(n, m, a)$, then

$$\left|\frac{\Delta(j)}{m}\right| \leq \frac{\sqrt{a}(n - m)^{1/4}}{\sqrt{m}},$$

and it follows that

$$|S_2| \leq \frac{1}{n(1 - x)} \frac{\sqrt{a}(n - m)^{1/4}}{\sqrt{m}} \leq \frac{1}{n}\left(\frac{(n - m)^{1/4}}{\sqrt{a\,m}}\right).$$

Finally, it is clear that $S_1 \leq \frac{1}{n}$. A lower bound for the sum S_1 is given by

$$S_1 \geq \frac{1}{n} \sum_{\ell=\max\{0,(n-m)x-\delta(n,m,a)\}}^{\min\{n-m,(n-m)x+\delta(n,m,a)\}} \binom{n-m}{\ell}(x)^\ell(1-x)^{n-\ell}$$

$$= \frac{1}{n}\Pr\left\{|Y| \leq \frac{\delta(n,m,a)}{\sqrt{(n-m)x(1-x)}}\right\} \geq \frac{1}{n}\left(1 - \frac{\sqrt{n-m}}{am}\right),$$

where the random variable Y is as defined above. So it follows from (2.7) and the calculations above that for $k+1 = xn$, where $0 < a < x < 1-a < 1$, $n - m \to \infty$, we have

$$\Pr\{k_n = k + 1 \mid L_n = m\} = \frac{1}{n}(1 + \varepsilon(n,m,x)) \tag{2.8}$$

where $|\varepsilon(n,m,x)| \leq \max\{\frac{(n-m)^{1/4}}{\sqrt{am}}, \frac{\sqrt{n-m}}{am}\}$.

Finally, we consider the case when $n - m = i = O(1)$ and, as previously, fix $0 < a < \frac{1}{2}$ and $k + 1 = xn$, where $0 < a < x < 1 - a < 1$. Then

$$\Pr\{k_n = k + 1 \mid L_n = m\}$$

$$= \sum_{j=k+1-i}^{k+1} \frac{m-j}{nm}\binom{n-m}{k-j+1}\left(\frac{k+1}{n}\right)^{k-j+1}\left(1 - \frac{k+1}{n}\right)^{n-m-k+j-2}$$

$$= \sum_{t=0}^{i} \frac{m - mx + mx - xn + t}{nm}\binom{i}{t}(x)^t(1-x)^{i-t-1}$$

$$= \frac{1}{n} + \frac{1}{n(1-x)}\sum_{t=0}^{i}\frac{-ix+t}{m}\binom{i}{t}x^t(1-x)^{i-t} = \frac{1}{n}(1 + \varepsilon(n,m,x))$$

where, in this case, $|\varepsilon(n,m,x)| \leq \frac{i}{am}$. This together with (2.8) gives the desired result.

$\square$

3 Applications and discussion

In this section we prove our main result, Theorem 5. We also discuss some related applications of our results and suggest some directions for the future work. We begin by considering a random mapping digraph $\hat{G}_n$ where, for $m \geq 1$, $\hat{\sigma}_m = \sigma_m$, the uniform random permutation on $\{1, 2, ..., m\}$. If we

consider the component structure of $\hat{G}_n$ as a 'function' of the number of cyclic vertices, $\hat{L}_n$, then, in some sense, $\sqrt{n}$ is a threshold for the number of cyclic vertices. Specifically, if there exists $m(n) = o(\sqrt{n})$ such that $\Pr\{\hat{L}_n < m(n)\} \to 1$ as $n \to \infty$, then it follows from Theorems 1 and 2 that for any constant $C > 0$ we have

$$\Pr\left\{\frac{\hat{k}_n}{n} \geq 1 - \frac{C(m(n))^2}{n}\right\} \to 1 \quad \text{as } n \to \infty.$$

In other words, with high probability, $\hat{G}_n$ consists of one large component of order n and the remaining components are of order at most $(m(n))^2$. If $\hat{L}_n$ is of order $\sqrt{n}$ (as in the case of the uniform random mapping) then the conditional distribution of $\frac{\hat{k}}{n}$ is parameterised by the value of $\frac{\hat{L}_n}{\sqrt{n}}$ and the asymptotic distribution of $\frac{\hat{k}_n}{n}$ may depend on the distribution of $\hat{L}_n$ in a complicated way. Finally, if $\hat{L}_n$ is of order greater than $\sqrt{n}$ then the conditional distribution of $\frac{\hat{k}_n}{n}$ given $\hat{L}_n$ is asymptotically uniform on $[0, 1]$ and is 'independent' of the exact distribution of $\hat{L}_n$. We exploit this 'independence' of $\frac{\hat{k}_n}{n}$ and $\hat{L}_n$ to prove the our main result which completely characterises of the asymptotic joint distribution of the order statistics of the normalised component sizes of $\hat{G}_n$ in this case:

Theorem 5. *Suppose that $\hat{L}_1, \hat{L}_2, \ldots$ is a sequence of discrete random variables such that for each $n \geq 1$, $1 \leq \hat{L}_n \leq n$. Also, suppose that there exists $m(n)$ such that $\sqrt{n} = o(m(n))$ and such that $\alpha(n) \equiv \Pr\{\hat{L}_n < m(n)\} \to 0$ as $n \to \infty$. Finally, suppose that, for $m \geq 1$, $\hat{\sigma}_m = \sigma_m$, a uniform permutation on $\{1, 2, \ldots, m\}$. Then the joint distribution of the order statistics of the normalised component sizes of $\hat{G}_n$ converges to the Poisson-Dirichlet(1) distribution on the simplex*

$$\nabla = \left\{\{x_i\} : \sum x_i \leq 1, x_i \geq x_{i+1} \geq 0 \text{ for every } i \geq 1\right\}$$

as $n \to \infty$.

To describe the main steps in the proof, we need to introduce some notation and state a sufficient condition for convergence to the Poisson-Dirichlet(1) distribution on ∇.

First, given $\hat{G}_n$, let $\hat{\mathcal{K}}_n^{(1)}$ denote the component in $\hat{G}_n$ which contains vertex labelled 1. If $\hat{\mathcal{K}}_n^{(1)} \neq \hat{G}_n$, then let $\hat{\mathcal{K}}_n^{(2)}$ denote the component in $\hat{G}_n \setminus \hat{\mathcal{K}}_n^{(1)}$ which contains the smallest vertex; otherwise, set $\hat{\mathcal{K}}_n^{(2)} = \emptyset$. For $i > 2$, we define $\hat{\mathcal{K}}_n^{(i)}$ iteratively: If $\hat{G}_n \setminus (\hat{\mathcal{K}}_n^{(1)} \cup \ldots \cup \hat{\mathcal{K}}_n^{(i-1)}) \neq \emptyset$, then let $\hat{\mathcal{K}}_n^{(i)}$ denote the component in $\hat{G}_n \setminus (\hat{\mathcal{K}}_n^{(1)} \cup \ldots \cup \hat{\mathcal{K}}_n^{(i-1)})$ which contains the

smallest vertex; otherwise, set $\hat{\mathcal{K}}_n^{(i)} = \emptyset$. For $i \geq 1$, let $\hat{k}_n^{(i)} = |\hat{\mathcal{K}}_n^{(i)}|$ and define the sequence $(\hat{z}_n^{(1)}, \hat{z}_n^{(2)}, \ldots)$ by

$$\hat{z}_n^{(1)} = \frac{\hat{k}_n^{(1)}}{n}, \ \hat{z}_n^{(2)} = \frac{\hat{k}_n^{(2)}}{n - \hat{k}_n^{(1)}}, \ \ldots, \ \hat{z}_n^{(i)} = \frac{\hat{k}_n^{(i)}}{n - \hat{k}_n^{(1)} - \hat{k}_n^{(2)} - \ldots - \hat{k}_n^{(i-1)}}, \ \cdots$$

where $\hat{z}_n^{(i)} = 0$ if $n - \hat{k}_n^{(1)} - \hat{k}_n^{(2)} - \ldots - \hat{k}_n^{(i-1)} = 0$. For $i \geq 1$, we also define $\hat{d}_n^{(i)}$ to be the size of the ith largest component in $\hat{G}_n$. Finally, for the *uniform* random mapping digraph G_n, the definitions of $\mathcal{K}_n^{(i)}$, $k_n^{(i)}$, $z_n^{(i)}$, and $d_n^{(i)}$ are analogous to the definitions given above.

Now it is well-known (see, for example, Hansen [9] and references therein) that to show that the joint distribution of the normalized order statistics, $\left(\frac{d_n^{(1)}}{n}, \frac{d_n^{(2)}}{n}, \ldots\right)$, converges to the $\mathcal{PD}(1)$ distribution on ∇, it is sufficient to show that for each $t \geq 1$ and $0 < a_i < b_i < 1$, $i = 1, 2, \ldots, t$, we have

$$\lim_{n \to \infty} \Pr\left\{ a_i < \hat{z}_n^{(i)} \leq b_i : 1 \leq i \leq t \right\} = \prod_{i=1}^{t}(b_i - a_i). \quad (3.1)$$

The proof of (3.1) is by induction on t and we give a sketch of this proof below.

Sketch of proof of Theorem 5. First, suppose that $t = 1$ and let $\gamma_1 = \min(a_1, 1 - b_1)$. Then we have

$$\Pr\{a_1 \leq \frac{\hat{k}_n^{(1)}}{n} \leq b_1\}$$

$$= \sum_{m \geq m(n)} \Pr\{a_1 \leq \frac{\hat{k}_n^{(1)}}{n} \leq b_1 \,|\, \hat{L}_n = m\}\Pr\{\hat{L}_n = m\} + \alpha(a_1, b_1, n) \quad (3.2)$$

where $0 \leq \alpha(a_1, b_1, n) \leq \alpha(n)$. Now it follows from (1.3) and Theorem 4 that

$$\sum_{m \geq m(n)} \Pr\{a_1 \leq \frac{\hat{k}_n^{(1)}}{n} \leq b_1 \,|\, \hat{L}_n = m\}\Pr\{\hat{L}_n = m\}$$

$$= \sum_{m \geq m(n)} \Pr\{a_1 \leq \frac{k_n^{(1)}}{n} \leq b_1 \,|\, L_n = m\}\Pr\{\hat{L}_n = m\}$$

$$= \sum_{m \geq m(n)} \left(\frac{\lfloor b_1 n \rfloor - \lceil a_1 n \rceil}{n} + \varepsilon(a_1, b_1, n, m(n))\right)\Pr\{\hat{L}_n = m\}$$

$$= \left(\frac{\lfloor b_1 n \rfloor - \lceil a_1 n \rceil}{n}\right)\Pr\{\hat{L}_m \geq m(n)\} + \sum_{m \geq m(n)} \varepsilon(a_1, b_1, n, m)\Pr\{\hat{L}_n = m\}$$

$$(3.3)$$

where, for all sufficiently large n and $m \geq m(n)$,

$$|\varepsilon(a_1, b_1, n, m)| \leq \frac{(n - m(n))^{1/4}}{\sqrt{\gamma_1 m(n)}} \,.$$

Since $\Pr\{\hat{L}_n \geq m(n)\} \to 1$ and $\frac{(n-m(n))^{1/4}}{\sqrt{\gamma_1 m(n)}} \to 0$ as $n \to \infty$, the result for $t = 1$ follows from (3.2) and (3.3).

Next, we sketch the induction step by considering the proof of the case $t = 2$ given that the result holds for $t = 1$. First, as above, it is enough to consider

$$\Pr\left\{ a_1 \leq \frac{k_n^{(1)}}{n} \leq b_1, a_2 \leq \frac{k_n^{(2)}}{n - k_n^{(1)}} \leq b_2 \,\middle|\, L_n = m \right\}$$

for $m \geq m(n)$. Now for $i \geq 1$, let $\ell_n^{(i)}$ denote the number of cyclic vertices in the component $\mathcal{K}_n^{(i)}$ (where $\ell_n^{(i)} = 0$ if $\mathcal{K}_n^{(i)} = \emptyset$). The key to the induction is the observation (which follows from the definition of $\hat{G}_n$ using uniform random permutations) that if $k_n^{(1)} = xn$ for some $a_1 < x < b_1$ then

$$\Pr\left\{ a_2 \leq \frac{k_n^{(2)}}{n - xn} \leq b_2 \,\middle|\, \ell_n^{(1)} = l, k_n^{(1)} = xn, L_n = m \right\}$$

$$= \Pr\left\{ a_2 \leq \frac{k_{n-xn}^{(1)}}{n - xn} \leq b_2 \,\middle|\, L_{n-xn} = m - l \right\}. \tag{3.4}$$

Now it can be shown (by carefully considering (2.1)) that *given* $k_n^{(1)} = xn$ and $L_n = m \geq m(n)$, we have $m - \ell_n^{(1)} \geq \frac{(1-b_1)}{2} m(n)$ with (uniformly) high probability. Standard arguments using Theorem 4 and (3.4) yield

$$\left| \Pr\left\{ a_2 \leq \frac{k_n^{(2)}}{n - k_n^{(1)}} \leq b_2 \,\middle|\, a_1 \leq \frac{k_n^{(1)}}{n} \leq b_1, L_n = m \right\} - (b_2 - a_2) \right|$$

$$\leq \hat{\varepsilon}(a_1, b_1, a_2, b_2, n, m(n)) \tag{3.5}$$

for $m \geq m(n)$, where $\hat{\varepsilon}(a_1, b_1, a_2, b_2, n, m(n)) \to 0$ as $n \to \infty$. The result for $t = 2$ now follows from (3.5), and the result for $t = 1$. The general induction argument is similar to the argument sketched above, but more cumbersome to write down. $\qquad\square$

Another application of our results is given in a companion paper (Hansen and Jaworski [10]) which considers a cutting process for a uniform random mapping. In [10] we consider the component structure of a uniform random

mapping that has been altered by randomly cutting and deleting non-cyclic edges (and vertices) in G_n. The structure of the resulting random directed graph depends on both the (random) number of edges and vertices that have been deleted from G_n and on the initial number of cyclic vertices in G_n. The resulting 'trimmed' random mapping model is a random structure which is also (in some sense) sandwiched between a uniform random mapping and a uniform random permutation. Specifically, if no edges are cut, we have a uniform random mapping, whereas if all the trees are trimmed down to their roots, we have a random permutation on the root vertices. Recall that the joint distribution of the normalized order statistics of the component sizes in G_n converges to the $\mathcal{PD}(1/2)$ distribution as $n \to \infty$ (see [1]), whereas the joint distribution of the normalized order statistics of the cycle lengths in the uniform random permutation σ_m converges to the $\mathcal{PD}(1)$ distribution as $m \to \infty$ (see [17]). Given these results, one might suppose that if $m(n)$ edges are cut, where $m(n) \to \infty$ as $n \to \infty$, then the joint distribution of the normalized order statistics of the component sizes of the resulting 'trimmed' random mapping converges to the $\mathcal{PD}(\theta)$ distribution for some $1/2 < \theta < 1$ (where the value of θ may depend on how $m(n)$ goes to infinity). In fact, we show in [10] that there is no smooth transition from the $\mathcal{PD}(1/2)$ distribution to the $\mathcal{PD}(1)$ distribution as the number of edges cut in G_n increases relative to n. More precisely, we show that there is a 'phase transition' when $m(n) = \beta\sqrt{n}$, where $\beta > 0$ is a fixed parameter, and in this case we show that the limiting distribution cannot be $\mathcal{PD}(\theta)$.

In light of Theorem 5 and our results for 'trimmed' random mappings, it would be of interest to determine whether there is some $\hat{L}_n$ (of order $\sqrt{n}$) such that the joint distribution of the normalized order statistics of the component sizes of $\hat{G}_n$ (constructed using uniform permutations) converges to the $\mathcal{PD}(\theta)$ for some $1/2 < \theta < 1$. In another direction, we mention that functional central limit theorems have been obtained for the number of cycles in a uniform random permutation (see [5]) and for the number of components in G_n (see [8]). It is likely that a functional central limit theorem also holds for the number of components $\hat{G}_n$, and we expect that the normalization in the functional central limit theorem will depend on the distribution of $\hat{L}_n$.

Finally, we note that the results in this paper have been obtained under the assumption that the permutations used to construct $\hat{G}_n$ are *uniformly* distributed. If we consider Fact 1 above, we see that when $\hat{\sigma}_m$ is a uniform random permutation we can compute the values of $p_{m,j}$ exactly to obtain Corollary 2. One may ask how far we can perturb the product $\binom{m}{j}jp_{m,j}$ from 1 and still obtain the same asymptotic results. In another direction, it would be interesting to use permutations $\hat{\sigma}_m^\theta$ to construct $\hat{G}_n$, where $\hat{\sigma}_m^\theta$ is a random permutation on $\{1, 2, ..., m\}$ with cycle structure given by the

Ewens sampling formula (see [4], p.60) with parameter $\theta > 0$. We note that the case $\theta = 1$ corresponds to the uniform random permutation on $\{1, 2, ..., m\}$. So it would be interesting to investigate how the structure of G_n varies as the parameter θ varies and to determine the corresponding threshold for the number of cyclic vertices in this case as well.

References

[1] Aldous, D., *Exchangeability and related topics*, Lecture Notes in Math, **1117**, Springer-Verlag, New York (1985).

[2] Arratia, R. and Tavaré, S., Limit theorems for combinatorial structures via discrete process approximations, *Random Structures and Algorithms*, **3**, (1992) 321–345.

[3] Arratia, R., Stark, D. and Tavaré, S., Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies, *The Annals of Probability*, **23**, (1995) 1347–1388.

[4] Arratia, R., Barbour, A. D. and Tavaré, S., *Logarithmic Combinatorial Structures: a Probabilistic Approach*, European Mathematical Society, Zurich (2003).

[5] DeLaurentis, J. M. and Pittel, B., Random permutations and Brownian motion, *Pacific J. Math.*, **119**, (1985) 287–301.

[6] DeLaurentis, J. M., Components and cycles of a random function, *Lecture Notes in Computer Science*, **293**, Advances in Cryptology - CRYPTO'87 Proceedings, (1988) 231–242.

[7] Feller, W., *An Introduction to Probability Theory and its Applications*, Vol I, 3rd edition, John Wiley and Sons, New York (1970).

[8] Hansen, J. C., A functional central limit theorem for random mappings, *Ann. of Probab.*, **17**, (1989) 317–332.

[9] Hansen, J. C., Order statistics for decomposable combinatorial structures, *Random Structures and Algorithms*, **5**, (1994) 517–533.

[10] Hansen, J. C. and Jaworski, J., A cutting process for random mappings, *Random Structures and Algorithms*, **30**, (2007) 287–306.

[11] Harris, B., Probability distribution related to random mappings, *Ann. Math. Statist.* **31**, (1960) 1045–1062.

[12] Jaworski, J., Epidemic processes on digraphs of random mappings, *J. Appl. Prob.* **36**, (1999) 1–19.

[13] Kolchin, V. F., *Random Mappings*, Optimization Software Inc., New York (1986).

[14] Mutafchiev, L., On some stochastic problems of discrete mathematics, in: *Mathematics and Education in Mathematics*, Bulg.Akad.Nauk, Sofia, (1984) 57–80.

[15] Pollard, J. M., A Monte Carlo method for factorization, *BIT*, **15(3)**, (1975) 331-334.

[16] Quisquater, J. J. and Delescaille, J. P., How easy is collision search? Application to DES, *Lecture Notes in Computer Science*, **434**, Advances in Cryptology - CRYPTO'89 Proceedings, (1990) 429–434.

[17] Vershik, A. M. and Schmidt, A. A., Limit measures arising in the asymptotic theory of symmetric groups. I, *Theor. Probab. Appl.*, **22**, (1977) 70–85.

[18] van Oorschot, P. C., and Wiener, M. J., Parallel collision search with cryptanalytic applications, *Journal of Cryptology*, **12(1)**, (1999) 1–28.