Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Ann. I. H. Poincaré – PR 43 (2007) 353–374
ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.elsevier.com/locate/anihpb

The Markov chain asymptotics of random mapping graphs

Xinxing Chen, Jiangang Ying [*,1]

Department of Mathematics, Fudan University, Shanghai, China

Received 16 November 2005; accepted 30 May 2006

Available online 15 December 2006

Abstract

In this paper the limit behavior of random mappings with n vertices is investigated. We first compute the asymptotic probability that a fixed class of finite non-intersected subsets of vertices are located in different components and use this result to construct a scheme of allocating particles with a related Markov chain. We then prove that the limit behavior of random mappings is actually embedded in such a scheme in a certain way. As an application, we shall give the asymptotic moments of the size of the largest component.

Résumé

Dans cet article, nous étudions le comportement asymptotique des trasformations aléatoires à n vertex. A titre d'application nous calculons les moments asymptotiques de la taille de la plus grande composante.

MSC: 05C80; 60J10

Keywords: Random mapping graphs; Connection; Component; Scheme of allocating particles; Markov chain; Asymptotic behavior

1. Introduction

Let V be a set with n elements, say, $V = \{1, 2, \ldots, n\}$, and Ω_n the set of all mappings on V, which includes n^n elements. Let $\mathbf{P}_n$ be the classical probability on Ω_n. Any $f \in \Omega_n$ induces a directed graph G_f with the set of vertices $V(G_f) = V$ and edges $E(G_f) = \{(u, f(u)):\ u \in V\}$. The space $(\Omega_n, \mathbf{P}_n)$ is called the space of random mappings or random mapping graphs.

The most interesting problems on random mappings are their various asymptotic behavior, by which we means the limit distribution of various graph structures, for example, number of components, size of component, etc., as n goes to infinity. Most of papers on this field focused on these issues. In this paper we shall start with a direct computation of asymptotic connection probability and use it to construct a scheme of allocating particles and a related Markov chain.

* Corresponding author.
E-mail address: jgying@fudan.edu.cn (J. Ying).
1 The research of this author is supported in part by NSFC No. 10671036.

0246-0203/$ – see front matter
doi:10.1016/j.anihpb.2006.05.004

We prove that asymptotic behaviors of random mappings may be represented in the Markov chain in certain sense. Moreover we may use the limit theorems in theory of probability and powerful tools developed for martingales and Markov chains. As a main result we shall give an explicit answer for the asymptotic moments and distribution of the largest component size.

We should mention that in 1994, Aldous and Pitman [2], also see [6], proved that the uniform random mapping is asymptotically the Brownian bridge in some sense, by which, various limit distributions were obtained. Comparing to that, our approach is more direct and elementary, and involves less machinery.

Throughout the paper $\mathbf{N}$ denotes the set of natural numbers. For any $n \in \mathbf{N}$, $[n] = \{1, 2, \ldots, n\}$. For any set A, the notation $|A|$ denotes the cardinality of A. A partition $\{A_1, \ldots, A_m\}$ of a finite set $U \subset \mathbf{N}$ is said to be ordered if $\min A_1 < \min A_2 < \cdots < \min A_m$. The partition involved in this paper will always be ordered. The probability and expectation are taken in the probability space of random mappings with n elements are written as $\mathbf{P}_n, \mathbf{E}_n$. The symbol ':=' should be read as 'is defined to be'. We shall briefly introduce our strategy as follows. The key is to relate this asymptotics with that of a particular Chinese restaurant process. More precisely let $(\Omega_n, \mathbf{P}_n)$ be the probability space of random mappings on $[n]$ and G_n a sample. For $N \leqslant n$, let $A_1, \ldots, A_m$ be an ordered partition of $[N]$. Denote by $J_{A_1} \oplus \cdots \oplus J_{A_m}$ the event that $i, j \in [N]$ are connected if and only if they are in the same set A_l. In other words, $J_{A_1} \oplus \cdots \oplus J_{A_m}$ is the event that the trace of the connected components of G_n on $[N]$ is the partition $\{A_1, \ldots, A_m\}$ (let us call this the N-trace of G_n). Now let $(Z_k)_{k\geqslant 1}$ be the $(0, 1/2)$-Chinese restaurant process and $\mathbf{P}$ its probability (Z_k is the label number of the table where the k-th customer is seated). Set

$$\phi_{A_1,\ldots,A_m}(\cdot) := \sum_{l=1}^{m} l \cdot 1_{A_l}.$$

The asymptotic connection of two models is the following (see Theorem 3.1, though it is actually proved in Section 2)

$$\lim_n \mathbf{P}_n(J_{A_1} \oplus \cdots \oplus J_{A_m}) = \mathbf{P}(Z_k = \phi_{A_1,\ldots,A_m}(k), \quad k = 1, \ldots, N).$$

(Roughly speaking, the asymptotic distribution of the N-trace of G_n is given by the distribution of $Z_1, \ldots, Z_N$.) We then show in Theorem 3.2 that the asymptotic ($n \to \infty$) "difference" between the cardinalities of the N-trace and those of the connected components of G_n vanishes as N goes to infinity. By this way we obtain a connection between the asymptotic of the cardinalities of the connected components of G_n and the process (Z_k). As a consequence, we give examples in Section 3 to show how to derive from this approach some known results as in Pittel [7], Stepanov [11], Aldous and Pitman [2], etc. However the main application is Theorem 4.1 in which the asymptotic of all moments of the r-th largest connected component of G_n is given.

2. The limit probability of connection

Two vertices i and j are connected if there is a path of edges, ignoring direction, connecting them. This naturally induces a classification for each graph. Let U be a fixed subset of $[n]$ and $\{A_1, A_2, \ldots, A_m\}$ a fixed partition of U. Let $J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}$ be the event that for any $i, j \in U$ with $i \neq j$, i and j are connected if and only if $i, j \in A_l$ for some $1 \leqslant l \leqslant m$. Set $J_m^d := J_{\{1\}} \oplus J_{\{2\}} \oplus \cdots \oplus J_{\{m\}}$. Let $Y_i = Y_i(G_n)$ be the number of the vertices connected with vertex i for any $i \in V$, and 1_α be the indicator of α for any set α. In order to write with more convenience, we also use the notation $a_l := |A_l| - 1 \geqslant 0$ for each $l, 1 \leqslant l \leqslant m$ and set $a := a_1 + a_2 + \cdots + a_m \leqslant n - m$; and $(x)_r := x(x-1)\cdots(x-r+1)$.

Lemma 2.1.

$$\mathbf{P}_n(J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}) = \frac{1}{(n-m)_a} \mathbf{E}_n\left(\prod_{l=1}^{m} (Y_l)_{a_l}; J_m^d\right).$$

Proof. Let $\mathcal{S} = \{J_{S_1} \oplus J_{S_2} \oplus \cdots \oplus J_{S_m}\colon S_1, S_2, \ldots, S_m$ are mutually disjoint subsets of V and for $1 \leqslant l \leqslant m, l \in S_l$ and $|S_l| = |A_l|\}$, a set of events. For any event $\alpha \in \mathcal{S}$,

$$\mathbf{E}_n(1_\alpha) = \mathbf{P}_n(1_\alpha = 1) = \mathbf{P}_n\{\alpha\} = \mathbf{P}_n(J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}).$$

Hence

$$\mathbf{P}_n(J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}) = \frac{\sum_{\alpha\in\mathcal{S}} \mathbf{E}_n(1_\alpha)}{|\mathcal{S}|} = \frac{\mathbf{E}_n(\sum_{\alpha\in\mathcal{S}} 1_\alpha)}{|\mathcal{S}|}.$$

On the other hand, a direct calculation gives that

$$|\mathcal{S}| = \frac{(n-m)_a}{a_1!a_2!\cdots a_m!}$$

and

$$\sum_{\alpha\in\mathcal{S}} 1_\alpha = \begin{cases} \prod_{l=1}^m \frac{(Y_l)_{a_l}}{a_l!}, & \text{if } J_m^d \text{ occurs,} \\ 0, & \text{otherwise.} \end{cases}$$

Replacing them into the formula for $\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\}$, the conclusion follows. □

Refer to pages 129–137, [9] for the following lemma.

Lemma 2.2.

$$\mathbf{P}_n\{G_n \textit{ is connected}\} = \frac{(n-1)!}{n^n}\sum_{j=0}^{n-1}\frac{n^j}{j!} \sim \sqrt{\frac{\pi}{2n}}.$$

Lemma 2.3. *For each* $1 \leqslant l \leqslant m$, *let* $k_l \geqslant 0$, *with* $k := k_1 + k_2 + \cdots + k_m \leqslant n - m$. *Then*

$$\mathbf{P}_n\big(Y_1 = k_1, \ldots, Y_m = k_m; J_m^d\big) = \frac{(n-m)!}{n^n}\frac{(n-k-m)^{n-k-m}}{(n-k-m)!}\prod_{l=1}^m\sum_{j=0}^{k_l}\frac{(k_l+1)^j}{j!}.$$

Proof. Set $\mathcal{S} := \{J_{S_1} \oplus J_{S_2} \oplus \cdots \oplus J_{S_m}\colon S_1, \ldots, S_m$ are mutually disjoint and $l \in S_l \subset V$, $|S_l| = k_l + 1$, for any $1 \leqslant l \leqslant m\}$, and for $\alpha = J_{S_1} \oplus J_{S_2} \oplus \cdots \oplus J_{S_m} \in \mathcal{S}$, $1 \leqslant l \leqslant m$ we set $l \in \alpha_l = S_l$. And that $\alpha = J_{\alpha_1} \oplus J_{\alpha_2} \oplus \cdots \oplus J_{\alpha_m}$ for $\alpha \in \mathcal{S}$. For $\Lambda \subset V$ we set D_Λ be the event that none of the vertices in Λ is connected with any vertex in $V\backslash\Lambda$. Then

$$\begin{aligned}
\mathbf{P}_n(Y_1 = k_1, \ldots, Y_m = k_m; J_m^d) &= \sum_{\alpha\in\mathcal{S}} \mathbf{P}_n(\alpha; Y_1 = k_1, \ldots, Y_m = k_m) \\
&= \sum_{\alpha\in\mathcal{S}} \mathbf{P}_n(J_{\alpha_1} \oplus J_{\alpha_2} \oplus \cdots \oplus J_{\alpha_m}; Y_1 = k_1, \ldots, Y_m = k_m) \\
&= \sum_{\alpha\in\mathcal{S}} \mathbf{P}_n(J_{\alpha_1} \oplus J_{\alpha_2} \oplus \cdots \oplus J_{\alpha_m}; D_{\alpha_1}, D_{\alpha_2}, \ldots, D_{\alpha_m}) \\
&= \sum_{\alpha\in\mathcal{S}} \mathbf{P}_n(D_{\alpha_1}D_{\alpha_2}\cdots D_{\alpha_m})\mathbf{P}_n(J_{\alpha_1} \oplus J_{\alpha_2} \oplus \cdots \oplus J_{\alpha_m} \mid D_{\alpha_1}, D_{\alpha_2}, \ldots, D_{\alpha_m}) \\
&= \sum_{\alpha\in\mathcal{S}} \mathbf{P}_n(D_{\alpha_1}D_{\alpha_2}\cdots D_{\alpha_m})\prod_{l=1}^m \mathbf{P}_n(J_{\alpha_l}|D_{\alpha_l}).
\end{aligned}$$

By the definition of $\mathcal{S}$ and α, we know that $k_l = |\alpha_l| - 1$ for each l, $1 \leqslant l \leqslant m$. It is easy to see that

$$|\mathcal{S}| = \frac{(n-m)!}{k_1!\cdots k_m!(n-m-k)!}, \qquad \mathbf{P}_n(D_{\alpha_1}D_{\alpha_2}\cdots D_{\alpha_m}) = \frac{(n-k-m)^{n-k-m}\prod_{l=1}^m (k_l+1)^{k_l+1}}{n^n}.$$

Employing Lemma 2.2 we know that for each l, $1 \leqslant l \leqslant m$ it holds that

$$\mathbf{P}_n(J_{\alpha_l}|D_{\alpha_l}) = \frac{k_l!}{(k_l+1)^{k_l+1}}\sum_{j=0}^{k_l}\frac{(k_l+1)^j}{j!}.$$

Substituting the $m+2$ equations into $\mathbf{P}_n(Y_1 = k_1, \ldots, Y_m = k_m; J_m^d)$, we shall have the conclusion directly. □

Corollary 2.1. *Let $k_l \geqslant 0$, $1 \leqslant l \leqslant m$. If $k := k_1 + k_2 + \cdots + k_m < n - m$, then*

$$\mathbf{P}_n(Y_1 = k_1, \ldots, Y_m = k_m;\, J_m^d) = \frac{1}{2^m n^{m-1/2}(n-k-m)^{1/2}}\big(1 - \Delta_{k_1,\ldots,k_m} + \Delta^2_{n,k_1,\ldots,k_m}\big);$$

and if $k = n - m$, then

$$\mathbf{P}_n\big(Y_1 = k_1, \ldots, Y_m = k_m;\, J_m^d\big) = \frac{\sqrt{2\pi}}{2^m n^{m-1/2}}\big(1 - \Delta_{k_1,\ldots,k_m} + \Delta^2_{n,k_1,\ldots,k_m}\big),$$

where

$$\Delta_{k_1,\ldots,k_m} := \sum_{l=1}^m \frac{2}{3\sqrt{2\pi(k_l+1)}} \quad \textit{and} \quad \Delta^2_{n,k_1,\ldots,k_m} := \frac{C_{-1}}{n-m-k} + \frac{C_0}{n} + \sum_{l=1}^m \frac{C_i}{k_l+1},$$

with $C_{-1}, C_0, C_i, \ldots, C_m$ being bounded functions which depend on n, m and $k_1, \ldots, k_m$. Note that we set $\frac{C_{-1}}{n-m-k} = 0$ if $k = n - m$.

Proof. We prove for the case $k < n - m$ only. Employing the Stirling's formula $r! = (r/\mathrm{e})^r \sqrt{2\pi r}\, \mathrm{e}^{\theta_r}$, where $\frac{1}{12r+1} < \theta_r < \frac{1}{12r}$, we get that

$$\frac{(n-m)!}{n^n}\frac{(n-k-m)^{n-k-m}}{(n-k-m)!} = \frac{n^{1/2}\mathrm{e}^{-k-m}}{n^m(n-k-m)^{1/2}}\mathrm{e}^{\theta_n - \theta_{n-k-m}} \prod_{i=0}^{m-1}\left(1 - \frac{i}{n}\right)^{-1}.$$

Furthermore, we have the Ramanujan sequence [8]

$$\frac{1}{2}\mathrm{e}^r = 1 + \frac{r}{1!} + \frac{r^2}{2!} + \cdots + \frac{r^{r-1}}{(r-1)!} + \gamma_r \frac{r^r}{r!}$$

where γ_r is decreasing in r and of the form $\gamma_r = \frac{1}{3} + \frac{4}{135(r+\eta_r)}$ with $\frac{2}{21} \leqslant \eta_r \leqslant \frac{8}{45}$. Then

$$\prod_{l=1}^m \sum_{j=0}^{k_l} \frac{(k_l+1)^j}{j!} = \prod_{l=1}^m \left(\frac{\mathrm{e}^{k_l+1}}{2} - \gamma_{k_l+1}\frac{(k_l+1)^{k_l+1}}{(k_l+1)!}\right) = \frac{\mathrm{e}^{k+m}}{2^m}\prod_{l=1}^m\left(1 - \frac{2\gamma_{k_l+1}}{\sqrt{2\pi(k_l+1)}\mathrm{e}^{\theta_{k_l+1}}}\right).$$

After some simple but lengthy calculation, it follows that

$$\mathrm{e}^{\theta_n - \theta_{n-k-m}} \prod_{i=0}^{m-1}\left(1 - \frac{i}{n}\right)^{-1} \prod_{l=1}^m\left(1 - \frac{2\gamma_{k_l+1}}{\sqrt{2\pi(k_l+1)}\,\mathrm{e}^{\theta_{k_l+1}}}\right) = 1 - \Delta_{k_1,\ldots,k_m} + \Delta^2_{n,k_1,\ldots,k_m}$$

where $\Delta_{k_1,\ldots,k_m}$ and $\Delta^2_{n,k_1,\ldots,k_m}$ are defined as above. Therefore, substituting the three equations above into Lemma 2.3, we shall have

$$\mathbf{P}_n\big(Y_1 = k_1, \ldots, Y_m = k_m;\, J_m^d\big) = \frac{1}{2^m n^{m-1/2}(n-k-m)^{1/2}}\big(1 - \Delta_{k_1,\ldots,k_m} + \Delta^2_{n,k_1,\ldots,k_m}\big). \qquad \square$$

The following two theorems give an explicit expression for connection probability and its exact asymptotic behavior.

Theorem 2.1. *Let $A_1, A_2, \ldots, A_m$ be a partition of a fixed set $U \subset V$. Then*

$$\mathbf{P}_n(J_{A_1} \oplus \cdots \oplus J_{A_m}) = \frac{(n-m-a)!}{n^n}\sum_{k=0}^{n-m}\sum_{k_l + \cdots + k_m = k}\frac{(n-k-m)^{n-k-m}}{(n-k-m)!}\prod_{l=1}^m (k_l)_{a_l}\sum_{j=0}^{k_l}\frac{(k_l+1)^j}{j!}.$$

Proof. It is a direct consequence of Lemmas 2.1 and 2.3. $\square$

Theorem 2.2. *Let $A_1, A_2, \ldots, A_m$ be a partition of a fixed set $U \subset V$. Then*

$$\begin{aligned}&\mathbf{P}_n(J_{A_1} \oplus \cdots \oplus J_{A_m})\\&\quad= \frac{(2a_1)!!\cdots(2a_m)!!}{(2m+2a-1)!!} + \frac{(2a_1)!!\cdots(2a_m)!!}{(2m+2a-2)!!}\left(1-\sum_{i=1}^m \frac{(2a_i-1)!!}{(2a_i)!!}\right)\frac{\sqrt{2\pi}}{6}\,n^{-1/2} + \mathrm{o}\big(n^{-1/2}\big),\end{aligned}$$

and

$$\begin{aligned}&\mathbf{P}_n(J_{A_1} \oplus \cdots \oplus J_{A_m}) - \mathbf{P}_{n+1}(J_{A_1} \oplus \cdots \oplus J_{A_m})\\&\quad= \frac{(2a_1)!!\cdots(2a_m)!!}{(2m+2a-2)!!}\left(1-\sum_{i=1}^m \frac{(2a_i-1)!!}{(2a_i)!!}\right)\frac{\sqrt{2\pi}}{12}\,n^{-3/2} + \mathrm{o}\big(n^{-3/2}\big).\end{aligned}$$

as n goes to infinity. Note that we set $(-1)!! = 0!! = 1$.

Proof. The case $m = 1,\ a_1 = 0$ is trivial. We shall prove for case I: $m = 1, a_1 = 1$ and case II: $m = 2, a_1 = a_2 = 0$. The result for the rest cases can be proved by induction to m and the size of $A_1, A_2, \ldots, A_m$, and we omit the procedure which is something like the procedure that we shall do for case I and case II.

Case I: $m = 1, a_1 = 1$. Without loss of generality, we set $A_1 = \{1, 2\}$. What we want to prove is that

$$\mathbf{P}_n\{J_{\{1,2\}}\} = \frac{2}{3} + \frac{\sqrt{2\pi}}{12}\,n^{-1/2} + \mathrm{o}\big(n^{-1/2}\big),$$

$$\mathbf{P}_n\{J_{\{1,2\}}\} - \mathbf{P}_{n+1}\{J_{\{1,2\}}\} = \frac{\sqrt{2\pi}}{24}\,n^{-3/2} + \mathrm{o}\big(n^{-3/2}\big).$$

Case II: $m = 2$, $a_1 = a_2 = 0$. Without lost of generality, we set $A_1 = \{1\}$ and $A_2 = \{2\}$. What we want to prove is that

$$\mathbf{P}_n\{J_{\{1\}} \oplus J_{\{2\}}\} = \frac{1}{3} - \frac{\sqrt{2\pi}}{12}\,n^{-1/2} + \mathrm{o}\big(n^{-1/2}\big),$$

and

$$\mathbf{P}_n\{J_{\{1\}} \oplus J_{\{2\}}\} - \mathbf{P}_{n+1}\big\{J_{\{1\}} \oplus J_{\{2\}}\big\} = -\frac{\sqrt{2\pi}}{24}\,n^{-3/2} + \mathrm{o}\big(n^{-3/2}\big).$$

We prove for case I in three steps. At first we point out that $\mathbf{P}_n\{J_{\{1,2\}}\}$ converges to $\frac{2}{3}$. Secondly we prove the existence of limit $T = \lim_{n\to\infty}\sqrt{n}\,(\mathbf{P}_n\{J_{\{1,2\}}\} - 2/3)$ by induction on the base of $\lim_{n\to\infty}\sqrt{n}\,(\mathbf{P}_n\{J_{\{1\}}\} - 1) = 0$. At last we find out T, where we use the fact $\lim_{n\to\infty}\sqrt{n}\,(\mathbf{P}_n\{J_{\{1\}}\} - 1) = 0$ again, and prove the monotonicity of $\mathbf{P}_n\{J_{\{1,2\}}\}$ at the same time. We prove for case II straightforward by using the previous results above for $\mathbf{P}_n\{J_{\{1,2\}}\}$. (It may be seen as being proved by induction, too.)

Step 1. Using Theorem 2.1, we get

$$\mathbf{P}_n\{J_{\{1,2\}}\} = \frac{\mathbf{E}_n(Y_1)}{n-1} = \frac{(n-2)!}{n^n}\sum_{k=1}^{n-1}\frac{k(n-k-1)^{n-k-1}}{(n-k-1)!}\sum_{j=0}^{k}\frac{(k+1)^j}{j!}.$$

Set ${}_n\Lambda_k$, for each n, k, $n = 2, 3, \ldots$ and $k = 0, 1, 2, \ldots, n-1$,

$${}_n\Lambda_k = \frac{k}{n-1}\mathbf{P}_n(Y_1 = k) = \frac{(n-2)!}{n^n}\frac{k(n-k-1)^{n-k-1}}{(n-k-1)!}\sum_{j=0}^{k}\frac{(k+1)^j}{j!}.$$

Then $\mathbf{P}_n\{J_{\{1,2\}}\} = \sum_{k=1}^{n-1}{}_n\Lambda_k$. Employing Corollary 2.1, we get, for $0 \leqslant k \leqslant n-2$

$${}_n\Lambda_k = \frac{1}{2\sqrt{n}(n-1)}\frac{k}{\sqrt{n-k-1}}\left(1 - \frac{2}{3\sqrt{2\pi(k+1)}} + \frac{C_{-1}}{n-k-1} + \frac{C_0}{n} + \frac{C_1}{k+1}\right).$$

Therefore,

$$\mathbf{P}_n\{J_{\{1,2\}}\} = P(Y_1 = n-1) + \sum_{k=0}^{n-2} {}_n\Lambda_k \sim \sum_{k=0}^{n-2} \frac{1}{2\sqrt{n}(n-1)} \frac{k}{\sqrt{n-k-1}} \sim \frac{1}{2}\int_0^1 \frac{x}{\sqrt{1-x}}\,\mathrm{d}x = \frac{2}{3}.$$

Step 2. As it is known that $\mathbf{P}_n\{J_{\{1\}}\} = 1$, we get

$$(n-1)\sum_{k=1}^{n-1} \frac{{}_n\Lambda_k}{k} = \sum_{k=1}^{n-1} \mathbf{P}_n(Y_1 = k) = \mathbf{P}_n\{J_{\{1\}}\} - \mathbf{P}_n(Y_1 = 0) = 1 + \mathrm{O}\big(n^{-1}\big).$$

We shall separate $\mathbf{P}_n\{J_{\{1,2\}}\}$ into several parts.

$$\begin{aligned}
\mathbf{P}_n\{J_{\{1,2\}}\} &= \sum_{k=1}^{n-1} {}_n\Lambda_k = \sum_{k=1}^{n-1} \frac{n-1}{k}\,{}_n\Lambda_k - \sum_{k=1}^{n-1} \frac{n-k-1}{k}\,{}_n\Lambda_k \\
&= 1 - \sum_{k=1}^{n-1} \frac{\sqrt{n-k-1}}{2\sqrt{n}(n-1)} \left(1 - \frac{2}{3\sqrt{2\pi(k+1)}} + \frac{C_{-1}}{n-k-1} + \frac{C_0}{n} + \frac{C_1}{k+1}\right) + \mathrm{O}\big(n^{-1}\big) \\
&= 1 - \sum_{k=1}^{n-1} \frac{\sqrt{n-k-1}}{2\sqrt{n}(n-1)} + \sum_{k=1}^{n-1} \frac{\sqrt{n-k-1}}{2\sqrt{n}(n-1)} \frac{2}{3\sqrt{2\pi(k+1)}} + \mathrm{o}\big(n^{-1/2}\big) \\
&= 1 - \frac{1}{2}\int_0^1 \sqrt{x}\,\mathrm{d}x + \frac{1}{3\sqrt{2\pi}}\int_0^1 \sqrt{\frac{1-x}{x}}\,\mathrm{d}x \cdot n^{-1/2} + \mathrm{o}\big(n^{-1/2}\big) \\
&= \frac{2}{3} + \frac{\sqrt{2\pi}}{12} n^{-1/2} + \mathrm{o}\big(n^{-1/2}\big).
\end{aligned}$$

Hence $T = \lim_{n\to\infty} \sqrt{n}\,(\mathbf{P}_n\{J_{\{1,2\}}\} - 2/3)$ exists and $T = \frac{\sqrt{2\pi}}{12}$. However, it is not necessary to know the value of T. What we need only at present is to know its existence, and we will calculate its value by solving out a equation. Such method will do help when we prove the rest cases.

We compare ${}_{n+1}\Lambda_{k+1}$ with ${}_n\Lambda_k$, and get for each $1 \leqslant k \leqslant n-1$,

$$\begin{aligned}
\frac{{}_{n+1}\Lambda_{k+1}}{{}_n\Lambda_k} &= \frac{k+1}{k}\,\frac{n-1}{n}\,\frac{e\cdot n^{n+1}}{(n+1)^{n+1}}\,\frac{\sum_{j=0}^{k+1}\frac{(k+2)^j}{j!}}{\mathrm{e}\sum_{j=0}^{k}\frac{(k+1)^j}{j!}} \\
&= \left(1+\frac{1}{k}\right)\left(1 - \frac{3}{2n} + \mathrm{O}\left(\frac{1}{n^2}\right)\right)\left(1 + \frac{1}{3(2\pi)^{1/2}k^{3/2}} + \frac{C_k}{k^2}\right) \\
&= 1 + \frac{1}{k} - \frac{3}{2n} + \frac{1}{3(2\pi)^{1/2}k^{3/2}} + \frac{C_k}{k^2},
\end{aligned}$$

where C_k, which may vary from place to place, is a bounded function which depends on k. But it can be shown that

$$n^{3/2}\sum_{k=1}^{n-1} \frac{{}_n\Lambda_k}{k^{3/2}} \sim \sum_{k=1}^{n-2} \frac{1}{2\sqrt{k(n-k-1)}} \sim \frac{1}{2}\int_0^1 \frac{1}{\sqrt{x(1-x)}}\,\mathrm{d}x = \frac{\pi}{2}$$

and $\sum_{k=1}^{n-1} \frac{{}_n\Lambda_k}{k^2} = \mathrm{o}(n^{-3/2})$. Therefore,

$$\begin{aligned}
n^{3/2}\sum_{k=1}^{n-1}({}_{n+1}\Lambda_{k+1} - {}_n\Lambda_k) &= n^{3/2}\sum_{k=1}^{n-1}\frac{{}_n\Lambda_k}{k} - \frac{3n^{1/2}}{2}\sum_{k=1}^{n-1} {}_n\Lambda_k + \frac{n^{3/2}}{3(2\pi)^{1/2}}\sum_{k=1}^{n-1}\frac{{}_n\Lambda_k}{k^{3/2}} + \mathrm{o}(1) \\
&= n^{1/2} - \frac{3n^{1/2}}{2}\big(2/3 + T\cdot n^{-1/2}\big) + \frac{\sqrt{2\pi}}{12} + \mathrm{o}(1) \\
&= -\frac{3}{2}T + \frac{\sqrt{2\pi}}{12} + \mathrm{o}(1).
\end{aligned}$$

It follows that

$$n^{3/2}\big(\mathbf{P}_{n+1}\{J_{\{1,2\}}\} - \mathbf{P}_n\{J_{\{1,2\}}\}\big) = n^{3/2} \cdot {}_{n+1}\Lambda_1 + n^{3/2}\sum_{k=1}^{n-1}({}_{n+1}\Lambda_{k+1} - {}_n\Lambda_k) = -\frac{3}{2}T + \frac{\sqrt{2\pi}}{12} + \mathrm{o}(1).$$

On the other hand, we have

$$\lim_{n\to\infty}\sqrt{n}\,\big(\mathbf{P}_n\{J_{\{1,2\}}\} - 2/3\big) = 2\lim_{n\to\infty} n\sqrt{n}\,\big(\mathbf{P}_n\{J_{\{1,2\}}\} - \mathbf{P}_{n+1}\{J_{\{1,2\}}\}\big),$$

since the existence of $\lim_{n\to\infty} n\sqrt{n}(\mathbf{P}_n\{J_{\{1,2\}}\} - \mathbf{P}_{n+1}\{J_{\{1,2\}}\})$. Therefore,

$$-\frac{T}{2} = -\frac{3}{2}T + \frac{\sqrt{2\pi}}{12}$$

and we get $T = \sqrt{2\pi}/12$. For case II, it holds that $\mathbf{P}_n(J_{\{1\}} \oplus J_{\{2\}}) = \mathbf{P}_n(J_{\{1\}}) - \mathbf{P}_n(J_{\{1,2\}})$. Substituting the result of case I into this, we draw out the conclusion for case II. □

The following two results are the direct consequences of Theorem 2.2.

Corollary 2.2.

(1) *The probability that vertices in* $[m]$ *are totally disconnected*

$$\mathbf{P}_n\{J_m^d\} = \mathbf{P}_n\{J_{\{1\}} \oplus J_{\{2\}} \oplus \cdots \oplus J_{\{m\}}\} \sim \frac{1}{(2m-1)!!};$$

(2) *The probability that vertices in* $[m]$ *are connected*

$$\mathbf{P}_n\{J_{\{1,2,\ldots,m\}}\} \sim \frac{(2m-2)!!}{(2m-1)!!}.$$

Remark. At the end of this section, we would like to present a table of Monte Carlo simulation. In Table 1 we list some numerical values concerning probability $\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\}$ in some simple cases for comparing. 'precise' means 'precise value', 'revised' means 'revised value', 'test' means 'test value' and 'limit' means 'limit value'. The precise value is given by Theorem 2.1. The revised value is given by approximation

Table 1
Calculation of $\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\}$

$\mathbf{P}$	precise $n=10$	revised $n=10$	test $n=10$	precise $n=50$	revised $n=50$	precise $n=100$	revised $n=100$	limit $n\to\infty$
$J_{\{1,2\}}$	.7159	.7327	.7175	.6923	.6962	.6855	.6876	.6667
$J_{\{1,2,3\}}$	.5966	.6159	.5990	.5659	.5703	.5572	.5594	.5333
$J_{\{1,2,3,4\}}$	.5272	.5480	.5315	.4931	.4978	.4834	.4859	.4571
$J_{\{1,2,3,4,5\}}$	.4805	.5023	.4830	.4443	.4493	.4341	.4367	.4063
$J_{\{1\}} \oplus J_{\{2\}}$	.2841	.2673	.2825	.3077	.3038	.3145	.3124	.3333
$J_{\{1,2\}} \oplus J_{\{3\}}$	.1193	.1168	.1185	.1264	.1259	.1284	.1281	.1333
$J_{\{1,2\}} \oplus J_{\{3,4\}}$	.0365	–	.0430	.0376	–	.0378	–	.0381
$J_{\{1,2,3\}} \oplus J_{\{4\}}$	.0694	.0679	.0675	.0728	.0725	.0737	.0736	.0762
$J_{\{1,2,3\}} \oplus J_{\{4,5\}}$	.0166	.0176	.0145	.0170	.0172	.0170	.0171	.0169
$J_{\{1,2,3,4\}} \oplus J_{\{5\}}$	.0467	.0456	.0485	.0488	.0485	.0493	.0492	.0508
$J_{\{1\}} \oplus J_{\{2\}} \oplus J_{\{3\}}$	.0454	.0336	.0445	.0548	.0519	.0578	.0562	.0667
$J_{\{1,2\}} \oplus J_{\{3\}} \oplus J_{\{4\}}$	.0135	.0108	.0120	.0160	.0154	.0168	.0164	.0190
$J_{\{1,2\}} \oplus J_{\{3,4\}} \oplus J_{\{5\}}$	.0032	.0029	.0045	.0037	.0036	.0038	.0038	.0042
$J_{\{1,2,3\}} \oplus J_{\{4\}} \oplus J_{\{5\}}$	.0060	.0047	.0055	.0071	.0068	.0074	.0073	.0085
$J_{\{1\}} \oplus J_{\{2\}} \oplus J_{\{3\}} \oplus J_{\{4\}}$	.0050	.0013	.0060	.0068	.0058	.0074	.0069	.0095
$J_{\{1,2\}} \oplus J_{\{3\}} \oplus J_{\{4\}} \oplus J_{\{5\}}$	.0011	.0004	.0000	.0015	.0013	.0017	.0016	.0021
$J_{\{1\}} \oplus J_{\{2\}} \oplus J_{\{3\}} \oplus J_{\{4\}} \oplus J_{\{5\}}$	.0004	−.0003	.0010	.0006	.0004	.0007	.0006	.0011

$$\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\} \approx \frac{(2a_1)!! \cdots (2a_m)!!}{(2m+2a-1)!!} + \frac{(2a_1)!! \cdots (2a_m)!!}{(2m+2a-2)!!}\left(1 - \sum_{i=1}^{m} \frac{(2a_i-1)!!}{(2a_i)!!}\right)\frac{\sqrt{2\pi}}{6} n^{-1/2}.$$

The test value is given by the Monte Carlo simulation

$$\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\} \approx \frac{\text{times of the occurrence of } J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m} \text{ in } N \text{ times}}{N}$$

with $N = 2000$. The limit value is given by

$$\mathbf{P}_n\{J_{A_1} \oplus J_{A_2} \oplus \cdots \oplus J_{A_m}\} \to \frac{(2a_1)!! \cdots (2a_m)!!}{(2m+2a-1)!!}.$$

3. A related Markov chain and asymptotics of components

We shall present the essential relation between large components of a random mapping and large boxes of the related scheme of allocation in this section. We first introduce a probability space which describes a scheme of allocating particles and discuss interesting properties of some related random sequences. By Theorem 2.2 the following lemma is direct.

Lemma 3.1. *Assume that $A_1, \ldots, A_m$ is a fixed partition of $[N]$. Then*

$$\lim_{n\to\infty} \mathbf{P}_n\{J_{A_1} \oplus \cdots \oplus J_{A_m} \oplus J_{\{N+1\}} | J_{A_1} \oplus \cdots \oplus J_{A_m}\} = \frac{1}{2N+1}$$

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

$$\lim_{n\to\infty} \mathbf{P}_n\{J_{A_1} \oplus \cdots \oplus J_{A_{l-1}} \oplus J_{A_l \cup \{N+1\}} \oplus J_{A_{l+1}} \oplus \cdots \oplus J_{A_m} | J_{A_1} \oplus \cdots \oplus J_{A_m}\} = \frac{2|A_l|}{2N+1}.$$

The lemma inspires us to consider a scheme of random allocating particles. There are different particles called $P_1, P_2, \ldots$ and different boxes called $B_1, B_2, \ldots$. The P_n and B_n are also called the n-th particle and n-th box. We define the allocation by induction. First we place P_1 into B_1, and next we place P_2 with probability $\frac{2}{3}$ into B_1 and with probability $\frac{1}{3}$ into B_2. More generally, suppose that the first N particles have been placed. Let us place the next particle. Let the first m boxes be non-empty and have $q_1, \ldots, q_m$ particles in boxes from 1 to m respectively. Then $N = q_1 + \cdots + q_m$. At the next time we place P_{N+1} into the B_l with probability $\frac{2q_l}{2N+1}$ for $l = 1, \ldots, m$ and into B_{m+1} with probability $\frac{1}{2N+1}$. More precisely we have a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a sequence of random variables $\{Z_n\}$, which is defined by induction as follows

(1) $Z_1 = 1$;
(2) Suppose that $Z_1, \ldots, Z_N$ is defined for $N \geqslant 1$. Let $m := \max\{Z_1, \ldots, Z_N\}$ and $q_i := \sum_{k=1}^{N} 1_{\{Z_k = i\}}$. Z_{N+1} is equal to i with probability $\frac{2q_i}{2N+1}$ for $1 \leqslant i \leqslant m$ and to $m+1$ with probability $\frac{1}{2N+1}$, or in the form of conditional expectation

$$\mathbf{P}(Z_{N+1} = i | Z_1, \ldots, Z_N) = \frac{2q_i}{2N+1} 1_{\{1 \leqslant i \leqslant m\}} + \frac{1}{2N+1} 1_{\{i = m+1\}}.$$

The random variable Z_N records the number labelled on the box where the N-th particle is placed. The property (2) actually gives the conditional distribution of Z_{N+1} given $\{Z_1, \ldots, Z_N\}$ which determines the law $\mathbf{P}$. Let $\mathcal{H}_N := \sigma(Z_1, \ldots, Z_N)$ the filtration of $\{Z_N\}$ and $\mathcal{H}_\infty = \sigma(\mathcal{H}_N\colon N \geqslant 1)$. Assume again that $A_1, \ldots, A_m$ is a fixed ordered partition of $[N]$. We denote by $A_1 \oplus A_2 \oplus \cdots \oplus A_m$ the event that for any $i, j \in [N]$, P_i and P_j are contained in the same box if and only if $i, j \in A_l$ for some $1 \leqslant l \leqslant m$. Actually the event $A_1 \oplus \cdots \oplus A_m$ determines the value of $Z_1, \ldots, Z_N$ and vice versa. Indeed if $Z_1, \ldots, Z_n$ are given, then the partition is natural, and conversely if an ordered partition is given as above, $Z_k = l$ if $k \in A_l$. Then

$$\mathcal{H}_N = \sigma\big(A_1 \oplus \cdots \oplus A_m\colon \{A_1, \ldots, A_m\} \text{ is a partition of } [N]\big)$$

and $A_1 \oplus \cdots \oplus A_m$ is an atom of $\mathcal{H}_N$.

This process is actually a particular case of so-called (α,θ)-Chinese restaurant process (see, e.g. Chapter 3, [6]) with $\alpha=0$ and $\theta=1/2$. By a direct calculation, it follows that

$$\mathbf{P}\{A_1\oplus A_2\oplus\cdots\oplus A_m\}=\frac{(2a_1)!!\cdots(2a_m)!!}{(2N-1)!!}.$$

Therefore we have the following theorem which states that asymptotically the probability that vertices are connected in a random mapping is equal to the probability that particles are placed in the same box in the above scheme of allocating particles. This illustrates the essential connection between these two models.

Theorem 3.1. *For any N and if $A_1,\dots,A_m$ is a partition of $[N]$, then*

$$\lim_{n\to\infty}\mathbf{P}_n\{J_{A_1}\oplus\cdots\oplus J_{A_m}\}=\mathbf{P}\{A_1\oplus\cdots\oplus A_m\}.$$

Remark. For $N\leqslant n$, define similarly for random mapping graph

$$\mathcal{H}_{n,N}:=\sigma\big(J_{A_1}\oplus\cdots\oplus J_{A_m}\colon\ \{A_1,\dots,A_m\}\text{ is a partition of }[N]\big).$$

It is actually generated by components in $[N]$. An event K in $\mathcal{H}_{n,N}$ may be viewed as the corresponding event K in $\mathcal{H}_N$ in a natural way. The theorem says $\lim_n\mathbf{P}_n(K)=\mathbf{P}(K)$.

For $N,m\geqslant 1$, let D_N be the number of nonempty boxes after the N-th particle has been placed, T_m the first time that the box B_m is nonempty and Q_N^m the number of the particles in box B_m at the time when the N-th particle is placed. Precisely

$$\begin{aligned}
&D_N:=\max\{Z_1,\dots,Z_N\},\\
&Q_N^m:=\sum_{k=1}^N 1_{\{Z_k=m\}},\\
&T_m:=\inf\{k\colon\ Z_k=m\}.
\end{aligned}$$

Set $Q_N=(Q_N^m\colon\ m\geqslant 1)$. Then D_N is also equal to the length of non-zero elements in Q_N or simply the length of Q_N. Clearly $Q_1,\dots,Q_N$ and $Z_1,\dots,Z_N$ are uniquely determined mutually for any $N\geqslant 1$. Then $\{\mathcal{H}_N\}$ is also the filtration of (Q_N). Let S be the set of sequences of non-negative integers $\mathbf{x}=(x_n)$ satisfying that there exists $m\geqslant 1$ such that $x_n=0$ for $n>m$ and $x_n>0$ for $n\leqslant m$ with norm $|\mathbf{x}|=\sum_n x_n$. Particularly $\mathbf{e}_i:=(1_{\{n=i\}}\colon\ n\geqslant 1)$ the i-th unit vector. The following lemma can be checked by definition directly.

Lemma 3.2. *$\{Q_N\colon\ N\geqslant 1\}$ is a Markov chain starting from $\mathbf{e}_1$, with state space S. More precisely for $i=1,2,\dots,$*

$$\mathbf{P}(Q_{N+1}=Q_N+\mathbf{e}_i|\mathcal{H}_N)=\frac{2Q_N^i}{2|Q_N|+1}+\frac{1}{2|Q_N|+1}1_{\{Q_N^{i-1}>0,\,Q_N^i=0\}}.$$

For the length of Q_N, by the definition of the random allocating particles process, we have

$$\mathbf{P}(D_{N+1}=D_N+1|\mathcal{H}_N)=\frac{1}{2N+1},\quad\text{and}\quad \mathbf{P}(D_{N+1}=D_N|\mathcal{H}_N)=\frac{2N}{2N+1}.$$

Hence, $\{D_N\}$ is a Markov chain with independent (but not stationary) increments with respect to $\{\mathcal{H}_N\}$.

From the definition of T_m, it is actually the first hitting time to $\{m\}$ of $\{Z_n\}$ and hence we know that $\{T_m\}$ is a sequence of stopping times with respect to $\{\mathcal{H}_N\}$. Obviously $1=T_1<T_2<\cdots<T_m<\cdots$. As we shall see later, $\{T_m\}$ takes finite value a.s. From the definition of Q_N^m, we know, on $\{T_m<\infty\}$

$$\begin{aligned}
&\mathbf{P}\big(Q^m_{T_m+N}=Q^m_{T_m+N-1}+1|\mathcal{H}_{T_m+N-1}\big)=\frac{2Q^m_{T_m+N-1}}{2T_m+2N-1},\\
&\mathbf{P}\big(Q^m_{T_m+N}=Q^m_{T_m+N-1}|\mathcal{H}_{T_m+N-1}\big)=1-\frac{2Q^m_{T_m+N-1}}{2T_m+2N-1}.
\end{aligned}$$

Thus $\{Q^m_{T_m+N}\}$ is a Markov chain with respect to $\{\mathcal{H}_{T_m+N}\}$ for any $m\geqslant 1$. In particular, when $m=1$, $\{Q_N^1\colon\ N\geqslant 1\}$ is a Markov chain with respect to $\{\mathcal{H}_N\}$. We now present a few martingale properties related to this chain.

Lemma 3.3.

(1) *For any* $m \geqslant 1$, T_m *is finite a.s., and for* $N = 1, 2, \ldots$ *we have that*

$$\mathbf{P}(T_{m+1} = T_m + N | \mathcal{H}_{T_m}) = \frac{B(T_m + N - 1, \frac{3}{2})}{B(T_m, \frac{1}{2})}$$

where $B(\cdot, \cdot)$ *is the beta function.*

(2) *Moreover,* $\{\frac{3^m}{2T_m+1}\}$ *is a martingale with mean* 1.

(3) *For each* $m \geqslant 1$, $\{\frac{Q^m_{T_m+N}}{2T_m+2N+1}\colon N = 0, 1, \ldots\}$ *is a bounded martingale with respect to* $\{\mathcal{H}_{T_m+N}\}$. *Moreover*

$$\mathbf{E}\left\{\frac{Q^m_{T_m+N}}{2T_m + 2N + 1}\Big|\mathcal{H}_{T_m}\right\} = \frac{1}{2T_m + 1}, \quad \textit{and} \quad \mathbf{E}\left\{\frac{Q^m_{T_m+N}}{2T_m + 2N + 1}\right\} = \frac{1}{3^m}.$$

(4) $\{\frac{Q^m_N}{2N+1}\}$ *is a bounded sub-martingale with respect to* $\{\mathcal{H}_N\}$, *and*

$$\mathbf{E}\left\{\frac{Q^m_N}{2N + 1}\right\} \leqslant \frac{1}{3^m}.$$

Proof. (1) Using the strong Markov property, we get on $\{T_m < \infty\}$ that $D_{T_m} = m$ and

$$\begin{aligned}
\mathbf{P}(T_{m+1} = T_m + N|\mathcal{H}_{T_m}) &= \mathbf{E}\big(\mathbf{P}(D_{T_m+N} = m + 1, D_{T_m+N-1} = m|\mathcal{H}_{T_m+N-1})|\mathcal{H}_{T_m}\big) \\
&= \mathbf{E}\big(\mathbf{P}(D_{T_m+N} = D_{T_m+N-1} + 1|\mathcal{H}_{T_m+N-1}), D_{T_m+N-1} = m|\mathcal{H}_{T_m}\big) \\
&= \frac{1}{2T_m + 2N - 1}\mathbf{P}(D_{T_m+N-1} = m|\mathcal{H}_{T_m}) \\
&= \frac{1}{2T_m + 2N - 1}\mathbf{E}\big(\mathbf{P}(D_{T_m+N-1} = D_{T_m+N-2}|\mathcal{H}_{T_m+N-2}); D_{T_m+N-2} = m|\mathcal{H}_{T_m}\big) \\
&= \frac{1}{2T_m + 2N - 1} \cdot \frac{2T_m + 2N - 4}{2T_m + 2N - 3}\mathbf{P}(D_{T_m+N-2} = m|\mathcal{H}_{T_m}) \\
&= \frac{1}{2T_m + 2N - 1} \cdot \frac{2T_m + 2N - 4}{2T_m + 2N - 3} \cdots \frac{2T_m}{2T_m + 1} \\
&= \frac{B(T_m + N - 1, \frac{3}{2})}{B(T_m, \frac{1}{2})},
\end{aligned}$$

and then on $\{T_m < \infty\}$,

$$\begin{aligned}
\mathbf{P}(T_{m+1} < \infty|\mathcal{H}_{T_m}) &= \sum_{N=1}^{\infty} \mathbf{P}(T_{m+1} = T_m + N|\mathcal{H}_{T_m}) \\
&= \sum_{N=1}^{\infty} \frac{B(T_m + N - 1, \frac{3}{2})}{B(T_m, \frac{1}{2})} \\
&= \frac{1}{B(T_m, \frac{1}{2})} \sum_{N=1}^{\infty} \int_0^1 x^{T_m+N-2}(1 - x)^{\frac{1}{2}}\, \mathrm{d}x = 1.
\end{aligned}$$

Hence, $\mathbf{P}\{T_{m+1} < \infty|T_m < \infty\} = 1$ and it follows from $T_1 = 1$ that any T_m is finite a.s.

(2) Calculating the conditional expectation of $\frac{3^{m+1}}{2T_{m+1}+1}$ given $\mathcal{H}_{T_m}$ by using (1), we have

$$\mathbf{E}\left(\frac{3^{m+1}}{2T_{m+1} + 1}\Big|\mathcal{H}_{T_m}\right) = \sum_{N=1}^{\infty} \frac{3^{m+1}}{2T_m + 2N + 1} \frac{B(T_m + N - 1, \frac{3}{2})}{B(T_m, \frac{1}{2})} = \frac{3^m}{2T_m + 1}$$

and

$$\mathbf{E}\left(\frac{3^{m+1}}{2T_{m+1}+1}\right)=\frac{3}{2T_1+1}=1.$$

(3) The boundedness is obvious since $Q^m_{T_m+N}\leqslant N+1$. Using strong Markov property again, we have

$$\begin{aligned}\mathbf{E}\left\{Q^m_{T_m+N+1}|\mathcal{H}_{T_m+N}\right\}&=Q^m_{T_m+N}\mathbf{P}\left\{Q^m_{T_m+N+1}=Q^m_{T_m+N}|\mathcal{H}_{T_m+N}\right\}\\&\quad+\left(Q^m_{T_m+N}+1\right)\mathbf{P}\left\{Q^m_{T_m+N+1}=Q^m_{T_m+N}+1|\mathcal{H}_{T_m+N}\right\}\\&=Q^m_{T_m+N}\left(1-\frac{2Q^m_{T_m+N}}{2T_m+2N+1}\right)+\left(Q^m_N+1\right)\left(\frac{2Q^m_{T_m+N}}{2T_m+2N+1}\right)\\&=\frac{2T_m+2N+3}{2T_m+2N+1}Q^m_{T_m+N}.\end{aligned}$$

Hence

$$\mathbf{E}\left\{\frac{Q^m_{T_m+N+1}}{2T_m+2N+3}\Big|\mathcal{H}_{T_m+N}\right\}=\frac{Q^m_{T_m+N}}{2T_m+2N+1}$$

and it follows that

$$\mathbf{E}\left\{\frac{Q^m_{T_m+N}}{2T_m+2N+1}\Big|\mathcal{H}_{T_m}\right\}=\frac{1}{2T_m+1}$$

since $Q^m_{T_m}=1$. The second equation is a consequence of (2).

(4) The fact that $\{\frac{Q^m_N}{2N+1}\}$ is a sub-martingale with respect to $\{\mathcal{H}_N\}$ follows directly from (3) and the fact that $Q^m_N=0$ if $T_m>N$. That completes the proof. □

The following theorem gives us a more intuitive picture about the relation between these two models. We need to introduce a series of random variables for a random mapping graph with n vertices. A random mapping graph G_n gives a partition for $[n]$ which is the set of components and denoted by $\mathcal{E}_1,\mathcal{E}_2,\ldots,\mathcal{E}_{\nu_n}$ in order, where ν_n is the number of components. Clearly $\nu_n\leqslant n$. Given a natural number k, let $H^k_n:=|\mathcal{E}_k|$ for $k\leqslant\nu_n$, and $H^k_n:=0$ for $k>\nu_n$.

For $N\leqslant n$ we shall see how the partition is shown locally in $[N]$. Let $\mathcal{A}^k_N:=\mathcal{E}_k\cap[N]$, and

$$D_{n,N}:=\sup\left\{k:\ \mathcal{A}^k_N\neq\emptyset\right\}=\sup\left\{k:\ \min\mathcal{E}_k\leqslant N\right\}$$

and obviously $D_{n,N}\leqslant N$. With this notation, for a fixed partition $A_1,\ldots,A_m$ of $[N]$ with proper order,

$$J_{A_1}\oplus\cdots\oplus J_{A_m}=\left\{G_n:\ \mathcal{A}^k_N=A_k,\ 1\leqslant k\leqslant m\right\}.$$

For $1\leqslant k\leqslant D_{n,N}$, let $Q^k_{n,N}:=|\mathcal{A}^k_N|$ be the number of vertices in $\mathcal{A}^k_N$ and $H^k_{n,N}$ the number of vertices of the component which contains $\mathcal{A}^k_N$ in G_n. For $k>D_{n,N}$ set $Q^k_{n,N}=H^k_{n,N}=0$. By definition it is easy to see that

$$\begin{aligned}&N=\sum_{k=1}^{\infty}Q^k_{n,N},\qquad n=\sum_{k=1}^{\infty}H^k_{n,n}\geqslant\sum_{k=1}^{\infty}H^k_{n,N},\\&Q^{N+1}_{n,N}=Q^{N+2}_{n,N}=\cdots=0,\\&H^k_{n,N}=H^k_n\cdot 1_{\{k\leqslant D_{n,N}\}}.\end{aligned}$$

It follows from Theorem 3.1 that $(Q^k_{n,N}:\ k\geqslant 1)$ converges to $(Q^k_N:\ k\geqslant 1)$ in law in the sense that for any $k\geqslant 1$ and $l_1,\ldots,l_k\geqslant 1$, it holds that

$$\lim_n\mathbf{P}_n\left(Q^1_{n,N}=l_1,\ldots,Q^k_{n,N}=l_k\right)=\mathbf{P}\left(Q^1_N=l_1,\ldots,Q^k_N=l_k\right),$$

which shall be written as $(Q^k_{n,N}:\ k\geqslant 1)\xrightarrow{\mathrm{d}}(Q^k_N:\ k\geqslant 1)$.

As we see above, some information of a random mapping graph will be reflected in a subset. For $N\leqslant n$, a property of a random mapping graph with n vertices recorded in its subset $[N]$ is called its local image, which is asymptotically embedded into a nice probability space. The theorem below says that such a property (i.e., ratio $\frac{H_n^k}{n}$ of the k-th component in G_n) may be approximated by its local image (ratio $\frac{Q_{n,N}^k}{N}$ of the part of it located in $[N]$). To prove it, we mainly use the techniques developed above.

Theorem 3.2.

$$\lim_{N\to\infty}\limsup_{n\to\infty}\mathbf{E}_n\left(\sum_{k=1}^{\infty}\left(\frac{Q_{n,N}^k}{N}-\frac{H_n^k}{n}\right)^2\right)=0.$$

Proof. Set $T_{n,1}=1$ and for any $G_n\in\Omega_n$,

$$T_{n,m}(G_n):=\min\mathcal{E}_m=\inf\{j>T_{n,m-1}\colon\ J_{\{T_{n,1}\}}\oplus\cdots\oplus J_{\{T_{n,m-1}\}}\oplus J_{\{j\}}\text{ occurs in }G_n\},$$

$(\inf\emptyset=\infty)$ for $m\geqslant 2, n\geqslant 1$, i.e., $T_{n,m}$ is the least numbered vertex in the m-th component of G_n. By definition, we have $D_{n,T_{n,m}}=m$, $\mathcal{A}_{T_{n,m}}^m=\{T_{n,m}\}$, $Q_{n,N}^m=H_{n,N}^m=0$ on $\{T_{n,m}>N\}$, and $H_n^m=H_{n,T_{n,m}}^m=H_{n,N}^m$ on $\{T_{n,m}\leqslant N\}$, since the component which contains vertex $T_{n,m}$ always contains vertices in A_m.

Employing Theorem 3.1, we can easily get when $n\to\infty$, $T_{n,m}$ and $Q_{n,N}^m$ asymptotically converge to T_m and Q_N^m in distribution respectively, namely, for any fixed m,k,N,

$$\lim_{n\to\infty}\mathbf{P}_n(T_{n,m}=k)=\mathbf{P}(T_m=k),$$
$$\lim_{n\to\infty}\mathbf{P}_n\big(Q_{n,N}^m=k\big)=\mathbf{P}\big(Q_N^m=k\big),$$

since $\{T_{n,m}=k\}$, $\{T_m=k\}$, $\{Q_{n,N}^m=k\}$, $\{Q_N^m=k\}$ can be decomposed into finite union of events of the form $J_{A_1}\oplus\cdots\oplus J_{A_m}$ or $A_1\oplus\cdots\oplus A_m$ in respective probability space. Furthermore, by the dominated convergence theorem and Lemma 3.3, we have for any fixed m,N,

$$\lim_{n\to\infty}\mathbf{E}_n\left(\frac{1}{2T_{n,m}+1}\right)=\mathbf{E}\left(\frac{1}{2T_m+1}\right)=\frac{1}{3^m},$$
$$\lim_{n\to\infty}\mathbf{E}_n\big(Q_{n,N}^m\big)=\mathbf{E}\big(Q_N^m\big)\leqslant\frac{2N+1}{3^m}.$$

For any $1>\delta>0$ and $m\in\mathbf{N}$, there exist N_m such that, $\mathbf{P}(T_m\leqslant N_m)\geqslant 1-\frac{\delta}{2}$ since T_m is a finite stopping time. Then we have

$$\mathbf{P}_n(T_{n,m}\leqslant N_m)\geqslant 1-\delta,$$
$$\mathbf{E}_n\left(\frac{1}{2T_{n,m}+1}\right)\leqslant\frac{1+\delta}{3^m},$$
$$\mathbf{P}_n(J_{A_1}\oplus\cdots\oplus J_{A_m}\oplus J_{\{M+1\}}|J_{A_1}\oplus\cdots\oplus J_{A_m})\leqslant\frac{1+\delta}{2M+1}$$

for any $n>n_1$, where n_1 depends on m, N_m, and $\{A_1,\ldots,A_m\}$ is any partition of $[M]$ with $M\leqslant N_m$.

Let us now estimate the expectation in question by several steps. The easy part is that for n large,

$$\mathbf{E}_n\left\{\sum_{k=1}^{\infty}\left(\frac{Q_{n,N}^k}{N}-\frac{H_n^k}{n}\right)^2;\ T_{n,m}>N_m\right\}\leqslant 2\mathbf{P}_n(T_{n,m}>N_m)\leqslant 2\delta.$$

To estimate the other case $T_{n,m}\leqslant N_m$, we separate the sum into two parts: the tail $k>m$ and the main body $k\leqslant m$. To estimate the tail of $\{Q_{n,N}^k\colon\ k\geqslant 1\}$ first, we have

$$\lim_{n\to\infty}\mathbf{E}_n\left(\sum_{k=m+1}^{\infty}Q_{n,N}^k\right)=\lim_{n\to\infty}\mathbf{E}_n\left(\sum_{k=m+1}^{N}Q_{n,N}^k\right)=\mathbf{E}\left(\sum_{k=m+1}^{N}Q_N^k\right)$$
$$=\sum_{k=m+1}^{N}\mathbf{E}\big(Q_N^k\big)\leqslant\sum_{k=m+1}^{N}\frac{2N+1}{3^k}<\frac{2}{3^m}N.$$

Hence there exists n_2 which depends on m, N, such that for $n > n_2$, we have

$$\mathbf{E}_n\left(\sum_{k=m+1}^{\infty} Q_{n,N}^k\right) \leqslant \frac{2}{3^m}N.$$

To estimate the tail of $\{H_n^k\colon k \geqslant 1\}$ is a little harder. Clearly for $T_{n,m} < \infty$, we have

$$\sum_{k=m+1}^{\infty} H_n^k = \sum_{j=T_{n,m}+1}^{n} 1_{J_{\mathcal{A}^1_{T_{n,m}}}\oplus\cdots\oplus J_{\mathcal{A}^m_{T_{n,m}}}\oplus J_{\{j\}}},$$

where the notation (and similar in the sequel) $J_{\mathcal{A}^1_{T_{n,m}}} \oplus \cdots \oplus J_{\mathcal{A}^m_{T_{n,m}}} \oplus J_{\{j\}}$ denotes the set of G_n satisfying

$$G_n \in J_{\mathcal{A}^1_{T_{n,m}}(G_n)} \oplus \cdots \oplus J_{\mathcal{A}^m_{T_{n,m}}(G_n)} \oplus J_{\{j\}},$$

namely, the event that vertex j is not connected with the first m components.

For any $M > m \geqslant 1$ with $M \leqslant N_m$, and an ordered partition $\{A_1, \ldots, A_{m-1}, A_m\}$ of $[M]$ with $A_m = \{M\}$, it holds that

$$\begin{aligned}
&\mathbf{P}_n\big(J_{\mathcal{A}^1_{T_{n,m}}} \oplus \cdots \oplus J_{\mathcal{A}^m_{T_{n,m}}} \oplus J_{\{T_{n,m}+1\}} | T_{n,m} = M, \mathcal{A}^1_{T_{n,m}} = A_1, \ldots, \mathcal{A}^m_{T_{n,m}} = A_m\big)\\
&\quad = \mathbf{P}_n(J_{A_1} \oplus \cdots \oplus J_{A_m} \oplus J_{\{M+1\}} | J_{A_1} \oplus \cdots \oplus J_{A_m})\\
&\quad \leqslant \frac{1+\delta}{2M+1},
\end{aligned}$$

for $n > n_1$. Then, on $\{T_{n,m} \leqslant N_m\}$,

$$\mathbf{P}_n\big(J_{\mathcal{A}^1_{T_{n,m}}} \oplus \cdots \oplus J_{\mathcal{A}^m_{T_{n,m}}} \oplus J_{\{T_{n,m}+1\}} | T_{n,m}, \mathcal{A}^1_{T_{n,m}}, \ldots, \mathcal{A}^m_{T_{n,m}}\big) \leqslant \frac{1+\delta}{2T_{n,m}+1}.$$

It follows that

$$\begin{aligned}
&\mathbf{E}_n\left(\sum_{k=m+1}^{\infty} H_n^k; T_{n,m} \leqslant N_m\right)\\
&\quad = \mathbf{E}_n\left(\mathbf{E}_n\left(\sum_{k=m+1}^{\infty} H_n^k | T_{n,m}, \mathcal{A}^1_{T_{n,m}}, \ldots, \mathcal{A}^m_{T_{n,m}}\right); T_{n,m} \leqslant N_m\right)\\
&\quad = \mathbf{E}_n\left(\mathbf{E}_n\left(\sum_{j=T_{n,m}+1}^{n} 1_{J_{\mathcal{A}^1_{T_{n,m}}}\oplus\cdots\oplus J_{\mathcal{A}^m_{T_{n,m}}}\oplus J_{\{j\}}} | T_{n,m}, \mathcal{A}^1_{T_{n,m}}, \ldots, \mathcal{A}^m_{T_{n,m}}\right); T_{n,m} \leqslant N_m\right)\\
&\quad = \mathbf{E}_n\left(\sum_{j=T_{n,m}+1}^{n} \mathbf{E}_n\big(1_{J_{\mathcal{A}^1_{T_{n,m}}}\oplus\cdots\oplus J_{\mathcal{A}^m_{T_{n,m}}}\oplus J_{\{j\}}} | T_{n,m}, \mathcal{A}^1_{T_{n,m}}, \ldots, \mathcal{A}^m_{T_{n,m}}\big); T_{n,m} \leqslant N_m\right)\\
&\quad = \mathbf{E}_n\big((n - T_{n,m})\mathbf{E}_n\big(1_{J_{\mathcal{A}^1_{T_{n,m}}}\oplus\cdots\oplus J_{\mathcal{A}^m_{T_{n,m}}}\oplus J_{\{T_{n,m}+1\}}} | T_{n,m}, \mathcal{A}^1_{T_{n,m}}, \ldots, \mathcal{A}^m_{T_{n,m}}\big); T_{n,m} \leqslant N_m\big)\\
&\quad = \mathbf{E}_n\left((n - T_{n,m})\frac{1+\delta}{2T_{n,m}+1}; T_{n,m} \leqslant N_m\right)\\
&\quad \leqslant n(1+\delta)\mathbf{E}_n\left(\frac{1}{2T_{n,m}+1}\right)\\
&\quad \leqslant \frac{(1+\delta\,)^2}{3^m}n.
\end{aligned}$$

It then implies that

$$\mathbf{E}_n\left(\sum_{k=m+1}^{\infty} H_n^k\right) \leqslant \left(\frac{(1+\delta\,)^2}{3^m} + \delta\right)n.$$

Therefore we have a tail estimate

$$\mathbf{E}_n\left\{\sum_{k=m+1}^{\infty}\left(\frac{Q_{n,N}^k}{N}-\frac{H_n^k}{n}\right)^2\right\}\leqslant \mathbf{E}_n\left\{\sum_{k=m+1}^{\infty}\left(\frac{Q_{n,N}^k}{N}+\frac{H_n^k}{n}\right)\right\}\leqslant \frac{2}{3^m}+\delta+\frac{(1+\delta)^2}{3^m}.$$

We are now left to estimate the hard part

$$\mathbf{E}_n\left\{\sum_{k=1}^{m}\left(\frac{Q_{n,N}^k}{N}-\frac{H_n^k}{n}\right)^2;\, T_{n,m}\leqslant N_m\right\}.$$

Set $S(n,t,a,h)=\{\alpha\subseteq[n]\backslash[t]\colon |\alpha|=h-a\}$. Since the symmetric property of $\mathbf{P}_n$, we have, on $\{T_{n,m}\leqslant N_m\}$, for any $\alpha\in S(n,T_{n,m},|\mathcal{A}_{T_{n,m}}^k|,H_{n,T_{n,m}}^k)$, $1\leqslant k\leqslant m$,

$$\mathbf{P}_n(J_{\mathcal{A}_{T_{n,m}}^k\cup\alpha}|T_{n,m},\mathcal{A}_{T_{n,m}}^1\cdots\mathcal{A}_{T_{n,m}}^m,H_{n,T_{n,m}}^1,\ldots,H_{n,T_{n,m}}^m)=\frac{1}{|S(n,T_{n,m},|\mathcal{A}_{T_{n,m}}^k|,H_{n,T_{n,m}}^k)|}.$$

Hence on $\{T_{n,m}\leqslant N_m\}$, for $1\leqslant k\leqslant m$, $0\leqslant Q_k\leqslant N\leqslant n-N_m$, we have

$$\begin{aligned}
&\mathbf{P}_n\big\{Q_{n,N+T_{n,m}}^k=Q_k+\big|\mathcal{A}_{T_{n,m}}^k\big|\big|T_{n,m},\mathcal{A}_{T_{n,m}}^1\cdots\mathcal{A}_{T_{n,m}}^m,H_{n,T_{n,m}}^1,\ldots,H_{n,T_{n,m}}^m\big\}\\
&\quad=\mathbf{P}_n\left\{\bigcup_{\alpha_1,\alpha_2}J_{\mathcal{A}_{T_{n,m}}^k\cup\alpha_1\cup\alpha_2}|T_{n,m},\mathcal{A}_{T_{n,m}}^1\cdots\mathcal{A}_{T_{n,m}}^m,H_{n,T_{n,m}}^1,\ldots,H_{n,T_{n,m}}^m\right\}\\
&\quad=\frac{|S(N+T_{n,m},T_{n,m},|\mathcal{A}_{T_{n,m}}^k|,Q_k+|\mathcal{A}_{T_{n,m}}^k|)||S(n,N+T_{n,m},Q_k+|\mathcal{A}_{T_{n,m}}^k|,H_{n,T_{n,m}}^k)|}{|S(n,T_{n,m},|\mathcal{A}_{T_{n,m}}^k|,H_{n,T_{n,m}}^k)|}\\
&\quad=\frac{\binom{N}{Q_k}\binom{n-N-T_{n,m}}{H_{n,T_{n,m}}^k-Q_k-|\mathcal{A}_{T_{n,m}}^k|}}{\binom{n-T_{n,m}}{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}}\\
&\quad=\frac{\binom{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{Q_k}\binom{n-T_{n,m}-H_{n,T_{n,m}}^k+|\mathcal{A}_{T_{n,m}}^k|}{N-Q_k}}{\binom{n-T_{n,m}}{N}},
\end{aligned}$$

where $\bigcup_{\alpha_1,\alpha_2}$ means union for all

$$\alpha_1\in S\big(N+T_{n,m},T_{n,m},\big|\mathcal{A}_{T_{n,m}}^k\big|,Q_k+\big|\mathcal{A}_{T_{n,m}}^k\big|\big),\qquad \alpha_2\in S\big(n,N+T_{n,m},Q_k+\big|\mathcal{A}_{T_{n,m}}^k\big|,H_{n,T_{n,m}}^k\big),$$

i.e., the conditional distribution is hypergeometric. It follows that

$$\begin{aligned}
&\mathbf{E}_n\left\{\left(Q_{n,N+T_{n,m}}^k-\big|\mathcal{A}_{T_{n,m}}^k\big|-N\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}\right)^2\,\middle|\,T_{n,m},\mathcal{A}_{T_{n,m}}^1\cdots\mathcal{A}_{T_{n,m}}^m,H_{n,T_{n,m}}^1,\ldots,H_{n,T_{n,m}}^m\right\}\\
&\quad=N\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}\left(1-\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}\right)\frac{n-T_{n,m}-N}{n-T_{n,m}-1}\\
&\quad\leqslant N\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}.
\end{aligned}$$

Furthermore, on $\{T_{n,m}\leqslant N_m\}$, we have

$$\begin{aligned}
&\mathbf{E}_n\left\{\sum_{k=1}^{m}\left(\frac{Q_{n,N+T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{N}-\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}\right)^2\,\middle|\,T_{n,m},\mathcal{A}_{T_{n,m}}^1\cdots\mathcal{A}_{T_{n,m}}^m,H_{n,T_{n,m}}^1,\ldots,H_{n,T_{n,m}}^m\right\}\\
&\quad\leqslant\frac{1}{N}\sum_{k=1}^{m}\frac{H_{n,T_{n,m}}^k-|\mathcal{A}_{T_{n,m}}^k|}{n-T_{n,m}}=\frac{1}{N}\frac{\sum_{k=1}^m H_{n,T_{n,m}}^k-T_{n,m}}{n-T_{n,m}}\leqslant\frac{1}{N}.
\end{aligned}$$

Therefore, it holds that

$$\mathbf{E}_n\left\{\sum_{k=1}^m\left(\frac{Q^k_{n,N+T_{n,m}}-|\mathcal{A}^k_{T_{n,m}}|}{N}-\frac{H^k_{n,T_{n,m}}-|\mathcal{A}^k_{T_{n,m}}|}{n-T_{n,m}}\right)^2;\, T_{n,m}\leqslant N_m\right\}\leqslant\frac{1}{N}\mathbf{P}_n(T_{n,m}\leqslant N_m)\leqslant\frac{1}{N}.$$

However, it is easy to see that (1) $|\mathcal{A}^k_{T_{n,m}}|\leqslant T_{n,m}$; (2) $H^k_{n,T_{n,m}}=H^k_n$ for $k\leqslant m$; (3) $0\leqslant Q^k_{n,N+T_{n,m}}-Q^k_{n,N}\leqslant T_{n,m}$. Hence if $T_{n,m}\leqslant N_m$, there exists n_3 which depends on N_m such that for any $n>N>n_3$ we have

$$\left|\sum_{k=1}^m\left(\frac{Q^k_{n,N+T_{n,m}}-|\mathcal{A}^k_{T_{n,m}}|}{N}-\frac{H^k_{n,T_{n,m}}-|\mathcal{A}^k_{T_{n,m}}|}{n-T_{n,m}}\right)^2-\sum_{k=1}^m\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2\right|\leqslant\delta.$$

This yields

$$\mathbf{E}_n\left(\sum_{k=1}^m\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2;\, T_{n,m}\leqslant N_m\right)\leqslant\frac{1}{N}+\delta.$$

Finally we have for $N>n_3$, $n>\max\{n_1,n_2,n_3,N\}$,

$$\begin{aligned}\mathbf{E}_n\left\{\sum_{k=1}^\infty\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2\right\}&\leqslant\mathbf{E}_n\left\{\sum_{k=1}^m\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2;\, T_{n,m}\leqslant N_m\right\}+\mathbf{E}_n\left\{\sum_{k=m+1}^\infty\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2\right\}\\&\quad+\mathbf{E}_n\left\{\sum_{k=1}^\infty\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2;\, T_{n,m}>N_m\right\}\\&\leqslant\frac{1}{N}+\delta+\frac{2}{3^m}+\delta+\frac{(1+\delta)^2}{3^m}+2\delta<\frac{1}{N}+10\delta,\end{aligned}$$

by choosing m large enough such that $3^{-m}<\delta$. Hence

$$\lim_N\limsup_n\mathbf{E}_n\left\{\sum_{k=1}^\infty\left(\frac{Q^k_{n,N}}{N}-\frac{H^k_n}{n}\right)^2\right\}\leqslant 10\delta.$$

The conclusion follows since δ may be arbitrarily small. □

Now we are at a position to uncover how our results can be used to discuss asymptotic behaviors of random mappings, which generally means the limit probability of a property or an event of a random mapping graph as the number of vertices goes to infinity. Our program runs like this: (1) a property of a random graph is approached by its local image; (2) the local image converges to a property in the scheme of allocation; (3) this property in the scheme shows some asymptotic behavior. Now a lemma, easy to prove, is prepared to bridge the last inch of the gap concerning asymptotic behaviors in two models. Let $\{X_n\}$ and $\{X_{n,N}\}$ be random sequences in $(\Omega_n,\mathcal{F}_n,\mathbf{P}_n)$, and $\{Y_N\}$ a random sequence in $(\Omega,\mathcal{F},\mathbf{P})$.

Lemma 3.4. *Assume that for any $\delta>0$ and real x,*

$$\lim_N\limsup_n\mathbf{P}_n\big(|X_n-X_{n,N}|>\delta\big)=0\quad\text{and}\quad\lim_n\mathbf{P}_n(X_{n,N}\leqslant x)=\mathbf{P}(Y_N\leqslant x).$$

If Y_N converges to Y in law, then $\lim_n\mathbf{P}_n(X_n\leqslant x)=\mathbf{P}(Y\leqslant x)$ for any real x. In this case we also say that X_n converges in law to Y, or simply $X_n\xrightarrow{\mathrm{d}}Y$.

Combining all results above, we can see that any asymptotic behavior of the Markov chain $\{Q_N\}$ leads to a similar behavior of random mappings. In other words, asymptotic behaviors concerning component size of random mappings are embedded in the related Markov chain. A direct consequence is that components of a random mapping are asymptotically organized as the blocks of a (0, 1/2)-Chinese restaurant process (refer to [6]). It follows from Lemma 3.3(4) that Q^k_N/N converges a.s. and in L^1 to a random variable, say, X_k.

Corollary 3.1. *As n goes to infinity, $(\frac{H^k_n}{n}\colon k\geqslant 1)\xrightarrow{\mathrm{d}}(X_k\colon k\geqslant 1)$.*

We shall give several examples to show how this program works to recover classical results.

Example 1. From the following theorem we can easily get one result of Pittel [7] on two-sided epidemic processes which was proposed by Gertsbakh [4], also see Stepanov [12]. The result states that for any fix $r \geqslant 1$, starting with r infected elements (or vertices) in a random mapping with n elements(or vertices), in a two-sided epidemic process, the percentage of eventually infected elements is asymptotically beta-distributed with parameters r and $\frac{1}{2}$ as n goes to infinity. In other words, the result gives the limit distribution of the ratio of the number of elements in components containing the first r vertices,

$$\lim_n \mathbf{P}_n\left(\frac{\sum_{k=1}^{\infty} H_{n,r}^k}{n} \leqslant x\right) = \frac{(2r-1)!!}{2(2r-2)!!}\int_0^x (1-y)^{-1/2} y^{r-1}\,\mathrm{d}y, \quad x \in [0,1],$$

where $H_{n,r}^k = H_n^k$ if the k-th component contains a vertex in $[r]$ and $H_{n,r}^k = 0$ otherwise. This follows from the theorem below, Theorem 3.2 and Lemma 3.4.

Theorem 3.3. *For $r \geqslant 1$, there exists a random variable ζ_r, such that $N^{-1}\sum_{k=1}^{D_r} Q_N^k \to \zeta_r$, L_1 & a.s. and ζ_r is beta-distributed with parameters r and $\frac{1}{2}$.*

Proof. Set $\widetilde{Q}_N^r := \sum_{k=1}^{D_r} Q_N^k$. Clearly, when $N = r$,

$$\mathbf{P}\big(\widetilde{Q}_r^r = r\big) = \mathbf{P}\big(Q_r^1 = r | Q_r^1 = r\big) = 1$$

and when $N > r$,

$$\begin{aligned}
\mathbf{P}\big(\widetilde{Q}_{N+1}^r = 1 + \widetilde{Q}_N^r | \mathcal{H}_N\big) &= \mathbf{E}\big(\mathbf{P}\big(\widetilde{Q}_{N+1}^r = 1 + \widetilde{Q}_N^r \big| D_r\big)\big| \mathcal{H}_N\big) \\
&= \mathbf{E}\left(\mathbf{P}\left(\sum_{k=1}^{D_r} Q_{N+1}^k = 1 + \sum_{k=1}^{D_r} Q_N^k \Big| D_r\right)\Bigg| \mathcal{H}_N\right) \\
&= \mathbf{E}\left(\sum_{k=1}^{D_r} \mathbf{P}\big(Q_{N+1}^k = 1 + Q_N^k \big| D_r\big)\big| \mathcal{H}_N\right) \\
&= \mathbf{E}\left(\sum_{k=1}^{D_r} \frac{2Q_N^k}{2N+1}\Bigg| \mathcal{H}_N\right) = \frac{2\widetilde{Q}_N^r}{2N+1}.
\end{aligned}$$

It follows that

$$\mathbf{P}\big(\widetilde{Q}_{N+1}^r = \widetilde{Q}_N^r \big| \mathcal{H}_N\big) = 1 - \frac{2\widetilde{Q}_N^r}{2N+1}.$$

It is then verified that $\{\frac{\widetilde{Q}_N^r}{2N+1}\colon N \geqslant r\}$ is a bounded martingale with respect to $\{\mathcal{H}_N\colon N \geqslant r\}$. Therefore, there exists a random variable ζ_r such that,

$$\frac{\widetilde{Q}_N^r}{2N+1} \to \frac{\zeta_r}{2}, \quad L_1 \text{ \& a.s.}$$

We need to show that ζ_r is beta-distributed with parameters r and $\frac{1}{2}$. It is seen that $\{\widetilde{Q}_N^r\}$ is a Markov chain with respect to $\{\mathcal{H}_N\colon N \geqslant r\}$. Moreover, the chain is independent of D_r. As a result,

$$\mathbf{P}\big(\widetilde{Q}_N^r \in \bullet\big) = \mathbf{P}\big(\widetilde{Q}_N^r \in \bullet | D_r = 1\big) = \mathbf{P}\big(Q_N^1 \in \bullet | Q_r^1 = r\big), \quad \forall N > r,$$

since $\{D_r = 1\} = \{Q_r^1 = r\}$. Set $\Lambda(k_1, \ldots, k_m) = \{A_1 \oplus \cdots \oplus A_m\colon |A_1| = k_1, \ldots, |A_m| = k_m$ and they are a partition of $[k_1 + \cdots + k_m]\}$ and $\Lambda_1(k_1, \ldots, k_m) = \{A_1 \oplus \cdots \oplus A_m\colon A_1 \supseteq [r]$, and $|A_1| = k_1, \ldots, |A_m| = k_m$ and they are a partition of $[k_1 + \cdots + k_m]\}$. And we shall write $\alpha(k_1, \ldots, k_m) = |\Lambda(k_1, \ldots, k_m)|$ and $\alpha_1(k_1, \ldots, k_m) = |\Lambda_1(k_1, \ldots, k_m)|$. For any $r \leqslant M < N$,

$$
\begin{aligned}
\mathbf{P}\big(Q_N^1 = M, Q_r^1 = r\big) &= \sum_{m=1}^{N} \sum_{k_2+\cdots+k_m=N-M} \mathbf{P}\big\{\Lambda_1(M, k_2, \ldots, k_m)\big\} \\
&= \sum_{m=1}^{N} \sum_{k_2+\cdots+k_m=N-M} \frac{\alpha_1(M, k_2, \ldots, k_m)(M-1)!(k_2-1)!\cdots(k_m-1)!2^{N-m}}{(2N-1)!!} \\
&= \sum_{m=1}^{N} \sum_{k_2+\cdots+k_m=N-M} \binom{N-r}{M-r} \frac{\alpha(k_2, \ldots, k_m)(M-1)!(k_2-1)!\cdots(k_m-1)!2^{N-m}}{(2N-1)!!} \\
&= \binom{N-r}{M-r} \frac{(M-1)!(2N-2M-1)!!}{2^{-M+1}(2N-1)!!} \\
&\quad \times \sum_{m=2}^{N} \sum_{k_2+\cdots+k_m=N-M} \frac{\alpha(k_2, \ldots, k_m)(k_2-1)!\cdots(k_m-1)!2^{(N-M)-(m-1)}}{(2N-2M-1)!!} \\
&= \binom{N-r}{M-r} \frac{(M-1)!(2N-2M-1)!!}{2^{-M+1}(2N-1)!!} \sum_{m=2}^{N} \sum_{k_2+\cdots+k_m=N-M} \mathbf{P}\big\{\Lambda(k_2, \ldots, k_m)\big\} \\
&= \frac{1}{2^{-2M+1}} \frac{N!N!}{N(2N)!} \frac{(2N-2M)!}{(N-M)!(N-M)!} \frac{(M-2)_{r-1}}{(N-2)_{r-1}}.
\end{aligned}
$$

Since $\mathbf{P}(Q_r^1 = r) = \frac{(2r-2)!!}{(2r-1)!!}$, it follows from Stirling formula that, as M is large enough,

$$
\mathbf{P}\big(\widetilde{Q}_N^r = M\big) = c_r \left(\frac{M}{N}\right)^{r-1} \frac{1}{\sqrt{N(N-M)}} \left(1 + \frac{C_1}{N} + \frac{C_2}{M} + \frac{C_2}{N-M}\right),
$$

where C_1, C_2, C_3 are bounded functions which depend on $N, M, N-M$ respectively. Hence

$$
\mathbf{P}\big(\widetilde{Q}_N^r \leqslant xN\big) = c_r \int_0^x (1-y)^{-1/2} y^{r-1} \, \mathrm{d}y + \mathrm{o}(1).
$$

That completes the proof. □

Example 2. The following theorem leads to a result of Stepanov [11] on L_n^1, the size of the largest component of a random mapping with n elements, also see Kolchin [5], which states that

$$
\mathbf{P}_n\left(\frac{L_n^1}{n} \geqslant x\right) \to \int_x^1 \frac{1}{2x\sqrt{1-x}} \, \mathrm{d}x, \quad x \in \left[\frac{1}{2}, 1\right].
$$

Theorem 3.4. *As N goes to infinity,*

$$
\mathbf{P}\left(\max_k \frac{Q_N^k}{N} \geqslant x\right) \to \int_x^1 \frac{1}{2x\sqrt{1-x}} \, \mathrm{d}x, \quad x \in \left[\frac{1}{2}, 1\right].
$$

Proof. It follows from a similar argument as in Theorem 3.3 that for any fixed N, M with $\frac{N}{2} < M \leqslant N$, it holds that

$$
\mathbf{P}\Big(\max_k Q_N^k = M\Big) = \frac{1}{2^{-2M+1}M} \frac{N!N!}{(2N)!} \frac{(2N-2M)!}{(N-M)!(N-M)!}.
$$

As M large enough, using Stirling's formula again, we have

$$
\mathbf{P}\Big(\max_k Q_N^k = M\Big) = \frac{1}{2M} \sqrt{\frac{N}{N-M}} \left(1 + \frac{C_1}{N} + \frac{C_2}{N-M}\right),
$$

where C_1, C_2 are bounded functions dependent on $N, N-M$ respectively, and this implies our assertion. □

Example 3. The study of the limit behavior of sequence concerning the size of components of random mappings

$$\left(\frac{H_n^1}{n},\dots,\frac{H_n^m}{n},\dots\right)$$

goes back at least to Stepanov [12], also see Aldous et al. [2,1], which states

$$\left(\frac{H_n^1}{n},\frac{H_n^2}{n},\dots,\frac{H_n^m}{n},\dots\right)\xrightarrow{\mathrm{d}}\left(\xi_1,\xi_2(1-\xi_1),\dots,\xi_m\prod_{k=1}^{m-1}(1-\xi_k),\dots\right),$$

where (ξ_k) are i.i.d. and beta-distributed with parameters 1 and $\frac{1}{2}$. This follows directly from Corollary 3.1, Theorem 3.3 and the self-similarity of (Q_N) as explained below, see also [6]. For $m \geqslant 1$, define $b_n^m := \sum_{k>m} Q_n^k$, the total number of particles in boxes beyond m after n particles are placed, which increases at most by 1 each time, and $\tau_N^m := \inf\{n\colon b_n^m = N\}$ for $N \geqslant 1$. Intuitively τ_N^m is the N-th particle placed beyond the first m boxes. Obviously τ_N^m is a finite stopping time since $\tau_N^m \leqslant T_{m+N}$. It is easy to check that $\{Q_{\tau_N}\}$ is a time change of $\{Q_N\}$. For any $\mathbf{x} \in S$, $\pi_m\mathbf{x} := (x_1,\dots,x_m)$, a projection, and $\xi_m\mathbf{x} := (x_{m+1},x_{m+2},\dots)$, a translation.

Theorem 3.5. *For any $m \geqslant 1$, $\{\xi_m(Q_{\tau_N^m})\colon N \geqslant 1\}$ is a Markov chain independent of $\{\pi_m(Q_{\tau_N^m})\colon N \geqslant 1\}$ and identical in law with $\{Q_N\colon N \geqslant 1\}$.*

4. The expected size of the largest component

In this section, we shall present the moments of size of the largest component in a random mapping graph. For $r, m \geqslant 1$, define

$$E(x) := \int_x^\infty \frac{\mathrm{e}^{-y}}{2y}\,\mathrm{d}y, \qquad G_{r,m} := \int_0^\infty x^{m-1}\frac{[E(x)]^{r-1}}{(r-1)!}\mathrm{e}^{-E(x)}\frac{\mathrm{e}^{-x}}{2}\,\mathrm{d}x.$$

Theorem 4.1. *Let L_r be the size of the r-th largest component in a random mapping graph in Ω_n. Then*

$$\lim_n \mathbf{E}_n(L_r/n)^m = \frac{2^m}{(2m-1)!!}G_{r,m}.$$

Particularly $\lim_n \mathbf{E}_n(L_1/n) \doteq 0.7575\dots.$

Proof. As we have known, there is a probability space $(\Omega,\mathcal{H},\mathbf{P})$ and $A_1 \oplus \cdots \oplus A_m$ is an atom of $\mathcal{H}_N$ where $A_1,\dots,A_m$ are an ordered partition of $[N]$, and $Q_N = (Q_N^k)$ where $Q_N^k = |A_k|, k = 1,\dots,$ and $\mathbf{P}$ satisfies that

$$\mathbf{P}(A_1 \oplus \cdots \oplus A_m) = \frac{(2a_1)!!\cdots(2a_m)!!}{(2N-1)!!},$$

where $a_l = |A_l| - 1$.

Define $\zeta_{i,N} := \#\{k\colon Q_N^k = i\}$ and $\zeta^{(N)} := (\zeta_{i,N}\colon i \geqslant 1)$. Suppose that $s = \{s_i\colon s_i \in \mathbf{Z}_+\}$ with

$$\upsilon(s) := \sum_{i=1}^\infty i s_i = N.$$

It is easily shown that

$$\begin{aligned}\mathbf{P}(\zeta_{i,N} = s_i\colon i = 1,\dots) &= \frac{N!}{\coprod_{i=1}^\infty (i!)^{s_i} s_i!}\,\frac{\coprod_{i=1}^\infty((2(i-1))!!)^{s_i}}{(2N-1)!!}\\ &= \frac{N!2^{\sum_{i=1}^\infty i s_i}\coprod_{i=1}^\infty (2i)^{-s_i}}{(2N-1)!!}\\ &= \frac{N!2^N}{(2N-1)!!}\prod_{i=1}^\infty \frac{(2i)^{-s_i}}{s_i!}.\end{aligned}$$

It implies the identity

$$\sum_{s:\ \upsilon(s)=N} \frac{N!2^N}{(2N-1)!!} \prod_{i=1}^{\infty} \frac{(2i)^{-s_i}}{s_i!} = 1. \tag{1}$$

The random variables $\{\zeta_{i,N}\colon\ i \geqslant 1\}$ would be independent if it were not for the condition on $\upsilon(s) = N$. Consider then a sequence $\zeta = \{\zeta_i\}$ of mutually independent nonnegative integer valued random variables, where for $i = 1, 2, \ldots$ the random variable ζ_i is Poisson distributed with mean $\frac{z^i}{2i}$, i.e.,

$$\mathbf{P}_z\{\zeta_i = s\} = \exp\left\{-\frac{z^i}{2i}\right\}\frac{(z^i/2i)^s}{s!}, \quad s = 0, 1, \ldots,$$

where, $z \in (0, 1)$.

Since $\mathbf{P}(\zeta_i \neq 0) = 1 - \exp\{-\frac{z^i}{2i}\} < \frac{z^i}{2i}$ and $\sum_{j=1}^{\infty} P_z(\zeta_i \neq 0)$ is finite, it follows from the Borel–Cantelli lemma, that $P_z(\zeta_i \neq 0,\ \text{infinitely often}) = 0$. Thus the random variable $\upsilon(\zeta) = \sum_{i=1}^{\infty} i\zeta_i$ is finite with probability 1, and the joint distribution of (ζ_i) may be written meaningfully as

$$\mathbf{P}_z(\zeta_i = s_i, i = 1, \ldots) = \prod_{i=1}^{\infty} \exp\left\{-\frac{z^i}{2i}\right\}\frac{(z^i/2i)^{s_i}}{s_i!} = \sqrt{1-z}\cdot z^{\upsilon(s)} \prod_{i=1}^{\infty} \frac{(2i)^{-s_i}}{s_i!}$$

for all sequences $s = (s_i)$ of nonnegative integers eventually 0. It is easy to see from (1) that the conditional distribution of the ζ's given $\upsilon(\zeta)$ does not depend on z, more precisely

$$\mathbf{P}_z(\zeta_i = s_i, i = 1, \ldots | \upsilon(\zeta) = N) = \begin{cases} \frac{N!2^N}{(2N-1)!!} \prod_{i=1}^{\infty} \frac{(2i)^{-s_i}}{s_i!}, & \text{if } \sum_{j=1}^{\infty} is_i = N, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore two probabilities are connected as follows

$$\mathbf{P}_z(\zeta_i = s_i, i = 1, \ldots | \upsilon(\zeta) = N) = \mathbf{P}(\zeta_{i,N} = s_i\colon\ i = 1, \ldots)$$

and the distribution of υ is

$$\mathbf{P}_z\big(\upsilon(\zeta) = N\big) = \sqrt{1-z}\cdot z^N \frac{(2N-1)!!}{N!2^N}, \quad N = 0, 1, \ldots.$$

Its expected value is $\mathbf{E}_z(\upsilon(\zeta)) = \sum_{i=1}^{\infty} \frac{z^i}{2} = \frac{z}{2-2z}$. Let Φ be a function on S. Then it follows that

$$\begin{aligned}\mathbf{E}_z\big(\Phi(\zeta)\big) &= \sum_{N\geqslant 0} \mathbf{E}_z\big(\Phi(\zeta)|\upsilon(\zeta) = N\big)\mathbf{P}_z\big(\upsilon(\zeta) = N\big) \\ &= \sum_{N\geqslant 0} \mathbf{E}\big(\Phi\big(\zeta^{(N)}\big)\big)\sqrt{1-z}\cdot z^N \frac{(2N-1)!!}{N!2^N}\end{aligned}$$

and then

$$\frac{1}{\sqrt{1-z}}\mathbf{E}_z\big(\Phi(\zeta)\big) = \sum_{N=0}^{\infty} z^N \frac{(2N-1)!!}{(2N)!!}\mathbf{E}\big(\Phi\big(\zeta^{(N)}\big)\big).$$

Given $r \geqslant 1$, define $\Phi(s) := \max\{i \geqslant 0\colon\ \sum_{j=i}^{\infty} s_j \geqslant r\}$ for $s \in S$, $L_r := \Phi(\zeta)$ and $L_{r,N} := \Phi(\zeta^{(N)})$. Then $L_{r,N}$ is the r-th large component in Q_N and we have

$$\frac{1}{\sqrt{1-z}}\mathbf{E}_z\big(L_r^m\big) = \sum_{N=0}^{\infty} z^N \frac{(2N-1)!!}{(2N)!!}\mathbf{E}\big(L_{r,N}^m\big).$$

It means that the left-hand side is equal to the generating function of $\{\frac{(2N-1)!!}{(2N)!!}\mathbf{E}(L_{r,N}^m)\colon\ N \geqslant 0\}$. An idea similar to that in [10] gives the following limit

$$\lim_{z\to 1}(1-z)^m\mathbf{E}_z\big(L_r^m\big) = G_{r,m}.$$

Thus

$$\lim_{z\to 1}\frac{(1-z)^{m+\frac{1}{2}}}{\Gamma(m+\frac{1}{2})}\cdot\frac{1}{\sqrt{1-z}}\mathbf{E}_z\big(L_r^m\big)=\frac{G_{r,m}}{\Gamma(m+\frac{1}{2})}.$$

Since $\{L_{r,N}\colon N\geqslant 1\}$ is obviously increasing, we may apply Karamata–Hardy–Littlewood Tauberian theorem (see, e.g., [3]), and Stirling's formula, and obtain

$$\lim_{N\to\infty}\mathbf{E}\bigg(\frac{L_{r,N}}{N}\bigg)^m=\frac{\sqrt{\pi}\,G_{r,m}}{\Gamma(m+\frac{1}{2})}=\frac{2^m}{(2m-1)!!}G_{r,m},$$

for the limiting form of the moments of $L_{r,N}/N$. The theorem follows from Theorem 3.2 and Lemma 3.4. The case $m=1, r=1$ gives the limit of

$$\lim_{n\to\infty}\mathbf{E}_n\bigg\{\frac{L_r}{n}\bigg\}=\int_0^\infty\exp\bigg\{-x-\int_x^\infty\frac{\mathrm{e}^{-y}}{2y}\,\mathrm{d}y\bigg\}\,\mathrm{d}x\approx 0.7575. \qquad \square$$

We now seek the limiting distribution of $L_{r,N}/N$. It is known that $\{L_{r,N}/N\colon N\geqslant 1\}$ converges a.s. Let $\eta_r:=\lim L_{r,N}/N$ with distribution $F_r(x), 0\leqslant x\leqslant 1$. By Lebesgue's dominated convergence theorem, the moments of η_r are given by

$$\mathbf{E}\eta_r^m=\int_0^1 x^m\,\mathrm{d}F_r(x)=\frac{2^m}{(2m-1)!!}G_{r,m},\quad m=0,1,\ldots.$$

Take a random variable X supported by $(0,\infty)$ with distribution

$$\mathbf{P}(X\in\mathrm{d}x)=\frac{[E(x)]^{r-1}}{(r-1)!}\,\mathrm{e}^{-E(x)}\frac{\mathrm{e}^{-x}}{2x}\,\mathrm{d}x.$$

Hence we have

$$\mathbf{E}\big(\eta_r^m\big)=\frac{2^m}{(2m-1)!!}\mathbf{E}\big(X^m\big).$$

Take a random variable Y which has a symmetric distribution on $(-\infty,\infty)$ about 0 with $Y^2\overset{d}{=}4X$, and we have

$$\mathbf{E}\big(\mathrm{e}^{\mathrm{i}Yz}\big)=\sum_{m=0}^\infty\frac{\mathbf{E}(Y^{2m})}{(2m)!}z^{2m}\mathrm{i}^{2m}=\sum_{m=0}^\infty\frac{\mathbf{E}(\eta_r^m)}{m!}\big(-z^2\big)^m=\mathbf{E}\big(\mathrm{e}^{-\eta_r z^2}\big).$$

On the other hand, take a standard Brownian motion $B=(B(t))$ on $\mathbf{R}$ independent of η, and then

$$\mathbf{E}\big(\mathrm{e}^{\mathrm{i}(\sqrt{2}B(\eta_r))z}\big)=\mathbf{E}\big(\mathbf{E}\big(\mathrm{e}^{\mathrm{i}(\sqrt{2}B(\eta_r))z}|\eta_r\big)\big)=\mathbf{E}\big(\mathrm{e}^{-\eta_r z^2}\big)=\mathbf{E}\big(\mathrm{e}^{iYz}\big).$$

It follows that $Y\overset{d}{=}\sqrt{2}B(\eta_r)$ or $X\overset{d}{=}\frac{1}{2}(B(\eta_r))^2$, i.e.,

$$\frac{[E(x)]^{r-1}}{(r-1)!}\,\mathrm{e}^{-E(x)}\frac{\mathrm{e}^{-x}}{2x}=\int_0^1\frac{2}{\sqrt{2x}}\frac{1}{\sqrt{2\pi t}}\,\mathrm{e}^{-\frac{x}{t}}f_r(t)\,\mathrm{d}t,$$

where f_r is the probability density of η_r. Let $\tau=\frac{1}{t}-1$, and set $g(\tau)\,\mathrm{d}\tau=-\frac{1}{\sqrt{t}}f_r(t)\,\mathrm{d}t$. It follows that

$$\int_0^\infty\mathrm{e}^{-x\tau}g(\tau)\,\mathrm{d}\tau=\frac{\sqrt{\pi}}{2}\frac{[E(x)]^{r-1}}{(r-1)!}\,\mathrm{e}^{-E(x)}\frac{1}{\sqrt{x}}=\frac{\sqrt{\pi}}{2}\sum_{m=0}^\infty\frac{(-1)^m(E(x))^{r+m-1}}{(r-1)!m!}\frac{1}{\sqrt{x}}.$$

Since

$$E(x)=\int_x^\infty\frac{\mathrm{e}^{-u}}{2u}\,\mathrm{d}u=\int_0^\infty\frac{\mathrm{e}^{-xu}}{2u}1_{\{u>1\}}\,\mathrm{d}u\quad\text{and}\quad\frac{1}{\sqrt{x}}=\int_0^\infty\mathrm{e}^{-xu}\frac{1}{\sqrt{\pi u}}\,\mathrm{d}u,$$

we have

$$\int_0^\tau g(x)\,\mathrm{d}x = \frac{1}{2}\sum_{m=0}^{\infty}\frac{(-1)^m}{(r-1)!m!}\int\cdots\int_{\Theta_{m+r-1}(\tau)}\frac{\mathrm{d}u_1}{2u_1}\cdots\frac{\mathrm{d}u_{m+r-1}}{2u_{m+r-1}}\frac{\mathrm{d}u_0}{\sqrt{u_0}}$$
$$= \frac{1}{2}\sum_{p=r-1}^{\infty}\frac{(-1)^{p-r+1}}{(r-1)!(p-r+1)!2^p}\int\cdots\int_{\Theta_p(\tau)}\frac{\mathrm{d}u_1}{u_1}\cdots\frac{\mathrm{d}u_p}{u_p}\frac{\mathrm{d}u_0}{\sqrt{u_0}}$$
$$= \frac{1}{2}\sum_{p=r-1}^{\infty}\frac{(-1)^{p-r+1}}{(r-1)!(p-r+1)!2^p}\int_0^\tau \mathrm{d}u\int\cdots\int_{\Theta_p^1(u)}\frac{1}{\sqrt{u-u_1-\cdots-u_p}}\frac{\mathrm{d}u_1}{u_1}\cdots\frac{\mathrm{d}u_p}{u_p},$$

where

$$\Theta_p(x) = \{(u_1,\ldots,u_p,u_0)\colon u_1\geqslant 1,\ldots,u_p\geqslant 1, u_0\geqslant 0, u_1+\cdots+u_p+u_0\leqslant x\},$$
$$\Theta_p^1(x) = \{(u_1,\ldots,u_p)\colon u_1\geqslant 1,\ldots,u_p\geqslant 1, u_1+\cdots+u_p\leqslant x\},\quad p\geqslant 1.$$

Hence we have

$$g(\tau) = \frac{1}{2}\sum_{p=r-1}^{\infty}\frac{(-1)^{p-r+1}}{(r-1)!(p-r+1)!2^p}\int\cdots\int_{\Theta_p^1(\tau)}\frac{1}{\sqrt{\tau-u_1-\cdots-u_p}}\frac{\mathrm{d}u_1}{u_1}\cdots\frac{\mathrm{d}u_p}{u_p}$$

and with substitution $\tau = \frac{1}{t}-1$, we obtain the asymptotic density of L_r/n

$$f_r(t) = g\left(\frac{1}{t}-1\right)t^{-3/2}$$
$$= t^{-3/2}\frac{1}{2}\sum_{p=r-1}^{p<1/t-1}\frac{(-1)^{p-r+1}}{(r-1)!(p-r+1)!2^p}\int\cdots\int_{\Theta_p^1(1/t-1)}\frac{1}{\sqrt{\frac{1}{t}-1-u_1-\cdots-u_p}}\frac{\mathrm{d}u_1}{u_1}\cdots\frac{\mathrm{d}u_p}{u_p}.$$

When $r=1$, $t>\frac{1}{2}$, we have a result of Stepanov [11]

$$f_1(t) = t^{-3/2}\frac{1}{2\sqrt{1/t-1}} = \frac{1}{2t\sqrt{1-t}}.$$

Remark. In Table 2 we list the numerical values concerning the expected size of largest component $n^{-1}\mathbf{E}_n(L_1)$ through Monte Carlo simulation given by

$$n^{-1}\mathbf{E}_n(L_1) \approx \frac{\text{sum of the size of largest component in } N \text{ times}}{n\cdot N}$$

with $N = 20000$.

Table 2
Calculation of $n^{-1}\mathbf{E}_n(L_1)$

$n=50$	$n=100$	$n=150$	$n=200$	$n=250$	$n=300$	$n=350$	$n=400$
0.7811	0.7724	0.7672	0.7673	0.7659	0.7650	0.7645	0.7656
$n=450$	$n=500$	$n=550$	$n=600$	$n=650$	$n=700$	$n=750$	$n=800$
0.7627	0.7659	0.7644	0.7652	0.7609	0.7617	0.7634	0.7634
$n=80$	$n=900$	$n=950$	$n=1000$	$n=1050$	$n=1100$	$n=1150$	$n=1200$
0.7623	0.7644	0.7634	0.7618	0.7648	0.7655	0.7606	0.7608
$n=1250$	$n=1300$	$n=1350$	$n=1400$	$n=1450$	$n=1500$	$n=1550$	$n=1600$
0.7618	0.7627	0.7623	0.7621	0.7605	0.7620	0.7612	0.7598
$n=1650$	$n=1700$	$n=1750$	$n=1800$	$n=1850$	$n=1900$	$n=1950$	$n=2000$
0.7611	0.7616	0.7614	0.7620	0.7603	0.7598	0.7612	0.7603

Acknowledgements

The authors would like to thank the referee for his corrections and very useful suggestions. Actually some of his comments is used in the introduction.

References

[1] D.J. Aldous, G. Miermont, J. Pitman, Brownian bridge asymptotics for random p-mappings, Electron. J. Probab. 9 (2004) 37–56.
[2] D.J. Aldous, J. Pitman, Brownian bridge asymptotics for random mappings, Random Structures Algorithms 5 (1994) 487–512.
[3] W. Feller, An Introduction to Probability Theory and its Applications, vol. I, third ed., John Wiley & Sons, 1968.
[4] I.B. Gertsbakh, Epidemic processes on a random graph: some preliminary results, J. Appl. Probab. 14 (1977) 427–438.
[5] V.F. Kolchin, Random Mappings, Optimization Software, New York, 1986 (Translation of Russian original).
[6] J. Pitman, Combitorial Stochastic Processes, Lecture Notes in Math., Springer, Berlin, in press. Available now via http://bibserver.berkeley.edu/csp/csp.html.
[7] B. Pittel, On distributions related to transitive closures of the random finite mappings, Ann. Probab. 11 (1983) 428–441.
[8] S. Ramanujan, Question 294, J. Indian Math. Soc. 3 (1911) 128.
[9] S.M. Ross, Introduction to Probability Models, seventh ed., Academic Press, 2000.
[10] L.A. Shepp, S.P. Lloyd, Ordered cycle lengths in random permutations, Trans. Amer. Math. Soc. 121 (1966) 340–357.
[11] V.E. Stepanov, Limit distributions of certain characteristics of random mappings, Theory Probab. Appl. 14 (1969) 612–626.
[12] V.E. Stepanov, Random mappings with a single attracting centre, Theory Probab. Appl. 16 (1971) 155–161.