ISRAEL JOURNAL OF MATHEMATICS 147 (2005), 221–243

COMPOSITIONS OF RANDOM TRANSPOSITIONS

BY

ODED SCHRAMM

Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA

In loving memory of my parents, Hanna and Mickey Schramm

ABSTRACT

Let $Y = (y_1, y_2, \ldots)$, $y_1 \geq y_2 \geq \cdots$, be the list of sizes of the cycles in the composition of cn transpositions on the set $\{1, 2, \ldots, n\}$. We prove that if $c > 1/2$ is constant and $n \to \infty$, the distribution of $f(c)Y/n$ converges to $PD(1)$, the Poisson–Dirichlet distribution with parameter 1, where the function f is known explicitly. A new proof is presented of the theorem by Diaconis, Mayer-Wolf, Zeitouni and Zerner stating that the $PD(1)$ measure is the unique invariant measure for the uniform coagulation-fragmentation process.

1. Introduction

Consider the composition $\pi_t = T_t \circ T_{t-1} \circ \cdots \circ T_2 \circ T_1$ of random, uniform, independent traspositions T_j of $V := \{1, 2, \ldots, n\}$. How large must t be in order for π_t to "look like" a random-uniform permutation π of V? As we will see, the answer depends on the precise meaning given to the term "look like".

It is easy to check that $\mathbf{P}[\pi(v) = v] = 1/n$ for all $v \in V$. Therefore, the expected number of fixed points of π is 1. However, if v does not appear in any of the transpositions $T_1, T_2, \ldots, T_t$, then $\pi_t(v) = v$. By the familiar solution of the coupon collector's problem, we see that when $t = o(n \log n)$, the probability that π_t has at most one fixed point is small. In this sense, π_t and π are rather different when $t = o(n \log n)$. On the other hand, when $t > cn \log n$, $c > 1/2$, the total variation distance between the law of π_t and that of π tends to zero as $n \to \infty$ [DS81].

We now consider the situation where $t \leq cn$ with $c < 1/2$. Let G^t be the graph on $V = \{1, 2, \ldots, n\}$ where $\{v, u\}$ is an edge in G^t if and only if the

Received May 11, 2004

I. Benjamini, O. Häggström (eds.), *Selected Works of Oded Schramm*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4419-9675-6_17, © Springer Science+Business Media, LLC 2011

transposition (v, u) appears in $\{T_1, \ldots, T_t\}$. Let V_G^t denote the set of vertices of the largest connected component of G^t (with arbitrary tie breaking if there is more than one). By the Erdős–Rényi Theorem, when $t \leq cn$, $c < 1/2$, we have $|V_G^t| = O(\log n)$ asymptotically almost surely (a.a.s.). It follows that the largest cycle (orbit) of π_t is also of size $O(\log n)$. When $c = 1/2$, the same holds, but with $\log n$ replaced by any function growing faster than $n^{2/3}$. This contrasts with the fact that for every $k \in \{1, 2, \ldots, n\}$ the probability that the cycle of π containing 1 has size $\leq k$ is precisely k/n. Thus, π_t is very different from π when $t/n \leq c \leq 1/2$.

Our main theorem deals with the case where $t/n \geq c > 1/2$. Confirming a conjecture by Aldous, we prove that in this range, the large cycles of π_t, when normalized by their total length, have a distribution that is close to that of the large cycles of π. A more precise statement of this result will be given shortly.

Let σ be some permutation of V. Let $X(\sigma)$ denote the set of cycles (orbits) of elements of V under σ. The **cycle structure** $\mathfrak{X}(\sigma)$ is then the sorted list of the lengths of the cycles, that is, the list $(|C| : C \in X(\sigma))$ sorted in nonincreasing order. Thus, $\mathfrak{X}_i(\sigma)$ denotes the size of the i-th largest cycle of σ. If i is larger than the number of cycles of σ, then we set $\mathfrak{X}_i(\sigma) = 0$, by convention. Since each T_j is chosen uniformly among the transpositions, it follows that for each fixed permutation σ of V the distribution of π_t is the same as that of $\sigma \circ \pi_t \circ \sigma^{-1}$. Thus, the distribution of π_t is determined by the distribution of the conjugacy class of π_t. Now, the conjugacy class of π_t is determined by $\mathfrak{X}(\pi_t)$. Consequently, the distribution of $\mathfrak{X}(\pi_t)$ determines the distribution of π_t.

We are now ready to state our main theorem, which gives a positive answer to a conjecture by David Aldous as stated in [BD].

THEOREM 1.1: *Let $c > 1/2$, and take $t \geq cn$. As $n \to \infty$, the law of $\mathfrak{X}(\pi_t)/|V_G^t|$ converges weakly to the Poisson–Dirichlet distribution $PD(1)$ with parameter 1 (which is defined below).*

A more explicit statement of the theorem is as follows. Given $c > 1/2$ and $\epsilon > 0$, there is an $n(c, \epsilon)$ such that for every $n > n(c, \epsilon)$ and every $t \geq cn$ there is a coupling of the sequence of transpositions T_j and a $PD(1)$ sample Y such that

$$\mathbf{P}[\|Y - \mathfrak{X}(\pi_t)/|V_G^t|\|_\infty < \epsilon] > 1 - \epsilon.$$

Weak convergence has several equivalent formulations (see [Dud89]), and we have opted to use the coupling version here.

The $PD(1)$ distribution is a probability measure on the infinite dimensional

simplex

$$\Omega := \{y \in [0,1]^{\mathbb{N}_+} : \sum_i y_i = 1, y_1 \geq y_2 \geq \cdots\}$$

and may be defined as follows. Let $U_1, U_2, \ldots$ be an i.i.d. sequence of random variables uniformly distributed in $[0,1]$. Set $x_1 := U_1$ and inductively $x_j := U_j(1 - \sum_{i=1}^{j-1} x_i)$. Let (y_i) be the sequence (x_i) sorted in nonincreasing order. The $PD(1)$ distribution is defined as the law of (y_i). See, for example, [Hol01] for other definitions and a discussion of some of the properties of the Poisson–Dirichlet distributions.

The behaviour of the size of the largest cluster of G^t, which is the normalizing quantity $|V_G^t|$ in the theorem, is known precisely. The Erdős–Rényi theorem (see, e.g., [AS00]) tells us that

$$|V_G^t|/n \to z(2t/n) \tag{1.1}$$

in probability as $n \to \infty$, where $z(s)$ is the survival probability of a Galton–Watson branching process with offspring distribution which is the Poisson random variable with mean s. Moreover, $z(s)$ is the positive solution of the equation $1 - z = \exp(-sz)$ when $s > 1$ and $z(s) = 0$ for $s \in [0,1]$.

Berestycki and Durrett [BD] have analysed other aspects of the chain π_t which exhibit a phase transition near $t = n/2$: they investigate the minimal number of transpositions necessary to write π_t as a composition.

In [DMP95], Diaconis, McGrath and Pitman discuss the Riffle shuffle, which is another example where the large cycles appear relaxed well before the permutation is uniformly distributed.

The evolution of $\mathfrak{X}(\pi_t)$ is also known as the discrete uniform coagulation-fragmentation process. Let us briefly describe the transition from $\mathfrak{X}(\pi_t)$ to $\mathfrak{X}(\pi_{t+1})$. Suppose that T_{t+1} is the transposition (a,b). Then a and b are selected uniformly from V, and are "almost independent". (We could also allow $a = b$; then $T = (a,a)$ would be the identity transposition, and a and b would be independent. That would not change anything significant in the following.) Let $X_i, X_j \in X(\pi_t)$ satisfy $a \in X_i$, $b \in X_j$. Then X_i and X_j are size biased selections from $X(\pi_t)$, and are nearly independent given π_t. If $X_i \neq X_j$, then in π_{t+1} the two cycles X_i and X_j are replaced by the single cycle whose vertices are $X_i \cup X_j$. If $X_i = X_j$, then this cycle splits into two cycles of π_{t+1}. If $k = |X_i|$ and $m \in \mathbb{N}_+$ is the least positive integer satisfying $\pi_t^m(a) = b$, then the resulting two cycles of π_{t+1} are $(a, \pi_t(a), \ldots, \pi_t^{m-1}(a))$ and $(b, \pi_t(b), \ldots, \pi_t^{k-m-1}(b))$. Note that given X_i and given $X_i = X_j$, the resulting two new cycles have sizes m and $|X_i| - m$, where m is chosen uniformly in $\{1, 2, \ldots, |X_i| - 1\}$.

There is a similar continuous coagulation-fragmentation process, which is a discrete time Markov chain on the infinite dimensional simplex Ω. The transition kernel M of the chain operates as follows. Given $Y = (Y_1, Y_2, \ldots) \in \Omega$, we choose two indices $i, j \in \mathbb{N}_+$ independently, with $\mathbf{P}[i = k|Y] = \mathbf{P}[j = k|Y] = Y_k$. If $i \neq j$, then let Y' be obtained from Y by replacing the two entries Y_i and Y_j with the single entry $Y_i + Y_j$ and resorting. If $i = j$, then given (Y, i, j), a random variable v is selected uniformly in $[0, Y_i]$ and Y' is obtained by splitting the entry Y_i into the two entries v and $Y_i - v$, and resorting. Then Y' is the new state of the Markov chain.

It is known that the probability measure $PD(1)$ is invariant under M. Apparently, this was first proved in [Wat76]; references for several other proofs of this fact are given in [DMWZZ]. Vershik conjectured that $PD(1)$ is the only invariant measure. Subsequently, this was proved by Diaconis, Mayer-Wolf, Zeitouni and Zerner:

THEOREM 1.2 ([DMWZZ]): *The invariant measure for M is unique.*

See [DMWZZ] for more information and bibliography regarding the history of the problem, including some earlier established special cases.

The proof of [DMWZZ] relies on coupling the discrete and the continuous coagulation-fragmentation processes, and using representation theory on the symmetric group to understand the discrete process. In the present paper, we use a different coupling to handle the continuous process directly, and thereby give a different proof of Theorem 1.2. Moreover, a slight modification of this coupling will be essential in the proof of Theorem 1.1.

The problems addressed in this paper are a mean-field version of a statistical physics model suggested by Tóth [Tót93], which may be described as follows. Consider a locally finite graph $G = (V, E)$, and fix a parameter $\beta > 0$. For each (unoriented) edge $e \in E$, let $Z_e \subset [0, 1]$ be an independent Poisson point process of intensity β on $[0, 1]$. Let $v_0 \in V$. We now describe a walk $v(t)$ starting at $v(0) = v_0$. Let t_1 be the first $t > 0$ such that there is an edge $e_1 = [v_0, v_1]$ incident with v_0 such that $t_1 \in Z_{e_1} + \mathbb{Z}$. If there is no such t_1, then $v(t) = v_0$ for all $t \geq 0$. But if t_1 exists, then let $v(t) = v_0$ for $t \in [0, t_1)$ and $v(t_1) = v_1$. Inductively, assume that t_j and v_j are defined and $v(t_j) = v_j$. Let t_{j+1} be the first $t > t_j$ such that there is an edge $e_{j+1} = [v_j, v_{j+1}]$ incident with v_j such that $t \in Z_{e_{j+1}} + \mathbb{Z}$; set $v(t) = v_j$ for $t \in (t_j, t_{j+1})$ and $v(t_{j+1}) = v_{j+1}$.

In the case where G is the complete graph on V, it is easy to see that the orbit of 1 in π_t is analogous to the range of this walk starting at 1, where $\beta = t/n$. The essential difference between the two is the distinction between continuous

time and discrete time.

There are several known open problems regarding Tóth's model. Is it true that for (connected) bounded degree graphs G, the simple random walk on G is transient iff Tóth's walk v_j visits infinitely many vertices with positive probability for some $\beta > 0$? In particular, is this true for $G = \mathbb{Z}^d$? For finite graphs G, one may ask about the distribution of the size of the image of the walk $\{v_j\}$, for example. See [Ang03] for an analysis of Tóth's model on regular trees and for a list of some open problems, including those mentioned above.

Returning to the symmetric group, one may ask about the typical cycle structure near the transition point $t = n/2$. A very thorough analogous theory exists for the Erdős–Rényi transition. See, for example, [Spe94, As00, JLR00] and the references cited there.

NOTATIONS. For the convenience of the reader, we list here some of the notations used extensively, with hyperlinks and page numbers of the definitions, and a brief description, where appropriate.

V	$\{1,2,\ldots,n\}$	221
$T_1, T_2, \ldots$	i.i.d. uniform transpositions on V	221
π_t	$T_t \circ T_{t-1} \circ \cdots \circ T_1$	221
$X(\sigma)$	set of cycles of a permutation σ	222
X^s	$X(\pi_s)$	226
$\mathfrak{X}(\sigma)$	cycle structure of σ	222
$X^s(v)$	cycle of π_s containing v	226
$V_X^s(k)$	union of cycles of π_s of size at least k	226
G^t	graph whose edges correspond to transpositions $T_i, i \le t$	222
V_G^t	largest cluster in G^t	222
$V_G^t(k)$	union of clusters of G^t of size at least k	226
$z(s)$	function in the Erdős–Rényi theorem	223
Ω	$\{y \in [0,1]^{\mathbb{N}_+} : \sum_i y_i = 1, y_1 \ge y_2 \ge \cdots\}$	223
$PD(1)$	Poisson–Dirichlet distribution with parameter 1	223
M	coagulation-fragmentation transition kernel	224
$\tilde{M}$	the coupling	230
$I(Y,Z)$	indexes of matched entries	230
Q	sum of matched entries	232
$\tilde{Y}, \tilde{Z}, \hat{Y}, \hat{Z}$	partitions used in defining $\tilde{M}$	230
u, v	random variables used in the definition of $\tilde{M}$	230
$\bar{\epsilon}$	ϵ + fragments smaller than ϵ	233
N^t	unmatched entries larger than ϵ	233

y_1^t, z_1^t largest unmatched entries in Y^t and Z^t 233

2. Big pieces

The main goal of the present section is to show in a quantitative way that most vertices in V_G^t are in reasonably large cycles of π_t.

Suppose that π is a permutation on V and $T = (x, y)$ a transposition. If x and y are in different cycles in π, then in $T \circ \pi$ these two cycles are joined, and the other cycles remain unchanged. Now suppose that $C = (x_0, x_1, \ldots, x_m)$ is a cycle of π which contains x and y. Say, $x = x_j$, $y = x_i$, and $j < i$. Then in $T \circ \pi$ the cycle C is split into the cycles $(x_i, x_{i+1}, \ldots, x_{j-1})$ and $(x_j, x_{j+1}, \ldots, x_m, x_0, x_1, \ldots, x_{i-1})$. The other cycles remain unchanged, of course. This clearly implies the following

LEMMA 2.1: *Let π be a permutation of V and $s \in \mathbb{N}$. Let T be a uniform-random transposition on V. Then the probability that some cycle of π is split in $T \circ \pi$ into two cycles at least one of which has length $\leq s$ is at most $2s/(n-1)$.*

This will be used in the next lemma. Let $X^s = X(\pi_s)$ be the set of cycles of π_s and for $v \in V$ let $X^s(v)$ be the cycle in X^s containing v. Let $V_G^s(k) \subset V$ be the union of those connected components of G^s which have at least k vertices, and let $V_X^s(k) \subset V$ be the union of the cycles in X^s that have at least k vertices.

LEMMA 2.2:

$$\mathbf{E}|V_G^s(k) \setminus V_X^s(k)| \leq 4sk^2/(n-1)$$

holds for every $k, s \in \mathbb{N}$.

Proof: Let I be the set of $t \in \mathbb{N}$ such that there is a cycle $A \in X^{t-1}$ which splits into two nonempty cycles in X^t, $A = A_1 \cup A_2$, $A_1, A_2 \in X^t$ and at least one of these cycles, say A_1, satisfies $|A_1| \leq k$. The above lemma shows that $\mathbf{P}[t \in I] \leq 2k/(n-1)$ for every $t \in \mathbb{N}$, and hence $\mathbf{E}[|I \cap [0, s]|] \leq 2sk/(n-1)$.

Suppose that $C \in X^s$, $|C| < k$ and $C \subset V_G^s(k)$. There must be some vertex $u \in C$ and some time $t \leq s$ such that $|X^t(u)| < |X^{t-1}(u)|$; otherwise, C would be equal to a component of G^s. Among all such possible pairs (u, t), we choose one that maximizes t. Then we have $X^t(u) \subset C$. Consequently, $t \in I \cap [0, s]$ and at least one of the two elements of V transposed by T_t is in C. Therefore, the number of such C is at most $2|I \cap [0, s]|$. The statement of the lemma now follows from the above bound on $\mathbf{E}|I \cap [0, s]|$. ∎

The following lemma will tell us that if X^t has many vertices in reasonably large cycles at time $t = t_0$, then with high probability at a specified later time t_1 most of these vertices will be in cycles of size at least ϵn.

LEMMA 2.3: *Let $\delta \in (0,1]$, $t_0, j \in \mathbb{N}$, and $\epsilon \in (0, 1/8)$. (The lemma will be useful primarily when $(\log n)^2 \le 2^j \le n^\alpha$ with any constant $\alpha < 1/2$.) Assume that $2^j < \epsilon\delta n$ and that $\mathbf{P}[|V_X^{t_0}(2^j)| > \delta n] > 0$. Set $\rho := 2^j/n$ and*

$$t_1 := t_0 + \lceil 2^6 \delta^{-1} \rho^{-1} \log_2(\rho^{-1}) \rceil. \tag{2.1}$$

Then the number of vertices v that are in cycles of size at least 2^j at time t_0 but are not in cycles of size at least $\epsilon\delta n$ at time t_1 satisfies

$$\mathbf{E}[|V_X^{t_0}(2^j) \setminus V_X^{t_1}(\epsilon\delta n)| \,|\, |V_X^{t_0}(2^j)| > \delta n] \le O(1)\delta^{-1}\epsilon |\log(\epsilon\delta)| n, \tag{2.2}$$

where the constant implied in the $O(1)$ notation is universal.

Two important aspects of this lemma are that the right hand side of (2.2) does not depend on j and that t_1 does not depend on ϵ. (However, $t_1 - t_0$ depends primarily on j and the right hand side of (2.2) depends primarily on ϵ.)

Before we begin with the actual proof, here is an informal outline. Let $v \in V_X^{t_0}(2^j)$. Set $K := \lceil \log_2(\epsilon\delta n) \rceil$. We will choose a sequence of times $\tau_j, \tau_{j+1}, \ldots, \tau_K$. For $s = j, j+1, \ldots, K$, when $t \in [\tau_s, \tau_{s+1})$ we will "expect" the size of $X^t(v)$ to be at least 2^s. This can fail in either of two scenarios: it may happen because a transposition cuts the cycle of v, or it may happen because no transposition merges the cycle of v with a sufficiently large cycle. The probabilities for each of these unfortunate situations will be appropriately estimated. The choice of the time interval $\tau_{s+1} - \tau_s$ is somewhat delicate. If it is too long, then perhaps too many cycles will be cut, while if it is too short, then cycles will not have enough time to merge. It turns out that

$$a_s := 2^4 \delta^{-1} 2^{-s} (\lceil \log_2 n \rceil - s)(n-1)$$

is roughly the right choice, as will become clear in the course of the proof.

Proof: Within the proof below, expectations and probabilities will be conditioned on $|V_X^{t_0}(2^j)| > \delta n$. Let $K := \lceil \log_2(\epsilon\delta n) \rceil$, and let a_s be as above. For $s > j$ let $m_s := \lceil a_s \rceil$ and set $m_j := t_1 - t_0 - \sum_{s=j+1}^{K-1} m_s$. Set $\tau_s := t_0 + \sum_{i=j}^{s-1} m_i$. Note that $\tau_K = t_1$ and $a_s \le m_s \le O(a_s)$ for $s = j, j+1, \ldots, K-1$.

Let $s \in \{j, j+1, \ldots, K-1\}$ and $t \in \{\tau_s + 1, \tau_s + 2 \ldots, \tau_{s+1}\}$. Define $F^t \subset V$ to be the set of vertices $v \in V$ such that $|X^t(v)| < |X^{t-1}(v)|$ and $|X^t(v)| < 2^{s+1}$.

Lemma 2.1 shows that $\mathbf{E}|F^t| \le 2^{2s+4}/(n-1)$. We also set $\tilde{F}^t := \bigcup_{\tau=t_0+1}^{t} F^\tau$. Then

$$\mathbf{E}|\tilde{F}^{t_1}| \le \sum_{s=j}^{K-1} m_s 2^{2s+4}/(n-1) = O(\epsilon|\log(\epsilon\delta)|n).$$

We consider the vertices in F^t as vertices "failing" at time t. However, there are other ways in which vertices can fail. If at time $t \in \{\tau_s, \tau_s+1, \dots, \tau_{s+1}-1\}$ we have $|V_X^t(2^s)| < \delta n/2$, then we consider the whole process as failed, and we set $H^t := V$. Otherwise, take $H^t = \emptyset$. Also set $\tilde{H}^t := \bigcup_{t'=t_0}^{t} H^{t'}$.

The third and last way in which a vertex v may fail is if $X^t(v)$ does not grow in time. Let

$$B^s := V_X^{\tau_s}(2^s) \setminus (\tilde{F}^{\tau_{s+1}} \cup \tilde{H}^{\tau_{s+1}-1} \cup V_X^{\tau_{s+1}}(2^{s+1})),$$

and $\tilde{B}^s := \bigcup_{k=j}^{s} B^k$. The vertices in B^s are vertices whose cycles failed to grow sufficiently between time τ_s and time τ_{s+1}. It is clear that

$$V_X^{t_0}(2^j) \subset V_X^{t_1}(\epsilon\delta n) \cup \tilde{H}^{t_1} \cup \tilde{F}^{t_1} \cup \tilde{B}^K. \tag{2.3}$$

If $v \in B^s$, then it must be the case that for every $t \in [\tau_s, \tau_{s+1}-1]$ we have $|V_X^t(2^s)| \ge \delta n/2$ (since B^s is disjoint from $\tilde{H}^{\tau_{s+1}-1}$) and $v \in V_X^t(2^s) \setminus V_X^t(2^{s+1})$ (since B^s is disjoint from $\tilde{F}^{\tau_{s+1}}$). If we condition on $2^s \le |X^t(v)| < 2^{s+1}$ and on $|V_X^t(2^s)| > \delta n/2$, then there is probability at least

$$2^s(\delta n/2 - 2^{s+1})\binom{n}{2}^{-1} \ge 2^{s-3}\delta(n-1)^{-1}$$

that T_{t+1} transposes an element from $X^t(v)$ and an element from some other cycle of X^t whose size is at least 2^s. If that happens, then $v \in V_X^t(2^{s+1})$ and this implies that v cannot be in B^s. Consequently,

$$\begin{aligned}\mathbf{P}[v \in B^s] &\le (1 - 2^{s-3}\delta/(n-1))^{m_s} \\ &\le \exp(-2^{s-3}\delta m_s/(n-1)) \le O(2^s/n).\end{aligned}$$

Hence,

$$\mathbf{E}|\tilde{B}^K| \le O(1)n2^K n^{-1} = O(\epsilon\delta n).$$

It follows from the definition of H^t that in order for H^t to be nonempty, we must have $|\tilde{F}^t \cup \tilde{B}^{s-1}| \ge \delta n/2$. Therefore,

$$\mathbf{E}|\tilde{H}^{t_1}| \le n\mathbf{P}[|\tilde{F}^{t_1} \cup \tilde{B}^K| \ge \delta n/2] \le 2\delta^{-1}\mathbf{E}|\tilde{F}^{t_1} \cup \tilde{B}^K|.$$

When we combine this with (2.3) and the above estimates for $\mathbf{E}|\tilde{F}^{t_1}|$ and $\mathbf{E}|\tilde{B}^K|$, the lemma follows. ∎

LEMMA 2.4: *Fix some $c > 1/2$, and let $t \geq cn$, $t \in \mathbb{N}$. Let $\epsilon, \alpha \in (0, 1/8)$ and let N be the minimal number of cycles in X^t which cover at least $(1-\epsilon)|V_G^t|$ vertices of V_G^t. Then*

$$\mathbf{P}[N > \alpha^{-1} |\log(\alpha\epsilon)|^2] \leq C_1 \alpha$$

for all $n > n_1$, where C_1 is a constant which depends only on c, and n_1 may depend on c and ϵ.

Proof: First, suppose that $t \leq n^{5/4}$. Choose j such that $n^{1/4} \leq 2^j < 2n^{1/4}$. Let $\delta = z/2$, where $z = z(2t/n)$ is the Galton–Watson survival probability discussed in the introduction. Choose t_0 so that (2.1) holds with t in place of t_1. Note that $t - t_0 = O(n^{3/4} \log n)$. (Here and below, the constants in the $O(\cdot)$ notation may depend on c.) We apply the Erdős–Rényi theorem at time t_0 to conclude that a.a.s. $|V_G^{t_0}| - nz = o(n)$ and the second largest component of G^{t_0} has size less than $(\log n)^2$. Lemma 2.2 with $k = 2^j$ and $s = t_0$ implies that $|V_G^{t_0} \setminus V_X^{t_0}(2^j)| \leq n^{7/8}$ a.a.s. Note also that $|V_G^t \setminus V_G^{t_0}| \leq (t - t_0) O(\log n)^2$ a.a.s., because we know that the components of G^{t_0} other than the largest one are typically smaller than $(\log n)^2$. Hence $|V_G^t \setminus V_X^{t_0}(2^j)| < n^{7/8}$ a.a.s. Now, Lemma 2.3 implies that for every fixed $\epsilon' > 0$ and for every sufficiently large n

$$\mathbf{E}[|V_G^t \setminus V_X^t(\epsilon' n)|] < O(1)\epsilon' |\log \epsilon'| n. \tag{2.4}$$

(Note that $|V_G^{t_0}| \leq n$, and hence the conditioning in (2.2) may be ignored once n is large enough so that $\mathbf{P}[|V_X^{t_0}(2^j)| \leq \delta n] < \epsilon' |\log \epsilon'|$.)

Now, to show that (2.4) holds also without the assumption that $t \leq n^{5/4}$, we note that Lemma 2.3 may be applied with $j = 0$, $\delta = 1$ and t_0 chosen so that (2.1) holds with t in place of t_1. (In this case, we do not need to use Lemma 2.2.)

Set $a(k) := |V_G^t \setminus V_X^t(k)|$. Let i_0 be the smallest integer i such that $a(2^{-i}n) < \epsilon n/2$. Then N is bounded by the number of cycles in $V_X^t(2^{-i_0}n) \cap V_G^t$. Let i_1 be the least integer such that $2^{-i_1} < \alpha\epsilon / |\log(\alpha\epsilon)|$. Then (2.4) shows that $\mathbf{P}[i_0 > i_1] = O(\alpha)$. We may write

$$a(k) = \sum \{|A| : A \subset V_G^t, A \in X^t, |A| < k\}.$$

By considering the contribution of each cycle to the sum

$$S_m := \sum_{i=0}^{m} a(2^{-i}n) 2^i / n,$$

we find that $N = O(S_{i_0})$. On the other hand, (2.4) implies that

$$\mathbf{E}[S_{i_1}] \leq O(1) \sum_{i=0}^{i_1} i \leq O(1)(i_1)^2 \leq O(1)|\log(\alpha\epsilon)|^2.$$

Because $\mathbf{P}[i_0 > i_1] = O(\alpha)$, this completes the proof. ∎

3. Coupling

At this point, it seems likely that the proof of Theorem 1.1 can be completed using some of the results from the work of Diaconis, Mayer-Wolf, Zeitouni and Zerner [DMWZZ]. However, we prefer instead to use a different coupling argument to finish off the proof and also prove the main result of [DMWZZ].

We now describe a coupling in the continuous setting. A similar coupling will also apply to couple between the discrete and continuous setting, but the purely continuous setting avoids several annoying minor notational issues.

The coupling is between two Markov chains Y^t and Z^t starting at possibly different initial starting points $Y^0, Z^0 \in \Omega$ with each separately evolving according to the transition kernel M.

In this coupling, the evolution of (Y^t, Z^t) will also be Markov. Its transition kernel will be denoted by $\tilde{M}$.

The basic idea in the construction of $\tilde{M}$ is that if we have entries in Y^t that are equal to entries in Z^t, then we don't want to ruin this. Consequently, if we make a change to such an entry in Y^t, we want to make a corresponding change to the corresponding entry in Z^t. On the other hand, as much as we can, we do want to produce new entries in Y^t and Z^t that match. Our measure of the discrepancy between Y^t and Z^t will roughly be the number of large unmatched entries, and we will strive to reduce the discrepancy.

In order to define $\tilde{M}$, we need some more notations. Let $(Y, Z) \in \Omega^2$. We will need to match entries in Y with entries in Z of the same length, if such exist, and match as many entries as possible. The matching will be encoded via maps $f_{Z,Y}, f_{Y,Z}\colon \mathbb{N}_+ \to \mathbb{N}$, which are defined as follows. Let $i \in \mathbb{N}_+$, let H be the set of $j \in \mathbb{N}_+$ such that $Y_i = Z_j$, and let $k := |\{j \in \mathbb{N} : j \leq i, Y_j = Y_i\}|$. (Partly because we want to easily generalize to the discrete setting, we do not want to rule out the possibility that $Y_i = Y_j$ for some $i \neq j$.) If $|H| < k$, then set $f_{Y,Z}(i) = 0$. Otherwise, let $f_{Y,Z}(i)$ be the k'th smallest element in H. (By exchanging Y and Z, this also defines the map $f_{Z,Y}$.) Let

$$I(Y, Z) := f_{Y,Z}^{-1}(\mathbb{N}_+) = \{i \in \mathbb{N} : f_{Y,Z}(i) \neq 0\}.$$

The entries Y_i with $i \in I(Y,Z)$ will be referred to as matched. Likewise, Z_j, $j \in I(Z,Y)$, are the matched entries of Z. Observe that $f_{Z,Y} \circ f_{Y,Z}(i) = i$ for every $i \in I(Y,Z)$, $f_{Y,Z}(I(Y,Z)) = I(Z,Y)$ and $Z_{f_{Y,Z}(i)} = Y_i$ for every $i \in I(Y,Z)$. Let

$$Q = Q(Y,Z) := \sum\{Y_i : i \in I(Y,Z)\} = \sum\{Z_j : j \in I(Z,Y)\}.$$

We will now describe the transition kernel $\tilde{M}$. Given $(Y,Z) \in \Omega$, we need to perform one step of M for each of Y and Z, thereby generating new configurations Y' and Z'. We associate with Y and with Z partitions $\tilde{Y} = (\tilde{Y}_i : i \in \mathbb{N}_+)$ and $\tilde{Z} = (\tilde{Z}_i : i \in \mathbb{N}_+)$ of $[0,1]$ into closed intervals, as follows. (See also Figure 3.1.) The length of the interval $\tilde{Y}_i$ is Y_i. The intervals $\tilde{Y}_i$ with $i \in I(Y,Z)$ tile the interval $[1-Q,1]$, while the intervals with $i \notin I(Y,Z)$ tile the interval $[0,1-Q]$. Within each of these classes, let the intervals be ordered according to the indices; that is $\max \tilde{Y}_i \le \min \tilde{Y}_{i'}$ if $i < i'$ when $i,i' \in I(Y,Z)$ and when $i,i' \notin I(Y,Z)$. A partition $\tilde{Z} = (\tilde{Z}_j : j \in \mathbb{N}_+)$ is constructed in the same way. Note that necessarily $\tilde{Z}_{f_{Y,Z}(i)} = \tilde{Y}_i$ whenever $i \in I(Y,Z)$.

Let u and v be two independent uniform random variables in $[0,1]$. Let $a, a' \in \mathbb{N}_+$ be the indices satisfying $u \in \tilde{Y}_a$ and $u \in \tilde{Z}_{a'}$. In this way, u induces a size biased sample from Y and from Z. We will use v to induce a different size biased sample, based on different tilings of $[0,1]$. Let $\hat{Y}$ be the tiling $(\hat{Y}_i : i \in \mathbb{N}_+)$ of $[0,1]$ by intervals that is obtained from $\tilde{Y}$ by shifting the interval $\tilde{Y}_a$ to the beginning. (That is, $\hat{Y}_a = [0, Y_a]$, $\hat{Y}_i = Y_a + \tilde{Y}_i$ if $\max \tilde{Y}_i \le \min \tilde{Y}_a$ and $\hat{Y}_i = \tilde{Y}_i$ if $\max \tilde{Y}_a \le \min \tilde{Y}_i$.) Similarly, $\hat{Z}$ is the tiling obtained from $\tilde{Z}$ by shifting $\tilde{Z}_{a'}$ to the beginning. Note that $\hat{Z}_{f_{Y,Z}(i)} = \hat{Y}_i$ whenever $i \in I(Y,Z)$.

Let b and b' be the indices satisfying $v \in \hat{Y}_b$ and $v \in \hat{Z}_{b'}$. If $a \neq b$, let Y' be obtained from Y by replacing the two entries Y_a and Y_b by the single entry $Y_a + Y_b$ and resorting. If $a = b$, let Y' be obtained from Y by replacing Y_a with the two entries v and $Y_a - v$ and resorting. Similarly, if $a' \neq b'$, let Z' be obtained from Z by replacing the two entries $Z_{a'}$ and $Z_{b'}$ by the single entry $Z_{a'} + Z_{b'}$ and resorting. If $a' = b'$, let Z' be obtained from Z by replacing $Z_{a'}$ with the two entries v and $Z_{a'} - v$ and resorting. This completes the construction of the Markov transition kernel $\tilde{M}$.

Let us observe a few essential features of this coupling. If Y_i and Z_j are split, then one of the two new entries in each of Y' and Z' is equal to v. If $i \in I(Y,Z)$ and Y_i is split or merged, then the same happens to $Z_{f_{Y,Z}(i)}$. Similarly, if

$j \in I(Z,Y)$ and Z_j is split or merged, then the same happens to $Y_{f_{Z,Y}(j)}$.

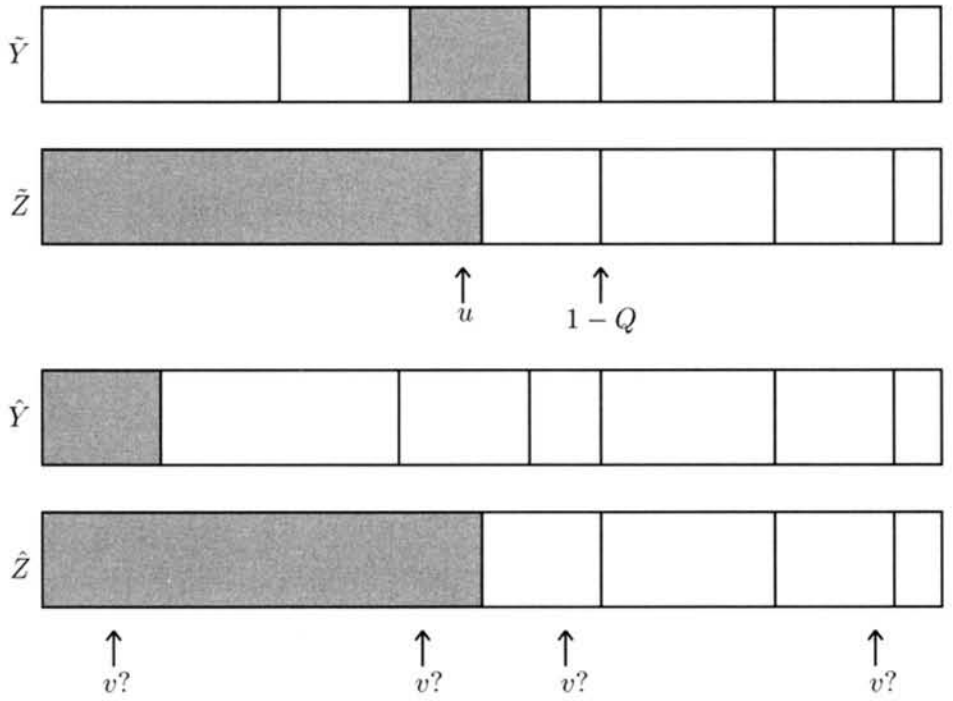

Figure 3.1. The random variable u chooses a segment in $\tilde{Y}$ and a segment in $\tilde{Z}$. The different illustrated choices for the random variable v yield a split in Y and in Z, a merge in Y and a split in Z, a merge in both where a matched segment is not involved, and merges involving matched segments, respectively.

We first informally describe the general behaviour of $\tilde{M}$, postponing the exact statements and proofs. When there are several unmatched reasonably large entries in Y^t and in Z^t, these merge and become few quite quickly. However, when they are very few, it is hard for them to dissappear completely. Suppose that there is one large unmatched entry in Y^t and two unmatched entries in Z^t. When the two unmatched entries in Z^t are merged, the single unmatched entry in Y^t is likely to be split. Thus, the situation does not improve so quickly. There is a parity phenomenon here: if the number of positive entries in Y^t is finite, then its parity either stays the same as that of t, or is opposite to that of t. Even if the number of positive entries is infinite, if it takes a long time for the smaller entries to be hit, the larger entries appear to follow this parity periodicity. One way to handle the parity issue would be to introduce

a delay to either Y^t or Z^t, but not both, in order to match up their parities. However, another phenomenon will be used instead. An unmatched entry in Y^t often splits into one matched entry and one unmatched entry. With any luck, the unmatched entry might be rather small. Thus, large unmatched entries are replaced by small unmatched entries. In effect, there is a diffusion of unmatched entries between different scales. Because of this, it is eventually unlikely to find a large unmatched entry, which is what we want to prove. However, this latter process is much slower than the first stage where large unmatched entries merge and become fewer. Thus, in time t the largest unmatched entry one can expect to find is of order roughly $1/\log t$.

Define

$$N_\epsilon(Y,Z) := |\{i \in \mathbb{N}_+ \setminus I(Y,Z) : Y_i > \epsilon\}|.$$

This is the number of entries in Y that are not matched by entries in Z and have size larger than ϵ.

LEMMA 3.1: *Let $\epsilon > 0$, and let $Y^0, Z^0 \in \Omega$. Let (Y^t, Z^t) be the Markov chain given by $\tilde{M}$ starting at (Y^0, Z^0). To abbreviate notations, set $N^t := N_\epsilon(Y^t, Z^t) + N_\epsilon(Z^t, Y^t)$, $Q^t = Q(Y^t, Z^t) = Q(Z^t, Y^t)$, $I^t := I(Y^t, Z^t)$ and $J^t := I(Z^t, Y^t)$. Also define*

$$\bar{\epsilon} := \epsilon + \sum\{Y_i^0 : Y_i^0 < \epsilon\} + \sum\{Z_i^0 : Z_i^0 < \epsilon\}.$$

Let $y_1^t := \max\{Y_i^t : i \notin I^t\}$ be the size of the largest unmatched entry of Y^t (set $y_1^t = 0$ if all entries are matched), and let z_1^t be the size of the largest unmatched entry of Z^t. Let q be a random variable with values in $\mathbb{N}$ which is independent from the evolution of the chain (Y^t, Z^t). Set

$$\eta := \max\{\mathbf{P}[q = t] : t \in \mathbb{N}\}.$$

Then

$$\mathbf{E}[(1 - Q^q)(1 - Q^q - \max\{y_1^q, z_1^q\})] \leq \frac{\eta}{2} N^0 + 4\bar{\epsilon}\mathbf{E}[q+1]. \tag{3.1}$$

When the right hand side in (3.1) is small, we know that with high probability either the sum of the unmatched entries in Y^q is only slightly larger than the largest unmatched entry, or this is true for Z^q.

Proof: Let $\mathcal{A}_s$ be the event that up to time s in every merging occurring both merged pieces are of size at least ϵ and in every splitting both resulting pieces are of size at least ϵ. Let $\mathcal{F}_s$ be the σ-field generated by $((Y^t, Z^t) : t = 0, 1, \ldots, s)$.

Conditioned on $\mathcal{F}_{t-1}$, the probability that at time t there is a split in any Y_i^{t-1} and one of the pieces is of size less than ϵ is at most 2ϵ. Conditioned on $\mathcal{F}_{t-1}$, the probability that there is any Y_i^{t-1} with $Y_i^{t-1} < \epsilon$ that is merged at time t with some other Y_j^{t-1} is at most $2\sum\{Y_i^{t-1} : Y_i^{t-1} < \epsilon\}$. Similar considerations apply to Z^t. Consequently, for $t \in \mathbb{N}_+$,

$$\mathbf{P}[\neg \mathcal{A}_t | \mathcal{A}_{t-1}, \mathcal{F}_{t-1}] \leq 4\bar{\epsilon}. \tag{3.2}$$

We now study the evolution of the quantity N^t, and consider several different cases for the transition from (Y^t, Z^t) to (Y^{t+1}, Z^{t+1}). In each case we assume that $\mathcal{A}_{t+1}$ holds.

1. The transition involves splitting in Y^t and merging in Z^t. Suppose that Y_i^t is split and Z_j^t is merged with $Z_{j'}^t$. Then necessarily $i \notin I^t$ and $j, j' \notin J^t$. Since $\mathcal{A}_{t+1}$ is assumed to hold, it follows that $N_\epsilon(Y^{t+1}, Z^{t+1}) \leq N_\epsilon(Y^t, Z^t) + 1$ and $N_\epsilon(Z^{t+1}, Y^{t+1}) \leq N_\epsilon(Z^t, Y^t) - 1$. Thus, in this case, $N^{t+1} \leq N^t$.
2. The transition involves splitting in Z^t and merging in Y^t. By symmetry, also in this case we have $N^{t+1} \leq N^t$.
3. The transition involves splitting in Y^t and splitting in Z^t. Note that by construction the size of one of the newly created split entries is the same for Y as for Z. Suppose that Y_i^t and Z_j^t are split. If $i \in I^t$ then also $j \in J^t$ and $Y_i^t = Z_j^t$. In that case, both new entries for Y are the same as the new entries for Z, and hence $N^{t+1} = N^t$. The same conclusion is obtained if $j \in J^t$. If $i \notin I^t$ and $j \notin J^t$, then in both Y^t and Z^t an unmatched entry is replaced by two entries at least one of which is matched. Thus $N^{t+1} \leq N^t$.
4. The transition involves merging in Z^t and merging in Y^t. Suppose that Y_i^t is merged with $Y_{i'}^t$. It is easy to verify, as above, that in this case also $N^{t+1} \leq N^t$. However, if $i, i' \notin I^t$, then the corresponding statement is also true for the merged entries in Z^t, and we actually have $N^{t+1} \leq N^t - 2$.

In summary, we see that on the event $\mathcal{A}_{t+1}$ we have $N^{t+1} \leq N^t$ and $N^{t+1} \leq N^t - 2$ when there is merging in both Y^t and Z^t and the merging does not involve matched entries.

Since $N^t \geq 0$, we obviously have

$$\sum_{t=0}^{\infty} (N^t - N^{t+1}) 1_{\mathcal{A}_{t+1}} \leq N^0,$$

and we have seen that all the summands are nonnegative. Since q is independent

from $(N^t - N^{t+1})1_{\mathcal{A}_{t+1}}$,

$$\begin{aligned}\mathbf{E}[(N^q - N^{q+1})1_{\mathcal{A}_{q+1}}] &= \sum_t \mathbf{E}[(N^t - N^{t+1})1_{\mathcal{A}_{t+1}}1_{q=t}] \\ &= \sum_t \mathbf{E}[(N^t - N^{t+1})1_{\mathcal{A}_{t+1}}]\mathbf{P}[q=t] \le \eta N^0.\end{aligned} \tag{3.3}$$

Set $a^t = 1 - Q^t - \max\{y_1^t, z_1^t\}$. Recall the random variables u and v used in the transition kernel $\tilde{M}$. If in the transition from (Y^t, Z^t) to (Y^{t+1}, Z^{t+1}) we have $u < 1 - Q^t$ and $\max\{y_1^t, z_1^t\} < v < 1 - Q^t$, then in both Y and Z we have merging of unmatched entries. Thus,

$$\mathbf{P}[N^t - N^{t+1} \ge 2 \text{ or } \neg\mathcal{A}_{t+1}|\mathcal{F}_t] \ge (1 - Q^t)a^t.$$

By applying this at time $t = q$ and taking expectations, we get

$$\begin{aligned}\mathbf{E}[(1 - Q^q)a^q] &\le \mathbf{P}[N^q - N^{q+1} \ge 2 \text{ or } \neg\mathcal{A}_{q+1}] \\ &\le \frac{1}{2}\mathbf{E}[(N^q - N^{q+1})1_{\mathcal{A}_{q+1}}] + \mathbf{P}[\neg\mathcal{A}_{q+1}].\end{aligned}$$

Consequently, (3.2) and (3.3) complete the proof of the lemma. ∎

Assuming that we can make the right hand side of (3.1) small, Lemma 3.1 tells us that with high probability either $1 - Q^q - y_1^q$ or $1 - Q^q - z_1^q$ is small. If we knew that both are small, it would follow that also $y_1^q - z_1^q$ is rather small, since $\sum_i Y_i^q = \sum_j Z_j^q = 1$. However, it might be the case that $1 - Q^q - z_1^q$ is small but $1 - Q^q - y_1^q$ is not. The next lemma tells us that in such a situation, with high probability, Y^q does not have more than two significant unmatched entries.

LEMMA 3.2: *With the setting and notations of Lemma 3.1, let y_2^t be the second largest unmatched entry in Y^t. For every $\rho \in (0,1)$*

$$\mathbf{P}[1 - Q^q - y_1^q - y_2^q > \rho] < 2^6\rho^{-4}\eta N^0 + 2^9\bar{\epsilon}\rho^{-4}\mathbf{E}[q+2]. \tag{3.4}$$

Proof: Let $\mathcal{D}$ be the event $\{1 - Q^q - y_1^q - y_2^q > \rho\}$ and let $\mathcal{R}$ be the event $\{1 - Q^q - z_1^q < \rho/4\}$. Assume that $\mathcal{D} \cap \mathcal{R}$ holds. Then $z_1^q \ge 3\rho/4 + y_1^q + y_2^q$. Let $\mathcal{U}$ be the event that the random variables u and v used in the transition from (Y^q, Z^q) to (Y^{q+1}, Z^{q+1}) satisfy $u < 3\rho/4$ and $z_1^q - \rho/2 < v < z_1^q - \rho/4$. On $\mathcal{D} \cap \mathcal{R} \cap \mathcal{U}$, the largest unmatched entry in Z^q will be split and the transition from Y^q to Y^{q+1} would involve a merge (of unmatched entries), because $z_1^q - \rho/2 > y_1$. Consequently, a.s. on $\mathcal{D} \cap \mathcal{R} \cap \mathcal{U}$ the two new entries of Z^{q+1} will be unmatched in Y^{q+1}, and in particular, $1 - Q^{q+1} \ge z_1^q$. Moreover, each of the new entries

of Z^{q+1} would be larger than $\rho/4$. Clearly, $y_1^{q+1} \le y_1^q + y_2^q$. Consequently, $1 - Q^{q+1} - y_1^{q+1} \ge z_1^q - y_1^q - y_2^q \ge 3\rho/4$ and $1 - Q^{q+1} - z_1^{q+1} \ge \rho/4$. Thus, on $\mathcal{D} \cap \mathcal{R} \cap \mathcal{U}$, we have $1 - Q^{q+1} - \max\{y_1^{q+1}, z_1^{q+1}\} \ge \rho/4$. Now,

$$\begin{aligned}
&\mathbf{E}[(1 - Q^{q+1})(1 - Q^{q+1} - \max\{y_1^{q+1}, z_1^{q+1}\})] \\
&\quad \ge \mathbf{E}[(1 - Q^{q+1} - \max\{y_1^{q+1}, z_1^{q+1}\})^2 | \mathcal{D}, \mathcal{R}, \mathcal{U}] \mathbf{P}[\mathcal{D}, \mathcal{R}, \mathcal{U}] \\
&\quad \ge (\rho^2/16)\mathbf{P}[\mathcal{D}, \mathcal{R}, \mathcal{U}].
\end{aligned}$$

Lemma 3.1 with q replaced by $q+1$ therefore gives

$$\mathbf{P}[\mathcal{D}, \mathcal{R}, \mathcal{U}] \le 8\rho^{-2}\eta N^0 + 2^6 \bar{\epsilon}\rho^{-2}\mathbf{E}[q+2].$$

Clearly, $\mathbf{P}[\mathcal{U} | \mathcal{D}, \mathcal{R}] = 3\rho^2/16$, and hence

$$\mathbf{P}[\mathcal{D}, \mathcal{R}] \le (16/3)\rho^{-2}\mathbf{P}[\mathcal{D}, \mathcal{R}, \mathcal{U}].$$

On the other hand, on $\mathcal{D} \setminus \mathcal{R}$ we have $(1 - Q^q)(1 - Q^q - \max\{y_1^q, z_1^q\}) > \rho^2/4$. Thus, applying Lemma 3.1 again gives

$$\mathbf{P}[\mathcal{D} \setminus \mathcal{R}] \le 2\rho^{-2}\eta N^0 + 16\bar{\epsilon}\rho^{-2}\mathbf{E}[q+1].$$

Since $\mathbf{P}[\mathcal{D}] = \mathbf{P}[\mathcal{D}, \mathcal{R}] + \mathbf{P}[\mathcal{D} \setminus \mathcal{R}]$, the above estimates combine to give (3.4), and complete the proof. ∎

LEMMA 3.3: *With the setting and notations of Lemma 3.1, let $\rho \in (0, 1/8)$ and assume that $0 < \epsilon < \rho$. Then for each $t \in \mathbb{N}_+$ and for every $n \in \mathbb{N}_+$ satisfying $2^n \le t\rho$*

$$t^{-1} \sum_{\tau=0}^{t-1} \mathbf{P}[y_1^\tau \ge \rho] \le O(\rho^{-1}n^{-1}) + O(2^{4n}/\rho^5)(N^0/t + \bar{\epsilon}t). \tag{3.5}$$

The basic idea of the proof of the lemma is to use the fact that conditioned on $y_1^\tau \ge \rho$ there is a significant enough probability that at a later time σ there will be some unmatched $Y_{i'}^\sigma \in [2^{-k}\rho, 2^{-k+1}\rho]$, since the unmatched piece at time τ of size $\ge \rho$ may be split immediately. Lemma 3.2 is then used to show that when we fix σ, with high probability the latter event occurs for at most three different k in the range $\{1, \dots, n\}$, if n is not too large. An appropriate summation over k and σ completes the proof.

Proof: For $\sigma > \tau$, $\sigma, \tau \in \mathbb{N}$, $k \in \mathbb{N}_+$, let $\mathcal{X}(\tau, \sigma, k)$ be the event that the transition between time τ and $\tau + 1$ produces a splitting in Y^τ and one of the

split pieces is unmatched, has size in the range $[2^{-k-1}\rho, 2^{-k}\rho)$, and this split piece is not modified up to time σ. Suppose that $y_1^\tau \geq \rho$ and that $Y_i^\tau = y_1^\tau$. If in the transition from τ to $\tau+1$ we have $u \in \tilde{Y}_i^\tau$ and $v \in (Y_i^\tau - 2^{-k}\rho, Y_i^\tau - 2^{-k-1}\rho)$, then Y_i^τ is indeed split, and it is easy to see that the resulting piece $Y_i^\tau - v$ is unmatched a.s. If that happens, the conditioned probability that up to time σ this piece is modified is bounded by $2(\sigma-\tau)2^{-k}\rho$, since the size of this piece is at most $2^{-k}\rho$. Thus,

$$\mathbf{P}[\mathcal{X}(\tau,\sigma,k)|y_1^\tau \geq \rho] \geq 2^{-k-1}\rho^2(1-2(\sigma-\tau)2^{-k}\rho),$$

which implies

$$\mathbf{P}[y_1^\tau \geq \rho] \leq 2^{k+2}\rho^{-2}\mathbf{P}[\mathcal{X}(\tau,\sigma,k)], \quad \text{if } \tau < \sigma \leq \tau + 2^{k-2}/\rho. \tag{3.6}$$

Let $\mathcal{V}^\sigma(n)$ be the event that there are at least 3 distinct $k \in \{0,1,\ldots,n-1\}$ such that there is an unmatched Y_i^σ in the range $[2^{-k-1}\rho, 2^{-k}\rho)$. We now apply Lemma 3.2 with q chosen uniformly in $\{0,1,\ldots,2t-1\}$ and with ρ replaced by $2^{-n}\rho$ to get

$$\sum_{\sigma=0}^{2t-1} \mathbf{P}[\mathcal{V}^\sigma(n)] \leq 2^{4n+11}\rho^{-4}(N^0 + \bar{\epsilon}(t+2)^2). \tag{3.7}$$

Now, observe that

$$\sum_{\tau=0}^{\sigma-1}\sum_{k=0}^{n-1} 1_{\mathcal{X}(\tau,\sigma,k)} < 3 + 1_{\mathcal{V}^\sigma(n)}n.$$

Therefore, by taking expectations and applying (3.7) we get

$$\sum_{\sigma=0}^{2t-1}\sum_{\tau=0}^{\sigma-1}\sum_{k=0}^{n-1}\mathbf{P}[\mathcal{X}(\tau,\sigma,k)] \leq 6t + n\sum_{\sigma=0}^{2t-1}\mathbf{P}[\mathcal{V}^\sigma(n)] \leq O(t) + O(1)2^{4n}n\rho^{-4}(N^0+\bar{\epsilon}t^2).$$

We now assume that $2^n \leq t\rho$. Then the inequalities (3.6) may be applied to the above, giving

$$\sum_{k=0}^{n-1}\sum_{\tau=0}^{t-1}\sum_{\sigma=\tau+1}^{\tau+\lfloor 2^{k-2}/\rho\rfloor} 2^{-k-2}\rho^2\mathbf{P}[y_1^\tau \geq \rho] \leq O(t) + O(1)2^{4n}n\rho^{-4}(N^0+\bar{\epsilon}t^2).$$

This implies (3.5), and completes the proof. ∎

COROLLARY 3.4: *Let $\gamma \in (0,1/2)$. Let q be a random variable with values in $\mathbb{N}$ which is independent from the Markov chain (Y^t, Z^t). Set $\eta := \max\{\mathbf{P}[q = t] : t \in \mathbb{N}\}$, and suppose that $(\bar{\epsilon})^{1-\gamma} \leq \eta \leq (\bar{\epsilon})^{\gamma}/\max\{N^0, 1\}$. Then for all $\lambda \geq 1$ and $\rho > 0$*

$$\mathbf{P}[y_1^q \geq \rho] \leq \mathbf{P}[q > \lambda\eta^{-1}] + C(\lambda/\rho)|\log\bar{\epsilon}|^{-1},$$

where C is a constant depending only on γ.

Proof: Let $s := \lfloor \lambda\eta^{-1} \rfloor$. We have

$$\begin{aligned}\mathbf{P}[y_1^q \geq \rho] &\leq \mathbf{P}[q > \lambda\eta^{-1}] + \sum_{\tau=0}^{s} \mathbf{P}[q = t]\mathbf{P}[y_1^\tau \geq \rho] \\ &\leq \mathbf{P}[q > \lambda\eta^{-1}] + \eta \sum_{\tau=0}^{s} \mathbf{P}[y_1^\tau \geq \rho].\end{aligned}$$

Thus, the proof is completed by applying (3.5) with $t = s + 1$ and $n := \lfloor |\log\bar{\epsilon}|/C \rfloor$, provided that with sufficiently large C we have

$$2^{4n}\rho^{-5}(N^0/t + \bar{\epsilon}t) \leq C\rho^{-1}|\log\bar{\epsilon}|^{-1} \tag{3.8}$$

and $2^n \leq t\rho$. First, note that we may assume that $\lambda, \rho^{-1} < |\log\bar{\epsilon}|$ and $\bar{\epsilon} < 1/10$. Then $2^n \leq t\rho$ holds by the assumptions on η. It is also easy to verify that with an appropriate choice of C, (3.8) follows from our inequalities for η and assumptions about ρ and λ. ∎

THEOREM 3.5: *Let Y^0 and Z^0 be independent random samples from $PD(1)$, and let (Y^t, Z^t) denote their evolution under $\tilde{M}$. Let $t_0 \in \mathbb{N}_+$, and let $q \in \{0, 1, \ldots, t_0 - 1\}$ be chosen uniformly and independently from the evolution of the chain (Y^t, Z^t). Then for each $\rho > 0$,*

$$\mathbf{P}[\max\{y_1^q, z_1^q\} > \rho] \leq O(1)\rho^{-1}(\log t_0)^{-1}.$$

Proof: Set $\epsilon := (t_0)^{-2}$, and define $\bar{\epsilon}$ as in Lemma 3.1. Recall that a size biased sample from the $PD(1)$ sample Y^0 gives the uniform distribution on $[0,1]$ (this is well-known, but also easy to verify from the definition). Consequently, $\mathbf{E}[\bar{\epsilon}] = 3\epsilon$. Let $\mathcal{A}_1$ be the event that $\bar{\epsilon} \leq \epsilon^{3/4}$. Then $\mathbf{P}[\neg\mathcal{A}_1] \leq 3\epsilon^{1/4}$. Let $\mathcal{A}_2$ be the event that $N^0 \leq \epsilon^{-1/4}$. It is easy to see (e.g., using the description of $PD(1)$ from the introduction) that $\mathbf{P}[\neg\mathcal{A}_2] \leq O(\epsilon)$. (In fact, $N^0/|\log\epsilon|$ is very unlikely to be large.) Define η as in Corollary 3.4. Then $\eta = \epsilon^{1/2}$ and on $\mathcal{A}_1 \cap \mathcal{A}_2$ we have $(\bar{\epsilon})^{1-\gamma} \leq \eta \leq (\bar{\epsilon})^{\gamma}/\max\{N^0, 1\}$ with $\gamma = 1/5$, for example. On the event

$\mathcal{A}_1 \cap \mathcal{A}_2$, apply the corollary with $\lambda = 1$ and the corresponding statement with the roles of Y and Z switched, to get

$$\mathbf{P}[\max\{y_1^q, z_1^q\} > \rho | \mathcal{A}_1 \cap \mathcal{A}_2] \leq O(\rho^{-1})|\log\epsilon|^{-1}.$$

Now our estimates for $\mathbf{P}[\neg\mathcal{A}_1]$ and $\mathbf{P}[\neg\mathcal{A}_2]$ complete the proof. ∎

Proof of Theorem 1.2: The proof is similar to the proof of Theorem 3.5 Let μ be a measure that is invariant under M, and let Y^0 be a sample from μ. Let Z^0 be a sample from $PD(1)$ (which we may take to be independent from Y^0, though this is not important). Let $t_0 \in \mathbb{N}_+$, $t_0 > 5$, and let q be as in Theorem 3.5. As in the proof of that theorem, choose $\epsilon = t_0^{-2}$.

Note that for every $t \in \mathbb{N}$, Y^t is also a sample from μ, because μ is t invariant. The same also holds for Y^q, since q is independent from the chain (Y^t).

We now explain how to get bounds on the distributions of $\bar{\epsilon}$ and N^0 using continuous analogs of Lemmas 2.3 and 2.4. Let $\beta(s, Y) := \sum\{Y_i : Y_i \leq s\}$. Since $\lim_{s\searrow 0}\beta(s, Y^0) = 0$ a.s., we may choose $k = k(\epsilon) > 0$ sufficiently large so that $\mathbf{P}[\beta(2^{-k}, Y^0) > \epsilon] < \epsilon$ and $2^{-k} < \epsilon(1-\epsilon)/8$. Set $\delta = 1 - 2^{-k}$ and $t_1 = \lceil 2^6\delta^{-1}k2^k\rceil$. The proof of Lemma 2.3 applied to the continuous setting gives

$$\mathbf{E}[\beta(\epsilon, Y^{t_1})|\beta(2^{-k}, Y^0) \leq \epsilon] \leq O(1)\epsilon|\log\epsilon|.$$

By our choice of k this implies $\mathbf{E}[\beta(\epsilon, Y^{t_1})] \leq O(\epsilon)|\log\epsilon|$. Since Y^{t_1} and Y^0 have the same distribution, this gives

$$\mathbf{E}[\beta(\epsilon, Y^0)] \leq O(\epsilon)|\log\epsilon|, \tag{3.9}$$

and since $\mathbf{E}[\beta(\epsilon, Z^0)] = \epsilon$, as in the proof of Theorem 3.5, we conclude that $\mathbf{E}[\bar{\epsilon}] \leq O(\epsilon)|\log\epsilon|$.

We now adapt the latter part of the proof of Lemma 2.4. Let $m_0 := \lceil|\log_2\epsilon|\rceil$. On the one hand

$$\sum_{m=0}^{m_0} 2^m\beta(2^{-m}, Y^0) = \sum_{m=0}^{m_0}\sum\{2^mY_i : Y_i \leq 2^{-m}\} \leq 2|\{i \in \mathbb{N}_+ : Y_i^0 \geq \epsilon\}|.$$

On the other hand, (3.9) gives

$$\mathbf{E}\left[\sum_{m=0}^{m_0} 2^m\beta(2^{-m}, Y^0)\right] \leq O(1)\sum_{m=0}^{m_0} m = O(1)|\log\epsilon|^2.$$

Consequently, we have $\mathbf{E}[N^0] = O(1)|\log\epsilon|^2$. Now, the proof of Theorem 3.5 applies, and gives for all $\rho > 0$

$$\mathbf{P}[\max\{y_1^q, z_1^q\} > \rho] \leq O(1)\rho^{-1}(\log t_0)^{-1}.$$

We conclude that for every $\rho > 0$ there is a coupling of Y^0 and Z^0 so that $\mathbf{P}[\max\{y_1^0, z_1^0\} > \rho] < \rho$. This implies that $\mu = PD(1)$. ∎

4. Conclusion

Proof of Theorem 1.1: The proof is similar to the proof of Theorem 3.5. Let $\epsilon > 0$. Let q be uniformly chosen in $(2\mathbb{Z}) \cap [0, \epsilon^{-1/2}]$. Set $z = z(2t/n)$, as in (1.1). Let Z^0 be chosen according to $PD(1)$, and let $Y^\tau = \mathfrak{X}(\pi_{t+\tau})/(nz)$.

We now apply a coupling of Z^τ and Y^τ similar to the coupling $\tilde{M}$ given in Section 3. There are a few minor necessary modifications in the definition of the coupling. First, note that the entries of Y^τ do not sum to 1 but to $1/z$. Thus, the random variables u and v needed in the transition kernel for $\tilde{M}$ should be uniform in $[0, 1/z]$. In $\tilde{Y}^\tau$ and $\hat{Y}^\tau$ we put those segments corresponding to cycles that do not intersect V_G^t in the very end; that is, roughly in the interval $[1, 1/z]$. When u or v turn out to be outside of $[0, 1]$, we make no transition to Z; that is, $Z^{\tau+1} = Z^\tau$, in this case.

Another modification is necessary because the transitions of Y^τ are discrete. Thus, the actual size of the splits occurring in the transitions of Y would be determined with $\lceil nzv \rceil/(nz)$. The definition of the matching between entries in Y^τ and entries in Z^τ needs to be modified as well. When a split is made in both Y^τ and Z^τ, pieces which would have been exactly the same may differ slightly now, because of the discretization in the transition of Y. This difference is of order $1/n$, and may be safely ignored. Although these errors may accumulate over time, when matched pieces are merged and split, the total discrepancy would still be small, since we take n much larger than $\epsilon^{-1/2}$, which bounds q.

Lemma 2.4 gives us good control on N^0 while (2.4) gives a bound on the probability that $\bar{\epsilon}$ is large. Consequently, the proof of Theorem 3.5 shows that for all $\rho > 0$ if n is large

$$\mathbf{P}[\|Y^q - Z^q\|_\infty > \rho] \leq O(1)\rho^{-1}|\log \epsilon|^{-1}. \tag{4.1}$$

Thus, the statement of Theorem 1.1 is obtained with t replaced by $t + q$. If we consider ϵ and ρ as fixed, then q is bounded. Since q is even, the following lemma completes the proof. ∎

LEMMA 4.1: *Let $t \geq cn$, $c > 1/2$. As $n \to \infty$, the total variation distance between the law of $\mathfrak{X}(\pi_t)$ and the law of $\mathfrak{X}(\pi_{t+2})$ tends to zero.*

First, we give a slightly informal proof. Note that when the largest entry in $\mathfrak{X}(\pi_\tau)$ is not too small, there is probability bounded away from 0 and 1 that

$\mathfrak{X}(\pi_\tau) = \mathfrak{X}(\pi_{\tau+2})$, because that entry may split and then recombine. We know that for many $\tau \in [n/2, t]$ the largest entry is not small. Consequently, there is a random "delay", which implies the statement of the lemma. For readers who are not convinced yet, we offer a proof with more details.

Proof: Set $W^\tau = \mathfrak{X}(\pi_\tau)$. To prove that W^t and W^{t+2} have close distributions, we couple the chain (W^τ) with a chain (U^τ) which has the same distribution as (W^τ). In essence, the two chains will be the same; the significant difference involves a random shift in time. Set $\tau_0 = \tau_0' = 0$. Inductively, suppose that τ_i and τ_i' have been defined such that $W^{\tau_i} = U^{\tau_i'}$. Let $m = m_i$ be the largest integer such that $W^{\tau_i+2j} = W^{\tau_i}$ for all $j = 1, 2, \ldots, m$. The distribution of m conditioned on W^{τ_i} is geometric; that is, $\mathbf{P}[m = k|W^{\tau_i}] = (1-p)p^k$, where $p = p_i = \mathbf{P}[m > 0|W^{\tau_i}]$. Similarly, the largest integer m' such that $U^{\tau_i'+2j} = U^{\tau_i'}$ for all $j = 1, 2, \ldots, m'$ has the same conditioned distribution: $\mathbf{P}[m' = k|W^{\tau_i}] = \mathbf{P}[m' = k|U^{\tau_i'}] = (1-p)p^k$. We now couple m and m'. If $\tau_i' = \tau_i + 2$, take $m' = m$. Otherwise we couple m and m' so that $|m - m'| \leq 1$, but $m \neq m'$ happens quite frequently. For example, for all $k \in \mathbb{N}$ take $(m, m') = (k, k+1)$ with probability $p^{k+1}(1-p)/(1+p)$, $(m, m') = (k+1, k)$ with the same probability, and $(m, m') = (0, 0)$ with probability $(1-p)/(1+p)$, all conditioned on W^{τ_i}. In this case, the conditioned probability that $m' - m = \pm 1$ is $2p/(1+p)$. In either case, take $U^{\tau_i'+j} = W^{\tau_i+j}$ for $j = 1, 2, \ldots, \min\{2m, 2m'\}$. If $m' > m$ let $U^{\tau_i'+2m+1}$ be independent from the chain (W^τ) given (W^{τ_i}, m, m'), and similarly if $m > m'$. Clearly, $W^{\tau_i+2m} = U^{\tau_i'+2m'}$. Take $U^{\tau_i'+2m'+j} = W^{\tau_i+2m+j}$ for $j = 1, 2$, $\tau_{i+1} := \tau_i + 2m + 2$ and $\tau_{i+1}' := \tau_i' + 2m' + 2$. Then continue inductively. This completes the specification of the coupling.

It clearly suffices to prove that with probability tending to 1 as $n \to \infty$, we have $\tau_i' = \tau_i + 2$ with some $\tau_i < t$. First, observe that the p_i are bounded away from 1. This guarantees that a.a.s. $\tau_i = O(i)$. Now note that p_i is bounded away from zero by some positive function of $W_1^{\tau_i}/n$ (the largest entry normalized), because that largest entry may split in the next step, and then the same two parts may merge in the step after that. We know, for example from (2.4), that with high probability for most values of i such that $t \geq \tau_i \geq (cn + n/2)/2$, the largest entry of W^{τ_i} is not too much smaller than n. Consequently, a.a.s. we have p_i bounded away from zero for many values of i satisfying $\tau_i < t$. Similarly, $m_i \neq m_i'$ for many values of i. Note that $(\tau_i - \tau_i')/2$ is a martingale, and its increments are $\{-1, 0, 1\}$. By removing the 0 increment steps, the martingale may be coupled with a simple random walk on $\mathbb{Z}$. The martingale starts at 0. Thus, the probability that many ± 1 steps are performed and it never gets to 1

tends to 0. This completes the proof. ∎

ACKNOWLEDGEMENT: We have had the pleasure to benefit from conversations with Rick Durrett, Michael Larsen, Russ Lyons, David Wilson and Ofer Zeitouni in connection with this work.

References

[Ang03] O. Angel, *Random infinite permutations and the cyclic time random walk,* in *Random Walks and Discrete Potential Theory* (C. Banderier and C. Krattenthaler, eds.), Discrete Mathematics and Theoretical Computer Science, 2003, pp. 9–16, http://dmtcs.loria.fr/proceedings/html/dmAC0101.abs.html.

[AS00] N. Alon and J. H. Spencer, *The Probabilistic Method,* second edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. With an appendix on the life and work of Paul Erdős.

[BD] N. Berestycki and R. Durrett, *A phase transition in the random transposition random walk,* arXiv:math.PR/0403259.

[DMP95] P. Diaconis, M. McGrath and J. Pitman, *Riffle shuffles, cycles, and descents,* Combinatorica **15** (1995), 11–29.

[DMWZZ] P. Diaconis, E. Mayer-Wolf, O. Zeitouni and M. P. Zerner, *The Poisson–Dirichlet law is the unique invariant distribution for uniform split-merge transformations,* The Annals of Probability **32** (2004), 915–938.

[DS81] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions,* Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **57** (1981), 159–179.

[Dud89] R. M. Dudley, *Real Analysis and Probability,* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[Hol01] L. Holst, *The Poisson–Dirichlet distribution and its relatives revisited,* 2001, http://www.math.kth.se/matstat/fofu/reports/PoiDir.pdf, preprint.

[JLR00] S. Janson, T. Łuczak and A. Rucinski, *Random Graphs,* Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[Spe94] J. Spencer, *Ten lectures on the probabilistic method,* Volume 64 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, second edition, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[Tót93] B. Tóth, *Improved lower bound on the thermodynamic pressure of the spin 1/2 Heisenberg ferromagnet*, Letters in Mathematical Physics **28** (1993), 75–84.

[Wat76] G. A. Watterson, *The stationary distribution of the infinitely-many neutral alleles diffusion model*, Journal of Applied Probability **13** (1976), 639–651.