Electronic Journal of Probability

Vol. 16 (2011), Paper no. 5, pages 152–173.

Journal URL
http://www.math.washington.edu/~ejpecp/

Emergence of giant cycles and slowdown transition in random transpositions and k-cycles

Nathanaël Berestycki*

Abstract

Consider the random walk on the permutation group obtained when the step distribution is uniform on a given conjugacy class. It is shown that there is a critical time at which two phase transitions occur simultaneously. On the one hand, the random walk slows down abruptly: the acceleration (i.e., the second time derivative of the distance) drops from 0 to $-\infty$ at this time as $n \to \infty$. On the other hand, the largest cycle size changes from microscopic to giant. The proof of this last result is considerably simpler and holds more generally than in a previous result of Oded Schramm [19] for random transpositions. It turns out that in the case of random k-cycles, this critical time is proportional to $1/[k(k-1)]$, whereas the mixing time is known to be proportional to $1/k$.

Key words: random transpositions, random k-cycles, random permutations, cycle percolation, coalescence-fragmentation, random hypergraphs, conjugacy class, mixing time.

AMS 2000 Subject Classification: Primary 60J10, 60K35, 60B15; Secondary: 05C80, 05C12, 05C65.

Submitted to EJP on April 21, 2010, final version accepted December 26, 2010.

*Statistical Laboratory, Cambridge University. Wilberforce Rd., Cambridge CB3 0WB.

1 Introduction

1.1 Basic result

Let $n \geq 1$ and let $\mathcal{S}_n$ be the group of permutations of $\{1, \ldots, n\}$. Consider the random walk on $\mathcal{S}_n$ obtained by performing random transpositions in continuous time, at rate 1. That is, let $\tau_1, \ldots$ be a sequence of i.i.d. uniformly chosen transpositions among the $n(n-1)/2$ possible transpositions of the set $V = \{1, \ldots, n\}$, and for all $t \geq 0$, set

$$\sigma_t = \tau_1 \circ \ldots \circ \tau_{N_t}$$

where $(N_t, t \geq 0)$ is an independent Poisson process with rate 1 and $\circ$ stands for the usual composition of permutations. It is well-known that the permutation σ_t is approximately a uniform random permutation (in the sense of total variation distance) after time $(1/2)n \log n$ (see [9]). In particular, this means that at this time, most points belong to cycles which are of macroscopic size $O(n)$, while initially, in the permutation σ_0 which is the identity permutation, every cycle is microscopic (being of size 1). How long does it take for macroscopic cycles to emerge? Oded Schramm, in a remarkable paper [19], proved that the first giant cycles appear at time $n/2$.

More precisely, answering a conjecture of David Aldous stated in [3], he was able to prove the following: if $t = cn$ with $c > 1/2$, then there exists a number $\theta = \theta(c) \in (0, 1)$ and a (random) set $W \subset \{1, \ldots, n\}$ satisfying $\sigma_t(W) = W$, such that $|W| \sim \theta n$. Furthermore, the cycle lengths of $\sigma_t|_W$, rescaled by θn, converge in the sense of finite-dimensional distributions towards a Poisson-Dirichlet random variable. (The Poisson-Dirichlet distribution describes the limiting cycle distribution of a uniform random permutation, see [19]. The number $\theta(c)$ in this result is the survival probability of a Galton-Watson branching process with Poisson offspring with mean $2c > 1$.) In particular, this implies that σ_t contains giant cycles with high probability if $t = cn$ with $c > 1/2$. On the other hand it is easy to see that no macroscopic cycle can occur if $c < 1/2$.

His proof is separated into two main steps. The first step consists in showing that giant cycles do emerge prior to time cn when $c > 1/2$. The second step is a beautiful coupling argument which shows that once giant cycles exist they must quickly come close to equilibrium, thereby proving Aldous' conjecture. Of these two steps, the first is the more involved. Our main purpose in this paper is to give an elementary and transparent new proof of this fact. Let $\Lambda(t)$ denote the size of the largest cycle of σ_t. For $\delta > 0$, define

$$\tau_\delta = \inf\{t \geq 0 : \Lambda(t) > \delta n\}. \tag{1}$$

Theorem 1. *If $c > 1/2$, then $\tau_\delta < cn$ with probability tending to 1, where*

$$\delta = \frac{\theta(c)^2}{8} > 0,$$

where $\theta(c)$ is the survival probability of a Galton-Watson branching process with Poisson offspring with mean $2c > 1$.

The result also trivially holds if $c \leq 1/2$ since $\theta(c) = 0$ then. The number $\delta = \theta(c)^2/8$ is somewhat arbitrary. It is likely that the result holds with any $\delta < \theta(c)$, although our proof does not show this. Our proof is completely elementary and in particular requires few estimates. As a consequence, it

is fairly robust and it is reasonable to hope that it extends to further models. We demonstrate that this is indeed the case by applying it to more general random walks on $\mathcal{S}_n$, whose step distribution is uniform on a given conjugacy class of the permutation group (definitions will be recalled below). We also show that the emergence of giant cycles coincides with a phase transition in the *acceleration* of the random walk, as measured by the second derivative of the distance (on the graph metric) between the position of the random walk at time t, and its starting point. This phase transition in the acceleration is the exact analogue of the phase transition described in [3] for random transpositions.

We mention that Theorem 1 is the mean-field analogue of a question arising in statistical mechanics in the study of Bose-Einstein condensation and the quantum ferromagnetic Heisenberg model (see Tòth [20]). Very few rigorous results are known about this model on graphs with non-trivial geometry, with the exception of the work of Angel [1] for the case of a d-regular tree with d sufficiently large. We believe that the proof of Theorem 1 proposed here opens up the challenging possibility to prove analogous results on graphs that are "sufficiently high-dimensional" such as a high-dimensional hypercube, for which recent progress has been made about the emergence of a giant component under the percolation process: see, e.g., Borgs et al. [6].

1.2 Random walks based on conjugacy classes.

Fix a number $k \geq 2$, and call an element $\gamma \in \mathcal{S}_n$ a k-cycle, or a cyclic permutation of length k, if there exist pairwise distinct elements $x_1, \ldots, x_k \in \{1, \ldots, n\}$ such that $\gamma(x) = x_{i+1}$ if $x = x_i$ (where $1 \leq i \leq k$ and $x_{k+1} := x_1$) and $\gamma(x) = x$ otherwise. Thus for $k = 2$, a 2-cycle is simply a transposition. If σ is a permutation then σ can be decomposed into a product of disjoint cyclic permutations $\sigma = \gamma_1 \circ \ldots \circ \gamma_r$ where $\circ$ stands for the composition of permutations. (This decomposition being unique up to the order of the terms). A conjugacy class $\Gamma \subset \mathcal{S}_n$ is any set that is invariant by conjugacy $\sigma \mapsto \pi^{-1}\sigma\pi$, for all $\pi \in \mathcal{S}_n$. It easily seen that a conjugacy class of $\mathcal{S}_n$ is exactly a set of permutations having a given cycle structure, say $(c_2, \ldots, c_J)$, i.e., consisting of c_2 cycles of size 2, ..., c_J cycles of size J in their cycle decomposition (and a number of fixed points which does not need to be explicitly stated). Note that if Γ is a fixed conjugacy class of $\mathcal{S}_n$, and $m > n$, Γ can also be considered a conjugacy class of $\mathcal{S}_m$ by also considering the set of permutatinons in S_m with c_2 cycles of size 2, ..., c_J cycles of size J.

Let Γ be a fixed conjugacy class, and consider the random walk in continuous time on $\mathcal{S}_n$ where the step distribution is uniform on Γ. That is, let $(\gamma_i, i \geq 1)$ be an i.i.d. sequence of elements uniformly distributed on Γ, and let $(N_t, t \geq 0)$ be an independent rate 1 Poisson process. Define a random process:

$$\sigma_t := \gamma_1 \circ \ldots \circ \gamma_{N_t}, \quad t \geq 0, \tag{2}$$

where $\circ$ stands for the composition of two permutations. Thus the case where Γ consists only of transpositions (i.e. $c_2 = 1$ and $c_j = 0$ if $j \geq 2$) corresponds to the familiar random process on $\mathcal{S}_n$ obtained by performing random transpositions in continuous time, and the case where Γ contains only one nontrivial cycle of size $k \geq 2$ will be referred to as the random k-cycles random walk. The process $(\sigma_t, t \geq 0)$ may conveniently be viewed as a random walk on G_n, the Cayley graph of $\mathcal{S}_n$ generated by Γ. Note that if $K := \sum_{j=2}^{J}(j-1)c_j$ is odd, the graph G_n is connected but it is not when K is even: indeed, in that case, the product of random p-cycles must be an even permutation, and thus σ_t is then a random walk on the alternate group $\mathcal{A}_n$ of even permutations. This fact will be of no relevance in what follows.

In this paper we study the pre-equilibrium behaviour of such a random walk. Our main result in this paper for this process is that there is a phase transition which occurs at time $t_c n$, where

$$t_c = \left(\sum_{j=2}^{J} j(j-1)c_j \right)^{-1}. \tag{3}$$

This transition concerns two distinct features of the walk. On the one hand, giant cycles emerge at time $t_c n$ precisely, as in Theorem 1. On the other hand, the speed of the walk changes dramatically at this time, dropping below 1 in a non-differentiable way. We start with the emergence of giant cycles, which is analogue to Theorem 1. Recall the definition of τ_δ in (1).

Theorem 2. *Let $t < t_c$. Then there exists $\beta > 0$ such that $\mathbb{P}(\sup_{s \le tn} \Lambda(s) \le \beta \log n) \to 1$, where $\Lambda(s)$ is the maximal cycle size at time $s \ge 0$. On the other hand for any $t > t_c$ there exists $\delta > 0$ such that $\tau_\delta < tn$ with high probability.*

We now state our result for the speed. Denote by $d(x, y)$ the graph distance between two vertices $x, y \in \mathcal{S}_n$, and for $t \ge 0$, let

$$d(t) = d(o, \sigma_t).$$

where o is the identity permutation of $\mathcal{S}_n$. Recall that a sequence of random functions $X_n(t)$ converge uniformly on compact sets of $S \subset \mathbb{R}$ in probability (*u.c.p.* for short) towards a random function $X(t)$ if for all compact subsets $K \subset S$, $\mathbb{P}(\sup_{t \in K} |X_n(t) - X(t)| > \varepsilon) \to 0$ as $n \to \infty$ for all $\varepsilon > 0$.

Theorem 3. *Let t_c be as in* (3)*, and fix $t > 0$. Then there exists an interval $I \subset (t_c, \infty)$, bounded away from t_c and ∞, and a nonrandom continuous function $\varphi(t)$ satisfying $\varphi(t) = t$ for $t \le t_c$ and $\varphi(t) < t$ for $t > t_c$, such that*

$$\frac{1}{n} d(tn) \longrightarrow \varphi(t), \quad t \in \mathbb{R} \setminus I \tag{4}$$

uniformly on compact sets in probability as $n \to \infty$. Furthermore φ is $\mathscr{C}^\infty$ everywhere except at $t = t_c$, where the acceleration satisfies $\varphi''(t_c^+) = -\infty$. In the case of random k-cycles ($k \ge 2$), $I = \emptyset$ so the convergence holds uniformly on compact sets in $\mathbb{R}$.

Remark 4. We believe that $I = \emptyset$ in all cases, but our proof only guarantees this in the case of random k-cycles and a few other cases which we have not tried to describe precisely. Roughly speaking there is a combinatorial problem which arises when we try to estimate the distance to the identity in the case of conjugacy classes which contain several non-trivial cycles of distinct sizes (particularly when these are coprime). This is explained in more details in the course of the proof. The above result is enough to prove that there is a phase transition for $d(tn)$ when $t = t_c$, but does not prevent other phase transitions after that time.

In the case of random k-cycles, we have $t_c = 1/[k(k-1)]$ and the function φ has the following explicit expression:

$$\varphi(t) := 1 - \sum_{s=0}^{\infty} \frac{((k-1)s+1)^{s-2}}{s!} (kt)^s e^{-kt(s(k-1)+1)} \tag{5}$$

It is a remarkable fact that for $t \le t_c$ a cancellation takes place and $\varphi(t) = t$. One can also check with Stirling's formula that φ is differentiable at t_c. The case $k = 2$ of random transpositions

matches Theorem 4 from [3]. In the general conjugacy class case, φ may be described implicitly as follows. For $t \geq 0$ and $z \in [0,1]$, let $G_t(z) = \exp(t \sum_{j=2}^{J} j c_j (z^{j-1} - 1))$, and let $\rho(t)$ be the smallest solution of the equation (in $z \in [0,1]$): $G_t(z) = z$. Then φ is defined by

$$\varphi(t) = \frac{1}{K} \int_0^t \sum_{j=2}^{J} c_j (j\rho(s) - \rho(s)^j) ds, \tag{6}$$

where $K := \sum_{j=2}^{J} c_j (j-1)$. It is a fact that $\rho(t) < 1$ if and only if $t > t_c$, hence $\varphi(t) = t$ for $t \leq t_c$ and $\varphi(t) < t$ for $t > t_c$. As we will see later, $\rho(t)$ is the extinction probability of the Gatson-Watson process with generating function $G_t(z)$.

1.3 Heuristics

The k-cycle random walk is a simple generalization of the random transpositions random walk on $\mathcal{S}_n$, for which the phase transition in Theorem 3 was proved in [3]. Observe that any k-cycle $(x_1, \ldots, x_k)$ may always be written as the product of $k-1$ transpositions:

$$(x_1, \ldots, x_k) = (x_1, x_2) \ldots (x_{k-1}, x_k)$$

This suggests that, qualitatively speaking, the k-cycle random walk should behave as "random transpositions speed up by a factor of $(k-1)$", and thus one might expect that phase transitions occur at a time that is inversely proportional to k. This is for instance what happens with the mixing time

$$t_{\text{mix}} = \frac{1}{k} n \log n \tag{7}$$

for the total variation distance. This was recently proved in [4] and was previously known [17] for $k \leq 6$, the particular case $k = 2$ being the celebrated Diaconis-Shahshahani theorem [9]. See [14] and [8] for an excellent introduction to the general theory of mixing times, and [18] in particular for mixing times of random walks on groups. The comparison to mixing times is not fortuitous: it is explained in [4] how the value of the mixing time is connected to Schramm's result [19] about Poisson-Dirichlet structure within the giant component. It may therefore come as a surprise that $t_c = 1/[k(k-1)]$ rather than $t_c = 1/k$ (although note that, of course, mixing occurs on a different time scale as specified by (7)). As it emerges from the proof, the reason for this fact is as follows. We introduce a coupling of $(\sigma_t, t \geq 0)$ with a random hypergraph process $(H_t, t \geq 0)$ on $V = \{1, \ldots, n\}$, which is the analogue of the coupling between random transpositions and Erdős-Renyi random graphs introduced in [3]. As we will see in more details, hypergraphs are graphs where edges (or rather *hyperedges*) may connect several vertices at the same time. In this coupling, every time a cycle $(x_1, \ldots, x_k)$ is performed in the random walk, H_t gains a hyperedge connecting $x_1, \ldots, x_k$. This is essentially the same as adding the complete graph K_k on $\{x_1, \ldots, x_k\}$ in the graph H_t. Thus the degree of a typical vertex grows at a speed which is $k(k-1)/2$ faster than in the standard Erdős-Renyi random graph. This results in a giant component occurring $k(k-1)/2$ faster as well. This explains the formula $t_c^{-1} = k(k-1)$, and an easy generalisation leads to (3).

Organisation of the paper: The rest of the paper is organised as follows. We first give the proof of Theorem 1. In the following section we introduce the coupling between $(\sigma_t, t \geq 0)$ and the random hypergraph process $(H_t, t \geq 0)$. In cases where the conjugacy class is particularly simple

(e.g. random k-cycles), a combinatorial treatment analogous to the classical analysis of the Erdős-Renyi random graph is possible, leading to exact formulae. In cases where the conjugacy class is arbitrary, our method is more probabilistic in nature (relying primarily on martingale methods) and the formulae take a different form (H_t is then closer to the Molly and Reed model of random graphs with prescribed degree distribution, [15] and [16]). The proof is thus slightly different in these two cases (respectively dealt with in Section 3 and 4), even though conceptually there are no major differences between the two cases.

2 Emergence of giant cycles in random transpositions

In this section we give a full proof of Theorem 1. As the reader will observe, the proof is really elementary and is based on well-known (and easy) results on random graphs. Consider the random graph process $(G_t, t \geq 0)$ on $V = \{1, \ldots, n\}$ obtained by putting an edge between i and j if the transposition (i, j) has occurred prior to time t. Then every edge is independent and has probability $p_t = 1 - e^{-t/\binom{n}{2}}$, so G_t is a realisation of the Erdős-Renyi random graph $G(n, p_t)$.

For $t \geq 0$ and $i \in V$, let C_i denote the cycle that contains i. Recall that if $C_i = C_j$ then a transposition (i, j) yields a fragmentation of $C_i = C_j$ into two cycles, while if $C_i \neq C_j$ then the transposition (i, j) yields a coagulation of C_i and C_j. It follows from this observation that every cycle of σ_t is a subset of one of the connected components of G_t. Thus let $N(t)$ be the number of cycles of σ_t and let $\bar{N}(t)$ denote the number of components of G_t. Then we obtain

$$N(t) \geq \bar{N}(t), \quad t \geq 0. \tag{8}$$

Now it is a classical and easy fact that the number $\bar{N}(t)$ has a phase transition at time $n/2$ (corresponding to the emergence of a giant component at this time). More precisely, let $\theta(c)$ be the asymptotic fraction of vertices in the giant component at time cn, so $\theta(c)$ is the survival probability of a Poisson Galton-Watson process with mean offspring $2c$ (in particular $\theta(c) = 0$ if $c < 1/2$).

Let $c > 1/2$ and fix an interval of time $[t_1, t_2]$ such that $t_2 = cn$ and $t_1 = t_2 - n^{3/4}$. Our goal will be to prove that a cycle of size at least δn appears during the interval $I = [t_1, t_2]$, where $\delta = \theta(c)^2/8$.

Lemma 5. *For any $c > 1/2$, with t_1 and t_2 as above,*

$$\bar{N}(t_1) - \bar{N}(t_2) \sim (t_2 - t_1)[1 - \theta^2(c)]$$

in the sense that the ratio of these two quantities tends to 1 in probability as $n \to \infty$.

Proof. This lemma follows easily from the following observation. The total number of edges that are added during I is a Poisson random variable with mean $t_2 - t_1$. Now, each time an edge is added to G_t, this changes the number of components by -1 if and only if the two endpoints are in distinct components (otherwise the change is 0). Thus if a transposition is applied at time $s \in [t_1, t_2]$, then given $\mathscr{F}_s$ (where $\mathscr{F}_s$ denotes the σ-field generated by the entire process up to time s), the conditional probability that it will change the number of components is $1 - \sum_i x_i^2(s)$, where $x_i(s)$ denotes the rescaled component sizes at time s, ordered in decreasing sizes. We claim that this converges in probability to $1 - \theta(c)^2$ uniformly over $s \in [t_1, t_2]$. Indeed observe first that this quantity is monotone in s, so it suffices to establish the claim for $s = t_1$ and $s = t_2$. We only treat $s = t_2$ for simplicity. It is clear that $\sum_i x_i^2 \geq x_1^2 \to \theta(c)^2$. Moreover, $\mathbb{E}(\sum_i x_i^2) \to \theta(c)^2$ as the

second largest component has rescaled size smaller than $\beta \log n/n$ with high probability (see, e.g., Theorem 2.3.2 in [10]). Hence $\sum_{i\geq 2} x_i^2 \to 0$ in expectation and hence in probability, so the claim is proved. The law of large numbers now concludes the proof: fix $\varepsilon > 0$ and consider the event E that $\sum_i x_i^2 \in [\theta^2 - \varepsilon, \theta^2 + \varepsilon]$ for all $s \in [t_1, t_2]$, so $\mathbb{P}(E) \to 1$. On this event, the number of times that G_t changes (in which case it changes by -1) is stochastically dominated above and below by a Poisson random variable with mean $(t_2 - t_1)(1 - \theta^2(c) \pm \varepsilon)$. The lemma follows easily. □

Lemma 6.

$$\mathbb{E}\left(\sup_{t\leq cn} |N(t) - \bar{N}(t)|\right) \leq (4c+1)n^{1/2}.$$

Proof. We already know that $N(t) \geq \bar{N}(t)$ for all $t \geq 0$. It thus suffices to show that the excess number of cycles is never more than $(4c+1)n^{1/2}$ in expectation. Call a cycle good if it never experienced a fragmentation and let $N^{go}(t)$ be the number of good cycles at time t. Call it excess otherwise, and let $N^{ex}(t)$ be the number of such cycles at time t. It is clear that points in a standard cycle are part of the same connected component in G_t, which does not contain any additional vertex. Thus every good cycle may be associated to one connected component of G_t in such a way that no two good cycles are associated to the same component of G_t. Hence $N(t) = N^{go}(t) + N^{ex}(t) \leq \bar{N}(t) + N^{ex}(t)$, so that $|N(t) - \bar{N}(t)| \leq N^{ex}(t)$. Note also that there can never be more than $n^{1/2}$ cycles of size greater than $n^{1/2}$. Thus it suffices to count the number $N_{\downarrow}^{ex}(t)$ of excess cycles of size $\leq n^{1/2}$:

$$|N(t) - \bar{N}(t)| \leq N_{\downarrow}^{ex}(t) + n^{1/2}.$$

These excess cycles of size $\leq n^{1/2}$ at time t must have been generated by a fragmentation at some time $s \leq t$ where one of the two pieces was smaller than $n^{1/2}$. But at each step, the probability of making such a fragmentation is smaller than $2n^{-1/2}$. Indeed, given the position of the first marker i, there are at most $4n^{1/2}$ possible choices for j which result in a fragmentation where one of the two pieces is of size smaller than $n^{1/2}$. To see this, note that if a transposition (i, j) is applied to a permutation σ, and $C_i = C_j$, so $\sigma^k(i) = j$ for some $1 \leq k \leq |C|$, then the two pieces are precisely given by $(\sigma^0(i), \ldots, \sigma^{k-1}(i))$ and $(\sigma^0(j), \ldots, \sigma^{|C|-k-1}(j))$. Thus to obtain two pieces of size k and $|C| - k$ there are at most two possible choices, which are $\sigma^k(i)$ and $\sigma^{-k}(i)$. Thus $\mathbb{E}(F_{\downarrow}(cn)) \leq cn \cdot 4n^{-1/2}$, where $F_{\downarrow}(cn)$ is the total number of fragmentation events where one of the pieces is smaller than $n^{1/2}$ by time cn. Since

$$\sup_{t\leq cn} N_{\downarrow}^{ex}(t) \leq F_{\downarrow}(cn)$$

this finishes the proof. □

Proof of Theorem 1. Applying Markov's inequality in Lemma 6, we see that since $n^{1/2} \ll n^{3/4} = t_2 - t_1$, we also have

$$N(t_1) - N(t_2) \sim (t_2 - t_1)(1 - \theta^2(c))$$

in probability, by Lemma 5. On the other hand, $N(t)$ changes by -1 in the case of a coalescence and by +1 in the case of a fragmentations. Hence $N(t_1) - N(t_2) = \text{Poisson}(t_2 - t_1) - 2F(I)$, where $F(I)$ is the total number of fragmentations during the interval I. We therefore obtain by the law of large numbers for Poisson random variables:

$$F(I) \sim \frac{1}{2}(t_2 - t_1)\theta(c)^2.$$

But observe that if $F(I)$ is large, it cannot be the case that all cycles are small - otherwise we would very rarely pick i and j in the same cycle. Hence consider the decreasing events $E_s = \{\tau_\delta > s\}$ for $s \in [t_1, t_2]$ and let $E = E_{t_2}$. On E_s, the maximal cycle size up to time s is no more than δn. Hence at the next transposition, the probability of making a fragmentation is no more than δ. We deduce that on the event E, $F(I)$ is stochastically dominated by a Poisson random variable with mean $(t_2 - t_1)\delta$. Thus, on $E \cap E'$ where $\mathbb{P}(E') \to 1$, $F(I) \leq 2\delta(t_2 - t_1)$. Since $2\delta = \theta(c)^2/4$, it follows immediately that $\mathbb{P}(E) \to 0$ as $n \to \infty$. This completes the proof. □

Remark 7. *This proof is partly inspired by the calculations in Lemma 8 of* [2].

3 Random hypergraphs and Theorem 3.

We now start the proof of Theorem 3 in the case of random k-cycles. We first review some relevant definitions and results from random hypergraphs.

A *hypergraph* is a graph where edges can connect several vertices at the same time. Formally:

Definition 1. *A hypergraph $H = (V, E)$ is given by a set V of vertices and a subset E of $\mathscr{P}(V)$, where $\mathscr{P}(V)$ denotes the set of all subsets of V. The elements of E are called hyperedges. A d-regular hypergraph is a hypergraph where all edges connect d vertices, i.e. for all $e \in E$, $|e| = d$.*

For a given $d \geq 2$ and $0 < p < 1$, we call $\mathbf{G}_d(n, p)$ the probability distribution on d-regular hypergraphs on $V = \{1, \ldots, n\}$ where each hyperedge on d vertices is present independently of the other hyperedges with probability p. Observe that when $d = 2$ this is just the usual Erdős-Renyi random graph case, since a hyperedge connecting two vertices is nothing else than a usual edge. For basic facts on Erdős-Renyi random graphs, see e.g. [5].

The notion of a hypertree needs to be carefully formulated in what follows. We start with the d-regular case. The excess $ex(H)$ of a given d-regular hypergraph H is defined to be

$$ex(H) = (d-1)h - r \tag{9}$$

where $r = |H|$ and h is the number of edges in H.

Observe that if H is connected then $ex(H) \geq -1$ as can be easily seen by induction on h (each new hyperedge adds at most $d-1$ new vertices to H).

Definition 2. *We call a connected d-regular hypergraph H a* hypertree *if $ex(H) = -1$.*

Likewise if $ex(H) = 0$ and H is connected we will say that H is *unicyclic* and if the excess is positive we will say that the component is *complex*.

Remark 8. *This is the definition used by Karoński and Luczak in [13], but differs from the definition in their older paper [11] where a hypertree is a connected hypergraph such that removing any hyperedge would make it disconnected.*

In the case where H is not necessarily regular, the excess of a connected hypergraph H made up of the hyperedges $h_1, \ldots, h_n$ is defined to be $ex(H) = \sum_{i=1}^{n}(|h_i| - 1) - |H|$, where $|h_i|$ denotes the size of the hyperedge h_i and $|H|$ is the cardinality of the vertex set of H. Then $ex(H) \geq -1$ and H is said to be a hypertree if $ex(H) = -1$.

3.1 Critical point for random hypergraphs

We start by recalling a theorem by Karoński and Luczak [13] concerning the emergence of a giant connected component in a random hypergraph process $(H_t, t \geq 0)$ where random hyperedges of degree $d \geq 2$ are added at rate 1.

Theorem 9. *Let $c > 0$ and let $t = cn$.*

- *When $c < c_d = 1/[d(d-1)]$ then a.a.s then H_t contains only trees and unicyclic components. The largest component has size $O(\log n)$ with high probability.*
- *When $c > c_d$ then there is with high probability a unique complex component, of size θn asymptotically, where $\theta = \theta_d(c) > 0$. All other component are not larger than $O(\log n)$ with high probability.*

Note that with high probability, if $c < c_d$ then the number of unicyclic components is no more than $C' \log n$ for some $C' > 0$ which depends on c. Indeed, at each step the probability of creating a cycle is bounded above by $C \log n/n$ since the largest component is no more than $O(\log n)$ prior to time cn. Since there are $O(n)$ steps this proves the claim. We will need a result about the evolution of the number of components $\bar{N}(t)$ in $(H_t, t \geq 0)$.

Proposition 10. *Let $t > 0$. Then as $n \to \infty$,*

$$\frac{1}{n}\bar{N}(tn) \longrightarrow \sum_{h=0}^{\infty} \frac{((d-1)h+1)^{h-2}}{h!}(dt)^h e^{-dt(h(d-1)+1)} \tag{10}$$

uniformly on compacts in probability.

Proof. By Theorem 9 and since there are no more than $C \log n$ complex components, it is enough to count the number of hypertrees $\tilde{N}(s)$ in H_s where $s = tn$. We will first compute the expected value and then prove a law of large numbers using a second moment method.

Let $h \geq 0$, we first compute the number of hypertrees with h hyperedges ($h = 0$ corresponds to isolated vertices). These have $r = (d-1)h+1$ vertices. By Lemma 1 in Karoński-Luczak [12], there are

$$\frac{(r-1)!r^{h-1}}{h![(d-1)!]^h} \tag{11}$$

trees on $r = (d-1)h+1$ labeled vertices (this is the analogue to Cayley's (1889) well-known formula that there are k^{k-2} ways to draw a tree on k labeled vertices). If T is a given hypertree with h edges on vertices in $V = \{1, \ldots, n\}$, there are a certain number of conditions that must be fulfilled in order for T to be one of the components of H_s: (i) The h hyperedges of T must be open, (ii) $\binom{r}{d} - h$ hyperedges must be closed inside the rest of T, (iii) T must be disconnected from the rest of the graph. Thus hyperedges containing exactly $1, 2, \ldots, d-1$ vertices in T must be closed. This requires closing $\binom{r}{1}\binom{n-r}{d-1} + \binom{r}{2}\binom{n-r}{d-2} + \ldots + \binom{r}{d-1}\binom{n-r}{1} \sim r\binom{n-r}{d-1}$ hyperedges.

Now, remark that at time $s = tn$, because the individual Poisson clocks are independent, each hyperedge is present independently of the others with probability $p = 1 - \exp\left(-s/\binom{n}{d}\right) \sim d!t/n^{d-1}$. It follows that the probability that T is one of the components of H_t is

$$p^h(1-p)^{\binom{r}{d}-h+\binom{r}{1}\binom{n-r}{d-1}+\ldots+\binom{r}{d-1}\binom{n-r}{1}}. \tag{12}$$

Hence the expected number of trees in H_s with h edges is

$$\begin{aligned}\mathbb{E}[\tilde{N}_h(tn)] &= \binom{n}{r}\frac{(r-1)!r^{h-1}}{h![(d-1)!]^h}p^h(1-p)^{\binom{r}{d}-h+\binom{r}{1}\binom{n-r}{d-1}+\ldots+\binom{r}{d-1}\binom{n-r}{1}} \\ &\sim n\frac{r^{h-2}}{h!}(dt)^h e^{-drt}\end{aligned} \tag{13}$$

Write $\mathscr{C}$ for the set of connected components of H_s. Note that if T_1 and T_2 are two given hypertrees on V with disjoint vertex sets and with h hyperedges each, then

$$\mathbb{P}(T_1 \in \mathscr{C} \text{ and } T_2 \in \mathscr{C}) = \frac{\mathbb{P}(T \in \mathscr{C})^2}{(1-p)^{e(T_1,T_2)}},$$

where $e(T_1,T_2)$ denotes the number of hyperedges in V that intersect both T_1 and T_2. Hence $e(T_1,T_2) = \sum_{1\le j\le d-1}\sum_{1\le k\le d-j}\binom{r}{j}\binom{r}{k}\binom{n-r}{d-j-k} \sim r^2 n^{d-2}$. Since $p = O(n^{-(d-1)})$, we deduce that $\mathrm{cov}(\mathbf{1}_{\{T_1\in\mathscr{C}\}}, \mathbf{1}_{\{T_2\in\mathscr{C}\}}) \to 0$ uniformly in $h \le h_0$, for all fixed $h_0 \ge 0$. On the other hand, if T_1 and T_2 are distinct hypertrees that share at least one common vertex, then $\mathbb{P}(T_1 \in \mathscr{C}; T_2 \in \mathscr{C}) = 0$ and hence $\mathrm{cov}(\mathbf{1}_{\{T_1\in\mathscr{C}\}}, \mathbf{1}_{\{T_2\in\mathscr{C}\}}) \le 0$. Therefore, $\mathrm{var}(\tilde{N}_h(s)) = o(n^2)$. Thus, by Chebyshev's inequality, for all $h_0 \ge 0$:

$$\frac{1}{n}\sum_{h=0}^{h_0}\tilde{N}_h(s) \longrightarrow_p \sum_{h=0}^{h_0}\frac{((d-1)h+1)^{h-2}}{h!}(dt)^h e^{-dt(h(d-1)+1)}, \tag{14}$$

in probability as $n \to \infty$. Let $\tilde{N}_{>h_0}(s)$ be the number of trees greater than h_0. Then we obviously have $\tilde{N}_{>h_0}(s) \le n/h_0$. Choosing h_0 large enough that $1/h_0 < \varepsilon$ and the finite sum in the right-hand side of (14) lies within ε of the infinite series, the proof of the pointwise convergence in (10) follows directly. Since $n^{-1}\tilde{N}(tn)$ is a decreasing function of time for every n and since the limiting function in the right-hand side of (10) is continuous, the uniform convergence on compact sets in probability follows automatically (see, e.g., [7]). □

3.2 Bounds for the Cayley distance on the symmetric group

In the case of random transpositions we had the convenient formula that if $\sigma \in \mathscr{S}_n$ then $d(o,\sigma) = n - N(\sigma)$ where $N(\sigma)$ is the number of cycles of σ, a formula originally due to Cayley. In the case of random k-cycles with $k \ge 3$, unfortunately there is to our knowledge no exact formula to work with. However this formula stays approximately true, as shown by the following proposition.

Proposition 11. *Let $k \ge 3$ and let $\sigma \in \mathscr{S}_n$. (If k is odd, assume further that $\sigma \in \mathscr{A}_n$). Then*

$$\frac{1}{k-1}(n - N(\sigma)) \le d(o,\sigma) \le \frac{1}{k-1}(n - N(\sigma)) + C(k)|R_k(\sigma)|$$

where $C(k)$ is a universal constant depending only on k, and $R_k(\sigma)$ is the set of cycles of σ whose length $\ell \neq 1 \bmod k-1$.

Proof. We start by proving the lower-bound. Note that multiplication of σ by a k-cycle can increase the number of cycles by at most $k-1$. Hence, after p multiplications the resulting permutation cannot have more than $N(\sigma) + p(k-1)$ cycles. Therefore the distance must be at least that k_0 for

which $N(\sigma) + k_0(p-1) \geq n$, since the identity permutation has exactly n cycles. The lower-bound follows.

For the upper-bound, we need a few notations. Suppose $c = (x_1, \ldots, x_\ell)$ is a cycle of length $\ell = d(k-1)+r$, where $r \in [2, k-1]$. [This is not exactly the usual Euclidean division of ℓ by $k-1$, as we choose $r = k-1$ rather than $r = 0$ if ℓ is a multiple of $k-1$.] Let $D(c) = (x_1, \ldots, x_{d(k-1)+1})$, and let $R(c) = (x_{d(k-1)+1}, \ldots, x_\ell)$. Note that $c = D(c) \circ R(c)$, the product of the permutation $D(c)$ with $R(c)$. Since cycles with disjoint support commute, we may write $\sigma = D(\sigma) \circ R(\sigma)$ with $D(\sigma) = \prod_{c \in \sigma} D(c)$ and $R(\sigma) = \prod_{c \in \sigma} R(c)$. Now notice that for a cycle $c = (x_1, \ldots, x_\ell)$ of length $\ell = d(k-1) + r$ as above, if $r = 1$ we may write

$$c = (x_1, \ldots x_k) \circ (x_k, \ldots, x_{2(k-1)+1}) \circ \ldots \circ (x_{(d-1)(k-1)+1}, \ldots, x_{d(k-1)+1})$$

and thus obtain c as a product of k-cycles with exactly d factors. Thus the permutation $D(\sigma)$ may be constructed as a product of k-cycles $D(\sigma) = \pi_1 \ldots \pi_d$ containing exactly

$$d = \sum_{c \in \sigma} \frac{|D(c)| - 1}{k-1} \leq \sum_{c \in \sigma} \frac{|c| - 1}{k-1} = \frac{n - N(\sigma)}{k-1}$$

factors, where $|D(c)|$ denotes the length of the cycle $D(c)$. The permutation $R(\sigma)$, on the other hand, may be written as a product of k-cycles $R(\sigma) = \pi'_1 \ldots \pi'_r$, where $r \leq C(k)|R_k(\sigma)|$. To see this, note that only cycles $c \notin R_k(\sigma)$ contribute a term to $R(\sigma)$. This term is then a cycle of length $|R(c)| \in [2, k-1]$. Then if k is even, one can take $C(k)$ to be the maximal number of k-cycles needed to create a cycle of length between 2 and $k-1$, which is less than (say) the diameter of $\mathscr{S}_k$ generated by the set of k-cycles. If k is odd, so that $\sigma \in \mathscr{A}_n$, then note that $D(\sigma) \in \mathscr{A}_n$ and hence $R(\sigma) \in \mathscr{A}_n$. Thus there is an even number of cycles $c \in \sigma$ such that $|R(c)|$ is even. By considering such cycles in pairs, whose product is in $\mathscr{A}_n$, we see that we can take $C(k)$ to be the diameter of $\mathscr{A}_k$ generated by k-cycles. Hence we may construct a path of length $d + r$ between the identity and σ by considering $\sigma = \pi_1 \ldots \pi_d \pi'_1 \ldots \pi'_r$. Since $d \leq (n - N(\sigma))/(k-1)$ and $r \leq C(k)|R_k(\sigma)|$, the upper-bound is proved. □

3.3 Phase transition for the 3-cycle random walk

We now finish the proof of Theorem 3 in the case of random k-cycles.

Proof of Theorem 3 if $c_j = \mathbf{1}_{\{j=k\}}$. The proof follows the lines of Lemma 6. Let $N(t)$ be the number of cycles of σ and let $\bar{N}(t)$ be the number of components in H_t, where $(H_t, t \geq 0)$ is the random k-regular hypergraph process obtained by adding the edge $\{x_1, \ldots, x_k\}$ whenever the k-cycle $(x_1, \ldots, x_k)$ is performed. Note again that every cycle of σ_t is a subset of a connected component of H_t, so $N(t) \geq \bar{N}(t)$. (Indeed, this property is a deterministic statement for transpositions, and a sequence of random k-cycles can be decomposed as a sequence $(k-1)$ times as long of transpositions.)

Repeating the argument in Lemma 6, we see that

$$n^{-3/4} \left(\sup_{t \leq cn} |N(t) - \bar{N}(t)| \right) \to 0, \tag{15}$$

in probability. This is proved in greater generality (i.e., for arbitrary conjugacy classes) in Lemma 14. Moreover, any $c \in R_k(\sigma_t)$ must have been generated by fragmentation at some point (otherwise the length of cycles only increases by $k-1$ each time). Thus $R_k(\sigma_t) \leq N(t) - \bar{N}(t)$, and Theorem 3 now follows.

□

4 Proofs for general conjugacy classes

4.1 Random graph estimates

Let $\Gamma = (c_2, \ldots, c_J)$ be our fixed conjugacy class. A first step in the proof of Theorems 3 and 2 in this general case is again to associate a certain random graph model to the random walk. As usual, we put a hyperedge connecting $x_1, \ldots, x_k$ every time a cycle $(x_1 \ldots x_k)$ is applied as part of a step of the random walk. Let H_s be the random graph on n vertices that is obtained at time s. A first step will to prove properties of this random graph H_s when $s = tn$ for some constant $t > 0$. Recall our definition of t_c:

$$t_c^{-1} = \sum_{j=2}^{J} c_j j(j-1), \tag{16}$$

and that $1 - \theta(t)$ be the smallest solution of the equation (in z): $G_t(z) = z$, where

$$G_t(z) = \exp(t \sum_{j=1}^{J} j c_j (z^{j-1} - 1)). \tag{17}$$

We will see that $G_t(z)$ is the generating function of the degree of any vertex at time tn.

Lemma 12. *If $t < t_c$ then there exists $\beta > 0$ such that all clusters of H_{tn} are smaller than $\beta \log n$ with high probability. If $t > t_c$, then there exists $\beta > 0$ such that all but one clusters are smaller than $\beta \log n$ and the largest cluster $L_n(t)$ satisfies*

$$\frac{L_n(t)}{n} \longrightarrow \theta(t)$$

as $n \to \infty$ in probability.

Proof. We first consider a particular vertex, say $v \in V$, and ask what is its degree distribution in H_{tn}. Write $\sigma_t = \gamma_1 \ldots \gamma_{N_t}$ where $(\gamma_i, i \geq 1)$ is a sequence of i.i.d. permutations uniformly distributed on Γ, and $(N_t, t \geq 0)$ is an independent Poisson process. Note that for $t \geq 0$, $\#\{1 \leq i \leq N_t : v \in \mathrm{Supp}(\gamma_i)\}$ is a Poisson random variable with mean $t \sum_{j=2}^{J} j c_j / n$. Thus by time tn, the number of times v has been touched by one of the γ_i is a Poisson random variable with mean $t \sum_{j=2}^{J} j c_j$. For each such γ_i, the probability that v was involved in a cycle of size exactly ℓ is precisely $\ell c_\ell / \sum_{j=2}^{J} j c_j$. Thus, the number of hyperedges of size j that contain v in H_{tn} is P_j, where $(P_j, j = 2, \ldots, J)$ are independent Poisson random variables with parameter $t j c_j$. Since each hyperedge of size j corresponds to $j-1$ vertices, we see that the degree of v in in H_{tn}, D_v, has a distribution given by

$$D_v = \sum_{j=2}^{\ell} (j-1) P_j. \tag{18}$$

Now, note that by definition of t_c (see (16)),

$$\mathbb{E}(D_v) > 1 \iff t > t_c.$$

The proof of Theorem 3.2.2 in Durrett [10] may be adapted almost *verbatim* to show that there is a giant component if and only if $\mathbb{E}(D_v) > 1$, and that the fraction of vertices in the giant component is the survival probability of the associated branching process. Note that the generating function associated with the progeny (18) is

$$G_t(z) := \mathbb{E}(z^D) = \prod_{j=2}^{J} \mathbb{E}(z^{(j-1)P_j}) = \prod_{j=2}^{\ell} \exp(tjc_j(z^{j-1} - 1))$$

thus $1 - \theta(t)$ is the smallest root of the equation $G_t(z) = z$. From the same result one also gets that the second largest cluster is of size no more than $\beta \log n$ with high probability, for some $\beta > 0$. □

Let $\bar{N}(s)$ be the number of clusters at time s in H_s, and let

$$u(t) = 1 - \int_0^t \sum_{j=2}^{J} c_j(j\rho(s) - \rho(s)^j)ds, \tag{19}$$

where $\rho(s) = 1 - \theta(s)$. Note that that $u(0) = 1$, and for $t \leq t_c$ we have $u(t) = 1 - Kt$ where $K := \sum_{j=2}^{J} c_j(j-1)$ since $\rho(t) = 1$ for $t \leq t_c$. Moreover $u(t) > 1 - Kt$ for $t > t_c$.

Lemma 13. *As $n \to \infty$, we have*

$$\frac{1}{n}\bar{N}(tn) \longrightarrow u(t),$$

uniformly on compacts in probability.

Proof. Let H denote a hypergraph on $\{1, \ldots, n\}$, and let $h = h_1 \cup \ldots \cup h_\ell$ be a set of hyperedges. Denote by $H' = H + h$ the graph obtained from H by adding the hyperedges $h_1, \ldots, h_\ell$ to H. Let $(x_1, \ldots, x_n)$ be a discrete partition of unity, i.e., a non-increasing sequence of numbers such that $\sum_{i=1}^{n} x_i = 1$ and such that for all $1 \leq i \leq n$, nx_i is a nonnegative integer. Define a function $f(x_1, \ldots, x_n)$ as follows. Let H be any hypergraph for which x_i are the normalized cluster sizes. Let $h = h_1 \cup \ldots \cup h_\ell$ be a collection of hyperedges of sizes $2, 3, \ldots, J$ (with size j being of multiplicity c_j), where the hyperedges h_i are sampled uniformly at random without replacement from $\{1, \ldots, n\}$. Let $H' = H + h$. Then we define f by putting

$$f(x_1, \ldots, x_n) := \mathbb{E}(|H'| - |H|)$$

where $|H|$ denotes the number of clusters of H. Then we have that

$$M_t := \frac{1}{n}|H(tn)| - \int_0^t f(x_1(sn), \ldots, x_n(sn))ds \tag{20}$$

is a martingale, if $(x_1(s), \ldots, x_n(sn))$ denote the ordered normalized cluster sizes of $H(s)$. (Note that $M_0 = 1$.)

We claim that, as $n \to \infty$, for every s fixed,

$$f(x_1(sn), \ldots, x_n(sn)) \to -\sum_{j=2}^{J} c_j (j\rho(s) - \rho(s)^j) \tag{21}$$

in probability, where $\rho(s) = 1 - \theta(s)$. To see this, take $h = h_1 \cup \ldots \cup h_\ell$ a collection of random hyperedges as above, and for $1 \le i \le \ell$, write $h_i = \{x_1^i, \ldots, x_j^i\}$ if h_i is of size j. Let $\mathscr{A}_i$ be the event that no two points of h_i fall in the same nongiant component: that is, if x_q^i and $x_{q'}^i$ are not in the largest component and $q \neq q'$, then the components containing x_q^i and $x_{q'}^i$ are distinct. Let $\mathscr{A} = \cap_{i=1}^{\ell} \mathscr{A}_i$. Note that by Lemma 12, $\mathbb{P}(\mathscr{A}) \to 1$. Moreover, on $\mathscr{A}$, then $|H'| - |H| = \sum_{i=1}^{\ell} X_i$. Here $X_i = Y_i \wedge (j-1)$, where Y_i is the number of points of h_i that do not fall in the largest component of H (and j is the size of h_i). Thus

$$\mathbb{E}(X_i) \to \mathbb{E}((j-1) \wedge \text{Binomial}(j, \rho)) = j\rho - \rho^j.$$

Since there are c_j hyperedges of size j, (21) follows. Note that the quantity $f(x_1(sn), \ldots, x_n(sn))$ is a.s. monotone as a function of s because $H(sn)$ is a purely coalescing process, and that the right-hand side of (21) is continuous. We deduce immediately that the convergence (21) holds uniformly on compacts in probability, and hence also, trivially,

$$\int_0^t f(x_1(sn), \ldots) ds \longrightarrow -\int_0^t \sum_{j=2}^{J} c_j (j\rho(s) - \rho(s)^j) ds, \tag{22}$$

uniformly on compacts in probability.

Moreover, note that

$$\text{var}(|H'| - |H|) \le C \tag{23}$$

for some constant C which depends only on $(c_2, \ldots, c_J)$, since $|H'|$ may differ from $|H|$ only by a bounded amount. By Doob's inequality, if $\bar{M}_s = n(M_s - 1)$:

$$\begin{aligned} \mathbb{P}\left(\sup_{s \le t} |(M_s - 1)| > \varepsilon \right) &= \mathbb{P}\left(\sup_{s \le t} |\bar{M}_s|^2 > n^2 \varepsilon^2 \right) \\ &\le \frac{4 \,\text{var}(\bar{M}_t)}{n^2 \varepsilon^2} \\ &\le \frac{4Ct}{n\varepsilon^2}. \end{aligned} \tag{24}$$

The last line inequality is obtained by conditioning on the number of steps N between times 0 and tn, noting that after each step, the variance of $\bar{M}_t$ increases by at most C by (23). (The value of the constant C may change from line to line). Combining (24), (22) and the definition of M in (20), we obtain the statement of Lemma 13. □

4.2 Random walk estimates

Lemma 14. *Let $N(t)$ be the number of cycles of $\sigma(tn)$. Then we have, as $n \to \infty$:*

$$\frac{1}{n^{3/4}} (N(tn) - \bar{N}(tn)) \longrightarrow 0, \quad u.c.p. \tag{25}$$

Proof. This is very similar to Lemma 6. Say that a cycle is large or small, depending on whether it is bigger or smaller than $\sqrt{n}$. To start with, observe that there can never be more than $\sqrt{n}$ large cycles. As usual, we have that $N(t) \geq \bar{N}(t)$, and we let $N^{ex}(t)$ be the number of excess cycles, i.e., those which have experienced at least one fragmentation during their evolution up to time t. Then we have, as before, $N - \bar{N}(t) \leq N^{ex}(t)$. We can in turn decompose $N^{ex}(t) = N^{ex}_{\uparrow}(t) + N^{ex}_{\downarrow}(t)$, where the subscripts $\uparrow$ and $\downarrow$ refer to the fact that the cycles are either small or large. Thus we have

$$N^{ex}_{\uparrow}(t) \leq \sqrt{n},$$

and the problem is to control $N^{ex}_{\downarrow}(t)$. Writing every cycle of size j as a product of $j-1$ transpositions, we may thus write $\sigma_t = \prod_{i=1}^{m_t} \tau_i$, for a sequence of transpositions having a certain distribution (they are not independent). Then $N^{ex}_{\downarrow}(t) \leq F_{\downarrow}(t)$, where $F_{\downarrow}(t)$ is the number of times $1 \leq i \leq m_t$ that the transposition τ_i yields a fragmentation event for which one of the fragments is small. Now, conditionally on $\tau_1, \ldots, \tau_{i-1}$, the conditional probability that τ_i yields such a fragmentation is at most $4n^{1/2}/(n-(2K+1))$, where as before $K = \sum_{j=2}^{J}(j-1)c_j$. Indeed, even though the τ_j are not independent, the following holds. Write $\tau_j = (x_j, y_j)$ for all $j \geq 1$. Then given $\tau_1, \ldots, \tau_{i-1}$, and given x_i, then y_i is sampled uniformly from $\{1, \ldots, n\} \setminus I$ where $I = \{x_j, y_j, i' \leq j \leq i-1\} \cup \{x_i\}$ and $i' = \sup\{j \leq i, j \text{ multiple of } K\}$. (Thus I never contains more than $2K+1$ points). Moreover, at most $4\sqrt{n}$ choices for y_j will generate a fragmentation where one of the pieces is small.

Since $m_t = KN_t$, where N_t is a Poisson random variable with mean t, it follows that for n large enough

$$\mathbb{E}(\sup_{s \leq tn} F_{\downarrow}(s)) \leq Ktn \frac{4n^{1/2}}{n-(2K+1)} \leq 5Kt\sqrt{n}.$$

Thus by Markov's inequality,

$$\mathbb{P}\left(\sup_{s \leq tn} F_{\downarrow}(s) > n^{3/4}\right) \longrightarrow 0. \tag{26}$$

Hence, $n^{-3/4}|N(tn) - \bar{N}(tn)|$ converges to 0 u.c.p, which concludes the proof by Lemma 14. □

Note in particular that by combining Lemma 13 with Lemma 14, we get that

$$\frac{1}{n}N(tn) \to u(t), \quad u.c.p. \tag{27}$$

Lemma 15. *Let $t > t_c$. Then $\tau_\delta < tn$ with high probability, where*

$$\delta := \frac{1}{3Kt}\int_0^t \theta^2(s)ds > 0, \tag{28}$$

where $K = \sum_{j=2}^{J}(j-1)c_j$.

Remark 16. *Note that Lemma 15 immediately implies Theorem 2.*

Proof. The proof is very similar to the proof of Theorem 1, with an additional martingale argument to get sufficiently good concentration properties (this part of the argument being quite similar to Lemma 13).

Assume that a permutation σ has a cycle structure $(C_1, \dots, C_r)$ and that $x_1, \dots, x_r$ are the normalized cycle sizes, i.e., $x_i = |C_i|/n$. Define a function $g(x_1, \dots, x_r)$ by putting

$$g(x_1, \dots, x_r) := \mathbb{E}(N(\sigma') - N(\sigma)),$$

where $\sigma' = \sigma \circ \gamma$ and γ is a uniform random element from Γ, while as usual $N(\sigma)$ denotes the number of cycles of σ. Then if we define a process

$$M'_t = \frac{1}{n} N(tn) - \int_0^t g(x_1(sn), \dots) ds,$$

then $(M'_t, t \geq 0)$ is a martingale started from $M'_0 = 1$. Moreover, writing $\gamma = \tau_1 \circ \dots \circ \tau_K$, where τ_i are transpositions, and if we let $\sigma_i = \sigma \circ \tau_1 \dots \tau_i$, so that $\sigma_0 = \sigma$ and $\sigma_K = \sigma'$, then

$$g(x_1, \dots, x_r) = \sum_{i=1}^{K} \mathbb{E}(N(\sigma_i) - N(\sigma_{i-1})).$$

Recall that the transposition τ_i can only cause a coalescence or a fragmentation, in which case the number of cycles decreases or increases by 1. If the normalized cycle sizes of σ_{i-1} are given by $(y_1, \dots, y_r)$, it follows that

$$-1 \leq \mathbb{E}(N(\sigma_i) - N(\sigma_{i-1})) \leq -1 + 2y_i^* \frac{n}{n-i+1},$$

where $y_i^* = \max(y_1, \dots, y_r)$. Moreover, $y_i^* \leq i y_0^*$.

From this we obtain directly that with probability 1, for n sufficiently large (uniformly on $t \geq 0$)

$$\int_0^t g(x_1(sn), \dots) ds \leq \int_0^t K \left[-1 + 3K x^*(sn) \right] ds, \tag{29}$$

where $x^*(s) = \max(x_1(s), \dots, x_r(s))$. On the other hand, using Doob's inequality in the same way as (24), we also have:

$$\mathbb{P}\left(\sup_{s \leq t} |(M'_s - 1)| > \varepsilon \right) \leq \frac{4C}{n\varepsilon^2}. \tag{30}$$

Combining this information with (27), we obtain, with high probability uniformly on compact sets:

$$\int_0^t \left[-1 + 3K x^*(sn) \right] ds \geq \int_0^t -1 + \theta^2(s) ds. \tag{31}$$

From this we get, with high probability,

$$3Kt \sup_{s \leq tn} x^*(s) \geq \int_0^t \theta^2(s) ds, \tag{32}$$

i.e., $\tau_\delta \leq tn$. □

4.3 Distance estimates

We are now ready to prove that

$$d(\sigma_{tn}) \longrightarrow \varphi(t),$$

uniformly on compact sets in probability as $n \to \infty$ except possibly on some interval $I \subset (t_c, \infty)$ bounded away from t_c and ∞, where

$$\varphi(t) = \frac{1-u(t)}{K} = \int_0^t \sum_{j=2}^{J} c_j (j\rho(s) - \rho(s)^j) ds. \tag{33}$$

The proof is analogous but more complicated (especially the upper-bound) than that of Proposition 11.

Proof of lower-bound. Note that if σ is a permutation, every transposition can at most increase the number of cycles by 1. Hence if σ has $N(\sigma)$ cycles, after one step $s \in \Gamma$, σ has at most $N(\sigma) + K$ cycles. Thus after p steps, the number of cycles of σ is at most $N(\sigma) + Kp$. Since the identity permutation has exactly n cycles, we conclude that

$$d(\sigma) \geq \frac{1}{K}(n - N(\sigma)). \tag{34}$$

Together with Lemma 14 and the definition of $\varphi(t)$, this proves the lower bound in Theorem 3.

Note that the bound (34) would be sharp if we could find a path from σ to the identity o which produces K fragmentations at every step. The rest of this section is devoted to the upper-bound, which shows that indeed such a path may be found except for an additional $o(n)$ coagulation steps, which is also enough. To do this we use a kind of greedy algorithm. Before we can explain the idea, we need a few definitions. Call a component of H_t *good* if it is a hypertree and bad otherwise; a hyperedge is good if its component is good. Likewise, call a cycle C of $\sigma(t)$ good if its associated component $\bar{C}$ in H_t is a hypertree. Therefore, a good cycle is one which has never been involved in fragmentations, i.e., its history consists only of coagulation events. Given a good cycle C of $\sigma(t)$ we can write uniquely $C = c_1 \dots c_r$, where the c_i are the cycles whose coagulations yielded C, written in the order in which they appeared. We call the collection $(c_i)_{1 \leq i \leq r}$ the subcycles of C. Finally, say that a cycle is bad if it is not good.

4.3.1 Heuristics

We now proceed to describe heuristically our greedy algorithm for constructing a path between σ_{tn} and the identity e. Naively, we think of each step as a combination of transpositions, and try to use each such transposition to make a fragmentation. However, this has to be done carefully. Essentially, good cycles need to be destroyed (i.e., fragmented) by reversing the way they were created, as made precise by the following.

Fragmentation of good cycles. Let $C = c_1 \dots c_r$ be a good cycle, where $(c_i)_{1 \leq i \leq r}$ are its subcycles. For $1 \leq i \leq r$, let $c_i' = c_i^{-1}$, and consider $C \circ c_i'$: this is a permutation which has at most two non trivial cycles, and each of which may be written as a product of a total of $r-1$ subcycles, whose sizes are identical to $|c_1|, \dots, |c_r|$ except for $|c_i|$. We will say that we have *destroyed* the subcycle c_i, and still speak of the two cycles of $C \circ c_i'$ as good, and talk freely about their own subcycles. Repeating this process, we are able to fragment C entirely without making a single coagulation.

Fragmentation of bad cycles. On the other hand, to destroy a bad cycle $C = (x_1, \dots, x_\ell)$, where ℓ is typically very big, we can simply compose C with $(x_j, \dots, x_1)$ where $2 \le j \le J$. This has the effect of reducing the length of C by $j-1$. As a consequence, for an arbitrary cycle C, we get that

$$C \text{ can be destroyed in at most } \frac{|C|}{K} + O(1) \text{ steps from } \Gamma, \tag{35}$$

where the term $O(1)$ is nonrandom, uniformly bounded in $|C|$ and n (but depends on Γ).

Order in which cycles are destroyed. It is tempting to first destroy all the good cycles and then all the bad cycles. This is roughly what we will do. However, a difficulty arises in the case where Γ contains cycles of different sizes. To illustrate the point, imagine that Γ is such that $c_2 = c_3 = 1$ and there are no other nontrivial cycles in Γ. The subcycles of every good cycle consist of a number of 2-subcycles and a number of 3-subcycles. These numbers may not be equal for every cycle, but the overall frequency of 2-subcycles and 3-subcycles is equal when $t < t_c$. However, if $t > t_c$, then the graph H_{tn} contains a giant component, and nongiant components of H_{tn} are now relatively less likely to contain a 3-hyperedge. Thus the frequency of 3-subcycles will be strictly less than the frequency of 2-subcycles. Hence when $t > t_c$ it is impossible to destroy all good cycles without touching the bad cycles at the same time (because every random walk step must use the same number of 3-cycles and 2-cycles, as these are the allowed steps under Γ). We will have to do so in a way that minimizes the impact on good cycles, using bad cycles (which are large) as a "bank" when we need to complete a step and no good subcycle of the required size remains.

4.3.2 Subcycle count.

To start with, we need the following lemma, which tells us the relative number of good subcycles of size j for all $2 \le j \le J$.

Lemma 17. *Fix $t > 0$. Let $j \ge 2$ such that $c_j > 0$. Then the number $U_j(tn)$ of good subcycles of size j in σ_{tn}, satisfies*

$$\frac{U_j(tn)}{n} \to c_j t \rho(t)^j, \tag{36}$$

uniformly on compacts in probability (u.c.p.), as $n \to \infty$.

Proof. It suffices to prove this result with $U_j'(tn)$, the number of good hyperedges in H_{tn} of size j, in place of $U_j(tn)$. The number of j-edges that have been added to H_{tn} is a Poisson random variable with mean tnc_j. For each such edge, the probability that it is not in the giant component W converges to $\rho(t)^j$.

To see this, note that by Lemma 12, it suffices to check that none of the j points are in a cluster of size greater than $\beta \log n$ for $\beta > 0$ large enough. Checking this will involve revealing the connections of no more than $j\beta \log n$ vertices in total. The exploration of the clusters can be done by using the standard breadth-first search procedure (see, e.g., Chapter 2.2 in [10]). The breadth-first search is well-approximated by a branching process with offspring distribution D_v from (18) until a vertex is sampled twice by this procedure. By the birthday problem, this does not occur until more than $o(\sqrt{n})$ vertices have been exposed with high probability. Since we are exposing at most only $j\beta \log n$ vertices, we can ignore this difference. Thus, the probability that none of those j vertices is contained in a cluster of size greater than $\beta \log n$ is the probability that among j independent branching processes, noone has a total progeny greater than $\beta \log n$. This converges as $n \to \infty$ towards $\rho(t)^j$.

Thus $\mathbb{E}(U_j'(tn)) \sim tnc_j\rho(t)^j$. Also, if e and e' are two randomly chosen j-edges, $\mathbb{P}(e \notin W, e' \notin W)$ converges for the same reasons to $\rho(t)^{2j}$, so that $\mathrm{cov}(\mathbf{1}_{\{e\subset W\}}, \mathbf{1}_{\{e'\subset W\}}) \to 0$. Thus the pointwise convergence follows from the second moment method. The convergence in the u.c.p. sense follows in a similar way after observing that if an edge added at time s is good at time t_2 then it is also good at time t_1, where $s \le t_1 \le t_2$ are arbitrary. □

4.3.3 Proof of Theorem 3

We now describe our algorithm more formally. We will assume for simplicity that Γ is an odd conjugacy class that generates all of $\mathcal{S}_n$, and that $c_2 > 0$ (the arguments below can easily be adapted otherwise).

Stage 1: reduction. Let $\sigma = \sigma_{tn}$, and work conditionally given $\mathcal{F}_{tn}$. Our first step is to coagulate in $o(n)$ steps all bad cycles together, so that we work with good small cycles and one giant bad cycle. Fix $t > 0$ and write (by analogy with the proof of Proposition 11) $\sigma = D(\sigma) \circ R(\sigma)$, where $D(\sigma)$ is the product of all good cycles of σ while $R(\sigma)$ is the product of all bad cycles, say there are r of them. Note that by (26), and recalling that there can never be more than $\sqrt{n}$ cycles greater or equal to $\sqrt{n}$, we have $r \le n^{3/4}$ say on an event of high probability uniformly on compacts. (If this doesn't hold, declare the algorithm a *failure*). Moreover the total unnormalized mass of nontrivial cycles in $R(\sigma)$ (i.e., the combined sizes of all cycles of size greater or equal to 2) is

$$|R(\sigma)| = \theta(t)n + o(n), \tag{37}$$

where $o(1)$ stands for a term which, when divided by n, converges to 0 in probability, u.c.p. Therefore (since $r \le n^{3/4}$), in less than $o(n)$ random walk steps, we can transform σ into $\sigma' = D(\sigma) \circ R'(\sigma)$, where $R'(\sigma)$ is the permutation obtained from $R(\sigma)$ by coagulating all its nontrivial cycles.

Stage 2: Elimination of good cycles.

We remove one by one every (good) subcycle of size 2 from σ'. For each random walk step $s = \gamma_1 \dots \gamma_p \in \Gamma$ that is applied, where $p = \sum_{j=2}^{J} c_j$, we use γ_i, $1 \le i \le p$, to destroy a subcycle of the size of $\ell_i = |\gamma_i|$, as described above in subsection 4.3.1. If no good subcycle of that size is available, we then turn $\ell_i - 1$ points from the bad cycle (bank) into fixed points, thereby reducing the size of the bad cycle (or bank) by as much. We do so until there are no good 2-subcycles left. Declare this a *failure* if at any time the bank runs out, i.e., if the bad cycle is not large enough to remove the required number of points from it. If not, let R_i denote the state of the bad cycle after the application of i random walk steps, and let $R_0 = R'(\sigma)$.

Let σ'' be the resulting permutation. We also declare this a *failure* if σ'' has a nontrivial cycle other than R_m, where m is the total number of random walk steps performed in this stage of the algorithm.

Stage 3: elimination of bad cycle remainder.

If the algorithm has not failed during step 2, there is only one nontrivial cycle $C = R_m$ left. We have already argued that we can destroy C in no more than $|C|/K + O(1)$ steps by (35), and thus make no more than $O(1)$ coagulations in this final step. At the end of step 3, the resulting permutation is necessarily the identity permutation, as all nontrivial cycles have been fragmented. Moreover, the total number of coalescences is no more than $o(n) + O(1) = o(n)$ (all coming from the first and final steps).

Check that the algorithm is successful. It now suffices to prove that the algorithm does not fail with high probability if $t < t_c + \varepsilon$ for some $\varepsilon > 0$ or if t is large enough. There are two possible reasons for failure: one is the the bank runs out of mass, the other is that by the end of stage 2, there is at least one other nontrivial cycle apart from $C = R_m$. But note that by Lemma 36, $U_j/c_j > U_2/c_2$ for all $2 < j \leq J$ such that $c_j > 0$, with high probability. On this event, by the time all good 2-cubcycles have been removed, all subcycles of larger sizes have also been removed: thus σ'' has only one nontrivial cycle, R_m.

We now turn to the other possible reason for a failure, which is that the bank runs out of mass. The total mass that is required from the bank in this procedure, $|R_0 \setminus R_m|$, is precisely

$$M = \sum_{j=2}^{J} (j-1) \left(\frac{c_j}{c_2} U_2(tn) - U_j(tn) \right).$$

The term of index j in this sum represents how much mass is taken away from the bank to cover for the deficit in subcycles of size j from the good cycles: there are U_2 subcycles of size 2 that need to be removed. For each 2-subcycle we remove, we also need to remove c_j/c_2 subcycles of size j. However, we have only U_j of them, so the mass that is borrowed from the bank is the term of index j in this sum. (Indeed we only take away $j-1$ points from the bad cycle when we do so).

By (37), the initial mass of the bank is $|R_0| = \theta(t)n + o(n)$, hence it suffices to show that M defined above satisfies

$$M \leq \theta(t)n + o(n) \tag{38}$$

with high probability. It thus remains solely to prove (38) if $t < t_c + \varepsilon$ for some $\varepsilon > 0$ and if t is sufficiently large. We start by the former. By Lemma 36, and using $1 - (1-x)^\alpha < \alpha x$ if $\alpha \geq 1$ and $0 < x < 1$, we see that for all $t > t_c$

$$\begin{aligned} M &= tn \sum_{j=2}^{J} (j-1) c_j (\rho^2 - \rho^j) + o(n) \\ &\leq tn \sum_{j=2}^{J} c_j (j-1) \rho^2 (1 - \rho^{j-2}) + o(n) \\ &\leq t\rho^2 n \sum_{j=2}^{J} c_j (j-1)(j-2)\theta + o(n). \end{aligned} \tag{39}$$

Thus it suffices to prove that

$$t\rho^2 \sum_{j=2}^{J} c_j (j-1)(j-2) < 1 \tag{40}$$

for $t < t_c + \varepsilon$. It is easy to check this when $t \leq t_c$, since then $\rho = 0$, hence this extends to $t < t_c + \varepsilon$ for some $\varepsilon > 0$ by continuity (as $\rho(t)$ is a continuous function). Alternatively, without using continuity of $\rho(t)$, (40) holds at $t = t_c$ by using $\rho \leq 1$ and the value of t_c. It is then obvious that the inequality extends by continuity to a neighbourhood of t_c.

In the case $t \to \infty$, this comes from the fact that there exists $c > 0$ such that for t large enough $\rho(t) \leq e^{-ct}$, so (40) also holds in this case. In turn, this follows from the fact that $\theta(t)$ is the survival probability of a Galton-Watson process where the offspring distribution is (18) and can thus be bounded below stochastically by a Poisson random variable with mean t. This finishes the proof of Theorem 3.

Acknowledgements

This paper owes much to the kind support of Oded Schramm during a visit made to the Theory Group at Microsoft Research in August 2008. I am very grateful for their invitation during that time. I would also like to thank an anonymous referee for constructive comments.

References

[1] O. Angel. Random infinite permutations and the cyclic time random walk. In *Discrete random walks (Paris, 2003)*, Discrete Math. Theor. Comput. Sci. Proc., AC, pages 9–16 (electronic). Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2003. MR2042369

[2] N. Berestycki. The hyperbolic geometry of random transpositions. *Ann. Probab.*, 34(2):429–467, 2006. MR2223947

[3] N. Berestycki and R. Durrett. A phase transition in the random transposition random walk. *Probab. Theory Related Fields*, 136(2):203–233, 2006. MR2240787

[4] N. Berestycki, O. Schramm, and O. Zeitouni. Mixing times of random k-cycles and coagulation-fragmentation chains. *Ann. Probab.*, to appear.

[5] B. Bollobás. *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. MR0809996

[6] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. III. The phase transition for the n-cube. *Combinatorica*, 26(4):395–410, 2006. MR2260845

[7] H. E. Buchanan and T. H. Hildebrandt. Note on the convergence of a sequence of functions of a certain type. *Ann. of Math. (2)*, 9(3):123–126, 1908. MR1502360

[8] P. Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988. MR0964069

[9] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981. MR0626813

[10] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2007. MR2271734

[11] M. Karoński and T. Łuczak. Random hypergraphs. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 283–293. János Bolyai Math. Soc., Budapest, 1996. MR1395864

[12] M. Karoński and T. Łuczak. The number of connected sparsely edged uniform hypergraphs. *Discrete Math.*, 171(1-3):153–167, 1997. MR1454447

[13] M. Karoński and T. Łuczak. The phase transition in a random hypergraph. *J. Comput. Appl. Math.*, 142(1):125–135, 2002. Probabilistic methods in combinatorics and combinatorial optimization. MR1910523

[14] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson. MR2466937

[15] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*, volume 6, pages 161–179, 1995. MR1370952

[16] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7(3):295–305, 1998. MR1664335

[17] S. Roussel. Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique. *Colloq. Math.*, 86(1):111–135, 2000. MR1799892

[18] L. Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004. MR2023654

[19] O. Schramm. Compositions of random transpositions. *Israel J. Math.*, 147:221–243, 2005. MR2166362

[20] B. Tóth. Improved lower bound on the thermodynamic pressure of the spin 1/2 Heisenberg ferromagnet. *Lett. Math. Phys.*, 28(1):75–84, 1993. MR1224836