Electronic Journal of Probability

Vol. 13 (2008), Paper no. 14, pages 374–395.

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

Limiting behavior for the distance of a random walk

Nathanaël Berestycki* Rick Durrett†

Abstract

In this paper we study some aspects of the behavior of random walks on large but finite graphs before they have reached their equilibrium distribution. This investigation is motivated by a result we proved recently for the random transposition random walk: the distance from the starting point of the walk has a phase transition from a linear regime to a sublinear regime at time $n/2$. Here, we study the examples of random 3-regular graphs, random adjacent transpositions, and riffle shuffles. In the case of a random 3-regular graph, there is a phase transition where the speed changes from $1/3$ to 0 at time $3\log_2 n$. A similar result is proved for riffle shuffles, where the speed changes from 1 to 0 at time $\log_2 n$. Both these changes occur when a distance equal to the average diameter of the graph is reached. However in the case of random adjacent transpositions, the behavior is more complex. We find that there is no phase transition, even though the distance has different scalings in three different regimes.

Key words: random walk, phase transition, adjacent transpositions, random regular graphs, riffle shuffles.

AMS 2000 Subject Classification: Primary 60C05, 60G50, 60J10.

Submitted to EJP on March 5, 2007, final version accepted March 5, 2008.

*Statistical Laboratory, University of Cambridge. CMS – Wilberforce Rd., Cambridge CB3 0WB, U.K. N.Berestycki@statslab.cam.ac.uk.

†Department of Mathematics, Malott Hall, Cornell University, Ithaca, NY 14853, U.S.A. rtd1@cornell.edu.
Both authors were partially supported by a joint NSF-NIGMS grant DMS-0201037.

1 Introduction

Random walks on large finite graphs have been the subject of intense research over the last 25 years. First used as mathematical models for problems related to card shuffling, they have also recently found some applications in the field of large-scale genome evolution (see respectively, e.g., [11], and [15; 16]). Since the pioneering work of Diaconis and Shahshahani [14], much of the work has traditionally focused on the *cutoff phenomenon*, which describes the way random walks on finite graphs converge (with respect to a given metric on probability distributions on the underlying graph) to their equilibrium distribution in a dramatically short period of time. See, for instance, the excellent monographs by Diaconis [11] and Saloff-Coste [23]. However, much less is known about how a random walk behaves before it has reached its equilibrium distribution. The goal of this paper is to precisely study aspects of this question for different examples. The result of our analysis is that in some cases there is an intermediary phase transition (in the sense of the phase transition of [6], recalled below), while in some other cases there is a more complex system of transitions between different regimes with no precise cutoff.

The starting point of our investigation is a result we recently proved for a random walk on the permutation group $\mathcal{S}_n$ on n markers. Let $(X_t^n, t \geq 0)$ be the continuous-time random transposition random walk. This means that X_t^n is a permutation and that at rate 1, we change the current permutation by performing a transposition of two randomly chosen elements. X^n may be thought of as a continuous-time simple random walk on the Cayley graph of $\mathcal{S}_n$ generated by the set of transpositions. Let D_t^n be the graphical distance of X_t^n from its starting point, i.e., D_t^n is the minimal number of transpositions necessary to change X_t^n into X_0^n. The main result of Berestycki and Durrett [6] is that D_t^n has a phase transition at time $n/2$ as $n \to \infty$. Writing $\to_p$ for convergence in probability, Theorem 3 in [6] may be restated as follows.

Theorem 0. *Let $t > 0$. As $n \to \infty$, $n^{-1} D_{tn}^n \to_p f(t)$ where $f(t)$ is defined by:*

$$f(t) = \begin{cases} t & \text{for } t \leq 1/2 \\ 1 - \displaystyle\sum_{k=1}^{\infty} \frac{1}{2t} \frac{k^{k-2}}{k!} (2te^{-2t})^k & \text{for } t > 1/2. \end{cases}$$

The function $f(t)$ is differentiable but the second derivative blows up as $t \downarrow 1/2$. For $t > 1/2$ we have $f(t) < t$. In words, the distance from the starting point of the random walk X^n asymptotically has a phase transition from linear to sublinear speed at time $t = n/2$.

Having seen this result, we ask in what situations does a random walk on a finite graph have a similar phase transition.

Organization of the paper. The rest of the paper is organized as follows. In the rest of this section we state our results which concern four different examples: random walk on a high-dimensional hypercube, random walk on a large random 3-regular graph, random adjacent transposition random walk, and the Gilbert-Shannon-Reeds riffle shuffle. We then discuss some related results and open problems in section 2. The proofs are in section 3, 4 and 5.

1.1 Random walk on the hypercube

We start with a trivial example. Let X^n_t be the random walk on the hypercube $\{0,1\}^n$ that jumps at rate 1, and when it jumps the value of one randomly chosen coordinate is changed. We assume that $X^n_0 = 0$. By considering a version of the chain that jumps at rate 2, and when it jumps the new coordinate takes on a value chosen at random from $\{0,1\}$ it is easy to see that, when $n = 1$,

$$\mathbb{P}_0(X^1_t = 1) = (1 - e^{-2t})/2.$$

Let D^n_t be the distance from X^n_t to X^n_0, i.e., the number of coordinates that disagree. Since the coordinates in the continuous time chain are independent, we easily obtain the following result.

Proposition 1. *As $n \to \infty$, $n^{-1}D^n_{nt} \to (1 - e^{-2t})/2$ in probability.*

Although this result is very simple, we note that Diaconis et al. [13] have shown that a discrete-time variant of this random walk undergoes an intermediary phase transition whose features present striking similarities with Theorem 0. After t moves, let $W(t)$ be the distance below which the probability of the random walk being at a given vertex of the hypercube is above average, i.e. $\mathbb{P}_t(x) \geq 1/2^n$ (this W is well-defined because the probability of being at a particular vertex depends only the distance of that vertex to the origin). Then a consequence of the results of [13] is that there exists an $\alpha > 0$ such that if $t = \lambda n$ for some $\lambda > 0$,

$$n^{-1}W(t) \to \rho \text{ where } \begin{cases} \rho = \lambda \text{ for } \lambda \leq \alpha \\ \rho < \lambda \text{ for } \lambda > \alpha. \end{cases}$$

Moreover a numerical approximation of α and a parametric relationship between ρ and λ are given (see (3.10) in [13]). A similar but simpler parametric relationship exists between t and $f(t)$ in Theorem 0. However the precise relation to our work remains elusive at this point.

1.2 Random walk on a random 3-regular graph

A *3-regular* graph is a graph where all vertices have degree equal to 3, and by random 3-regular graph we mean a graph on n vertices chosen uniformly at random from all 3-regular graphs on the n vertices.

The key to our proof is a construction of random 3-regular graphs due to Bollobás and de la Vega [9] (see also Bollobás [8]). This construction goes as follows. We suppose n is even. Expand each vertex i into 3 "mini-vertices" $3i$, $3i+1$ and $3i+2$, and consider a random matching $\sigma(j)$ of the $3n$ mini-vertices. A random 3-regular graph G_n is then obtained by collapsing back the n groups of 3 mini-vertices into n vertices while keeping the edges from the random matching. We may end up with self-loops or multiple edges, but with a probability that is positive asymptotically, we do not, so the reader who wants a neat graph can condition on the absence of self-loops and multi-edges. In this case the resulting graph is a uniform 3-regular random graph.

Departing from our choices in the previous example, we consider the discrete time random walk $\hat{X}^n_k$, $k \geq 0$, that jumps from j to $\lfloor \sigma(3j+i)/3 \rfloor$ where i is chosen at random from $\{0,1,2\}$. (We have used this definition since it works if there are self-loops or multiple edges.) Let $\hat{D}^n_t$ be the distance from the starting point at time t.

Theorem 1. *For fixed $t > 0$*
$$\frac{\hat{D}^n_{\lfloor t\log_2 n\rfloor}}{\log_2 n} \to_p \min\left(\frac{t}{3}, 1\right).$$

An intuitive description of a random 3-regular graph, as seen from vertex 1, can be given as follows. Grow the graph by successively adding vertices adjacent to the current set. Branching process estimates will show that as long as the number of vertices investigated is $O(n^{1-\epsilon})$, this portion of the graph looks very much like a regular tree in which each vertex has 2 edges going away from the root and 1 leading back towards the root. Thus, until the distance of $\hat{X}_n$ from $\hat{X}_0$ is $\geq (1-\epsilon)\log_2 n$, $\hat{D}^n_k$ evolves like a biased random walk on the nonnegative integers, with transition probabilities $p(x, x+1) = 2/3$ and $p(x, x-1) = 1/3$, and reflection at 0. After k moves we expect this walk to be at distance $k/3$. On the other hand, once the walk reaches a distance corresponding to the diameter of the graph, which is $\log_2 n$ by Bollobàs and de la Vega [9], or Theorem 2.13 in Worwald [24], it should remain at this level. Indeed, it cannot go any further, since this is the diameter. On the other hand the tree structure below makes it hard for it to come down back toward the root.

Open Problem 1. The techniques developed for the random walk on a 3-regular graph should be useful when dealing with random walk on the giant cluster of a Erdős-Rényi random graph with $p = c/n$ and $c > 1$, which locally has the geometry of a "Poisson mean c Galton-Watson tree". We conjecture that the random walk exhibits a phase transition like the one in Theorem 1 but with a different constants in place of 3 and 1 on the right-hand side. One technical problem is that the diameter is strictly larger than the average distance between points $\log n/(\log c)$, see Chung and Lu [10], so we don't have the easy upper bound.

1.3 Random adjacent transpositions

Let X^n_t be the continuous time random adjacent transposition random walk on n markers. An intuitive description of the process is as follows. We are thinking of $X_t(j)$ as the location of particle j, but the dynamics are easier to formulate in terms of $Y_t(i) := X_t^{-1}(i)$, which is the number of the particle at location i. At rate 1, we change the permutation by picking $1 \leq i \leq n-1$ at random and exchanging the values of $Y^n_t(i)$ and $Y^n_t(i+1)$. Without loss of generality we can suppose X^n_0 is the identity permutation I. In other words, we have n particles numbered 1 to n initially sorted in increasing order, and at rate 1 we exchange two adjacent particles. More formally, at rate 1, we change the value of the permutation from $X^n_{t^-}$ to X^n_t by setting
$$X^n_t = \tau X^n_{t^-}, \tag{1}$$
where τ is equal to the transposition $(i, i+1)$ with probability $1/(n-1)$ for $1 \leq i \leq n-1$.

Given a permutation σ, there is a convenient formula which gives the distance $d_{\text{adj}}(\sigma)$ from σ to I, i.e., the minimum number of adjacent transpositions needed to build σ.
$$d_{\text{adj}}(\sigma) = \text{Inv}(\sigma) := \#\{1 \leq i < j \leq n : \sigma(i) > \sigma(j)\}. \tag{2}$$

$\text{Inv}(\sigma)$ is called the number of inversions of σ. This formula is a quiet result. See, e.g., Diaconis and Graham [12], which includes earlier references to Kendall [21] and Knuth [22, section 5.1.1.].

If we view the set of permutations $\mathcal{S}_n$ of $\{1, \dots, n\}$ as a graph where there is an edge between σ and σ' if and only if σ' can be obtained from σ by performing an adjacent transposition (in the sense defined above), then X_t has the law of simple random walk on this graph and $d_{\text{adj}}(X_t)$ is the length of the shortest path between the current state of the walk, X_t, and its starting point, the identity.

Erikkson et al. [19] and later Eriksen [18], who were also motivated by questions in comparative genomics, considered the problem of evaluating the distance for the discrete time chain $\hat{X}_k^n$. Relying heavily on formula (2) they were able to carry out some explicit combinatorial analysis, to obtain various exact formulae for this expected distance, such as this one:

$$\mathbb{E}d_{\text{adj}}(\hat{X}_k^n) = \sum_{r=0}^{k} \frac{(-1)^r}{n^r} \left[\binom{k}{r+1} 2^r C_r + 4 d_r \binom{k}{r} \right] \tag{3}$$

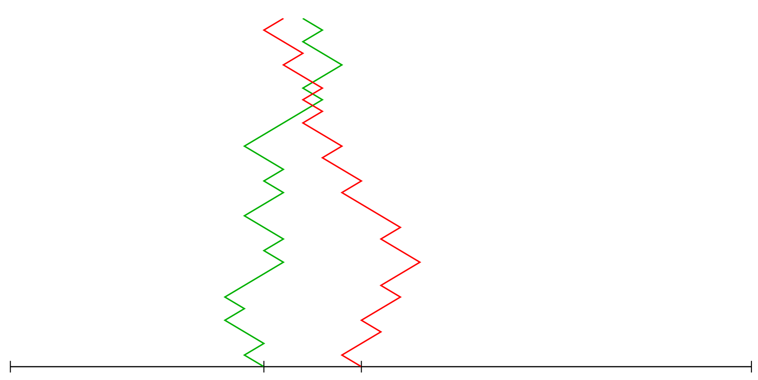
where C_r are the Catalan numbers and d_r is a less famous non-negative integer sequence, defined in [18].

While formula (3) is exact, it is far from obvious how to extract useful asymptotics from it. We will take a probabilistic approach based on the formula

$$D_t^n = d_{\text{adj}}(X_t^n) = \sum_{i<j} \mathbf{1}_{\{X_t^n(i) > X_t^n(j)\}}. \tag{4}$$

If $1 \le i \le n$ is fixed, the trajectory $X_t^n(i)$ of the i^{th} particle is a continuous time simple random walk on $\{1, \dots, n\}$ starting at i with jumps at rate $2/(n-1)$ and reflecting boundaries at 1 and n that cause the particle to stay put with probability 1/2.

Two such trajectories, say those of particles i and j with $i < j$, move by the nearest neighbor stirring process on $\{1, \dots, n\}$ (which for indistinguishable particles produces the *simple exclusion process*). When the particles are not adjacent, they perform independent simple random walks. When they are adjacent, the only things that can happen are an exchange of the two particles, or one of them moves away from the other (see figure below for an illustration). As the reader can probably guess, and Durrett and Neuhauser [17] have proved on **Z**, when n is large the random walks behave as if they are independent.



On different time scales we see three different behaviors. At small times, the behavior is given by a messy formula that defines a smooth function. For an integer $x \ge 0$, let T^x denote the hitting time of the level x by a rate 4 continuous time random walk on $\mathbf{Z}$ starting at 0. Let $\{Y(t), t \geq 0\}$ and $\{Y'(t), t \geq 0\}$ be moved by random stirring on $\mathbf{Z}$, with $Y(0) = 0$ and $Y'(0) = 1$. (That is, each edge of $\mathbf{Z}$ is endowed with a Poisson process with intensity 1. When the clock associated with an edge rings and that at least one end of the edge is occupied by a particle, we put that particle at the other end of the edge, while if both ends are occupied we simply exchange the particles – in particular, the particles never occupy the same site.) Let $p(u)$ be the probability that at time u the relative order of the two particles has changed, i.e., $p(u) = \mathbb{P}[Y(u) > Y'(u)]$, and note that this is the same as requiring the particles to have been exchanged an odd number of times. For all $t > 0$, let

$$f(t) := \sum_{x=0}^{\infty} \int_0^t \mathbb{P}[T^x \in ds] p(t-s) \tag{5}$$

and recall the formula for the distance D_n^t given in (4).

Theorem 2. *Let $t > 0$. Then $n^{-1} D_{nt}^n \to_p f(t)$ as $n \to \infty$ where f is the function defined by (5). $f(t)$ is infinitely differentiable, and moreover it has the asymptotic behavior*

$$\lim_{t \to \infty} \frac{f(t)}{\sqrt{t}} = \frac{1}{2} \mathbb{E} \left(\max_{0 \leq s \leq 1} B_{4s} \right) = \sqrt{\frac{2}{\pi}}$$

where B_t is a standard Brownian motion.

To check the second equality in the limit, recall that by the reflection principle, for all $x \geq 0$,

$$\mathbb{P} \left(\max_{0 \leq s \leq 1} B_{4s} > x \right) = 2\mathbb{P}(B_4 > x)$$

so integrating gives

$$\begin{aligned} \frac{1}{2} \mathbb{E} \left(\max_{0 \leq s \leq 1} B_{4s} \right) &= \int_0^\infty \mathbb{P}(B_4 > x) dx \\ &= \mathbb{E} B_4^+ = 2\mathbb{E} B_1^+ \\ &= \frac{2}{\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}}. \end{aligned}$$

An anonymous referee points out that the expression (5) which defines the function f describing the limiting behavior of the distance at small times may be simplified using well-studied special functions. From our point of view, the interest of that formula is that it is not the same as in other regimes, as we will see in Theorems 3 and 4.

The next result looks at the distance of the random walk at times of order n^3, i.e., when each particle has moved of order n^2 times, and hence has a significant probability of hitting a boundary. Let $p_t(u, v)$ denotes the transition function of $\bar{B}$, a one-dimensional Brownian motion run at speed 2 reflecting at 0 and 1.

Theorem 3. *Let $t > 0$.*

$$\frac{1}{n^2} D_{n^3 t}^n \to_p \int_0^1 du \int_u^1 dv \int_0^1 p_t(u, x) dx \int_0^x p_t(v, y) dy = \mathbb{P}[\bar{B}_1(t) > \bar{B}_2(t)]$$

where $\bar{B}_1$ and $\bar{B}_2$ are independent copies of $\bar{B}$ started uniformly on $0 \leq \bar{B}_1(0) < \bar{B}_2(0) \leq 1$ evolving independently.

In between the two extremes we have a simple behavior, anticipated by the limit in Theorem 2.

Theorem 4. *Let $s = s(n)$ with $s \to \infty$ and $s/n^2 \to 0$. Then*

$$\frac{1}{n\sqrt{s}} D^n_{ns} \to_p \sqrt{\frac{2}{\pi}}.$$

Recently, Angel et al. [2] (see also [3]) have also used the simple exclusion process to analyze a process on the Cayley graph of the symmetric group generated by adjacent transpositions, but this time in the context of sorting networks.

1.4 Riffle shuffles

The Gilbert-Shannon-Reeds shuffle, or riffle shuffle, is a mathematical model for the way card players shuffle a deck of cards. It can be viewed as a nonreversible random walk on the permutation group. We identify the order of the deck of n cards with an element of the permutation group $\mathcal{S}_n$ by declaring that $\pi(i)$ is the label of the card in position i of the deck (and hence $\pi^{-1}(i)$ is the position in the deck of the card whose label is i). The most intuitive way to describe this shuffle is to say that at each time step, σ_m is obtained from σ_{m-1} by first cutting the deck into two packets, where the position of the cut has a Binomial $(n, 1/2)$ distribution. The two packets are then riffled together in the following way: if the two packets have respective sizes a and b, drop the next card from the first packet with probability $a/(a+b)$ and from the second with probability $b/(a+b)$. This goes on until cards from both packets have been dropped, and the resulting deck forms σ_n. This shuffle has been extensively studied. See, e.g., Bayer and Diaconis [4] and Aldous [1], for results about the mixing time of this random walk, which is $(3/2)\log_2 n$ for the total variation distance.

Bayer and Diaconis [4] were able to prove the following remarkable exact formula for the probability distribution of the random walk after a given number of steps. Let $\pi \in \mathcal{S}_n$ be a permutation, viewed as an arrangement of cards, and let $r = R(\pi)$ be the number of *rising sequences* of π. A rising sequence of π is a maximal subset of cards of this arrangement consisting of successive face values displayed in order. For instance, if $n = 13$ and the deck consists of the following arrangement:

$$1 \quad 7 \quad 2 \quad 8 \quad 9 \quad 3 \quad 10 \quad 4 \quad 5 \quad 11 \quad 6 \quad 12 \quad 13$$

then there are two rising sequences:

$$\begin{array}{ccccccccccccc} 1 & & 2 & & & 3 & & 4 & 5 & & 6 & & \\ & 7 & & 8 & 9 & & 10 & & & 11 & & 12 & 13 \end{array}$$

Theorem 1 of Bayer and Diaconis [4] states that after m shuffles,

$$\mathbb{P}(\sigma_m = \pi) = \frac{1}{2^{mn}} \binom{2^m + n - R(\pi)}{n}. \tag{6}$$

Note that a consequence of the Bayer-Diaconis formula (6) is that the chance of π is positive if and only if $2^m - R(\pi) \geq 0$ or $m \geq \lceil \log_2 R(\pi) \rceil$. Therefore, the distance of a permutation π

to the identity (where here the distance means the minimal number of riffle shuffles needed to build σ_m), is given by the explicit formula

$$d_{RS}(\pi) = \lceil \log_2 R(\pi) \rceil. \tag{7}$$

Based on these ideas, it is easy to prove the following result. Let $D(m)$ be the distance to the identity of the random walk after m shuffles.

Theorem 5. *Let $t > 0$.*

$$\frac{D(\lfloor t \log_2 n \rfloor)}{\log_2 n} \to_p \min(t, 1).$$

Remark. It is interesting to note that this random walk reaches the average distance of the graph abruptly at time $\log_2 n$, while it reaches uniformity at time $(3/2)\log_2 n$ (see, e.g., Aldous [1] and Bayer-Diaconis [4]). This contrasts with the conjectured situation for random 3-regular graphs.

When $t < 1$, this result can be deduced from Fulman's recent paper [20], although our methods are much more elementary. However, Fulman obtains some much more precise results, which describe what happens near the transition point when $t = 1$, and convey some information about the fluctuations when $t \leq 1$. This result and related open problems will be discussed in the next section.

2 Related results and open problems

2.1 A theorem of J. Fulman.

Fix an $\alpha > 0$ and consider the state of the Gilbert-Shannon-Reeds riffle shuffle after $m = \lfloor \log_2(\alpha n) \rfloor$ shuffles. Part of Fulman's result may be reformulated in the following way. Let $R(\sigma_m)$ be the number of rising sequences of σ_m.

Theorem (Fulman [20]). Suppose $\alpha > 1/(2\pi)$.

$$\frac{1}{n}\mathbb{E}(R(\sigma_m)) \to \alpha - \frac{1}{e^{1/\alpha} - 1}. \tag{8}$$

To see why this indeed the same as his Proposition 4.5, simply note that his $R_{k,n}$ coincides with the law of σ_m^{-1} for $m = 2^k$. Since

$$R(\sigma) = \mathrm{Des}(\sigma^{-1}) + 1$$

where $\mathrm{Des}(\sigma)$ is the number of descents of σ, (8) follows immediately. Using (7), Fulman's result (8) has an immediate interpretation for the distance to the identity after m shuffles. Using in particular the Stein method, he also finds that the variance of $R(\sigma_m)$ is approximately $C_\alpha n$ for some explicit but complicated $C_\alpha > 0$, and that $R(\sigma_m)$ is asymptotically normally distributed with this mean and variance. For smaller values of m, in particular for $m = t \log_2 n$ and $t < 1$ (i.e., before the phase transition point of Theorem 5), he finds that the number of rising sequences is approximately $2^m - Z$, where Z is a Poisson random variable with mean $\lambda := 2^m/n$. This parallels in a striking way the Poisson and normal deviations observed by us in [6].

Open Problem 2. In (8), what is the behavior of $\mathbb{E}(R(\sigma_m))$ for values of α smaller than $1/(2\pi)$?

It is not clear at this point whether (8) also holds for $\alpha < 1/(2\pi)$, although it is tempting to let $\alpha \to 0$ and get that for small values of α the walk is "almost" linear (the fraction term with the exponential is much smaller than the other term).

2.2 Geometric interpretation of phase transitions.

The techniques used to prove the results in this paper, and other results such as Theorem 0 or Fulman's result, rely in general on *ad hoc* formulae for the distance of a point on the graph to a given starting point. For the moment there does not seem to be any general technique to approach these problems. However we note that a geometric approach has been proposed by [5]. The main result of [5] relates the existence of such phase transitions to some qualitative changes in the hyperbolic properties of the underlying graph of the random walk: the space *looks hyperbolic* to the random walk until a certain critical radius. Beyond this critical radius, space starts wrapping around itself and ceases to look hyperbolic. (While [5] is only concerned with the random transposition random walk, it is easy to rephrase this result for more general walks with phase transitions). This is for instance consistent with Theorem 1 for random 3-regular graphs, where we indeed expect hyperbolicity to break down at distance $\log_2 n$. However this hyperbolicity criterion seems hard to use in practice. Can a general method be developed?

2.3 Cyclic adjacent transpositions.

Following a question raised to us by W. Ewens, we now turn our attention to a process closely related to the random adjacent transpositions analyzed in the present paper. Suppose that we modify the dynamics of X^n_t defined by (1) by also allowing the transposition $\tau = (1\ n)$ in addition to $\tau = (1\ 2), (2\ 3), \ldots, (n-1\ n)$. That is, we only exchange adjacent particles, but now adjacent is interpreted in a cyclical way. Informally, we have n particles equally spaced on the unit circle, and exchange adjacent particles at rate 1. For a given configuration of particles σ, let $d(\sigma)$ be the number of moves it takes to put the particles back in order using only exchanges of adjacent particles, and let $D^n_t = d(X^n_t)$. We conjecture that in this setting, Theorems 2 and 4 still hold. For large times we conjecture the following result.

Conjecture 3. *Let $t > 0$.*

$$\frac{1}{n^2} D^n_{n^3 t} \to_p \frac{1}{4\pi} \mathbb{E}(|R_{4t}|) \tag{9}$$

where $|R_t|$ denotes the Lebesgue measure of the range of a Brownian motion with unit speed on the circle of radius 1.

We view this conjecture as an analogue of Theorem 3, since the scalings are the same in both statements. Moreover, this conjecture comes from a heuristics similar to the one leading to Theorem 3. However, the difficulty here is that there is no exact formula analogous to (2) for the distance of a permutation using cyclic transpositions.

3 Random walk on a random 3-regular graph

Let G_n be a random 3-regular graph as constructed in the introduction using the approach of Bollobás and de la Vega [9] (that is, G_n is a random multigraph which contains no multiple edges or self-loops with probability bounded away from zero). Let X_k be the discrete time random walk on G_n, where for simplicity we drop both the superscript n and the hat to indicate discrete time. We assume that $X_0 = 1$ and write D_k^n for the graph distance from X_k to X_0. Our goal is to prove Theorem 1, that is, for fixed $t > 0$

$$\frac{D^n_{\lfloor t \log_2 n \rfloor}}{\log_2 n} \to_p \min\left(\frac{t}{3}, 1\right). \tag{10}$$

3.1 Proof for the subcritical regime

Let v be a vertex at distance ℓ from the root. We say that v is a "good" vertex if it has two edges leading away from the root (at distance $\ell + 1$) and one leading back to distance $\ell - 1$. Otherwise we say that v is a "bad" vertex. Let $B(\ell)$ be the set of all bad vertices at distance ℓ.

Lemma 1. *Let $2 \leq v \leq n$ be a vertex distinct from the root. Given that v is at distance ℓ from the root, $\mathbb{P}(v \in B(\ell)) \leq 3 \cdot 2^\ell / n$.*

Proof. First consider the event that v has an edge leading to some other vertex at distance ℓ. Since it is at distance ℓ, it must have at least one edge leading backwards, so there are only two other edges left. In particular (taking into account that the root has three edges going out), there are at most $i := 3 \cdot 2^{\ell-1}$ vertices at distance ℓ. In G_n those i vertices at distance ℓ correspond to $2i$ unpaired mini-vertices, so the probability of a connection sideways to another vertex at distance ℓ is smaller than $2i/3n$.

When v has two edges leading forward, the probability that one of its children is connected to another vertex from level ℓ is also smaller than $2i/3n$ since there are at most $2i$ edges leading to level $\ell + 1$. Since v has at most 2 children, this gives a probability of at most $4i/3n$. Combining this with the estimate above gives $2i/3n + 4i/3n = 2i/n$. Since $i = 3 \cdot 2^{\ell-1}$ this terminates the proof of Lemma 1. $\square$

A simple heuristic now allows us to understand that with high probability the random walk will not encounter any bad vertex as long as we are in the subcritical regime. Before we encounter a bad vertex, the distance to the origin is a (2/3,1/3) biased random walk and hence the walk spends an average of 3 steps at any level. Hence, the expected number of bad vertices encountered until distance $(1 - \varepsilon) \log_2 n$ is smaller than

$$\sum_{\ell=1}^{(1-\varepsilon)\log_2 n} 9 \frac{2^{\ell-1}}{n} = O(n^{-\varepsilon}) \to 0.$$

To prove this rigorously, let A_k denote the event that by time k the random walk has never stepped on a bad vertex.

Lemma 2. *As $n \to \infty$, $\mathbb{P}(A_{3(1-\varepsilon)\log_2 n}) \to 1$.*

On this event, for each $1 \le j \le 3(1-\varepsilon)\log_2 n$, X_j has probability 2/3 to move away from the root and 1/3 to move back towards the root, and the first part of Theorem 1 follows easily.

Proof. By Lemma 1 the probability that some vertex within distance L of 1 is bad is

$$\le \sum_{\ell=1}^{L} 3 \cdot 2^{\ell-1} \frac{3 \cdot 2^{\ell-1}}{n} \le \frac{9}{2n} \cdot \frac{2^{2L}}{1-1/4} \to 0 \tag{11}$$

if $L = (1/3)\log_2 n$.

Since for each vertex there are at most two edges leading out and one leading back, the distance from the starting point is bounded above by a (2/3,1/3) biased random walk. Standard large deviations arguments imply that there are constants C and α depending on $\rho > 1/3$ so that

$$\mathbb{P}(d(X_k) > \rho k) \le Ce^{-\alpha k}. \tag{12}$$

Let $A'_k = A_k \cap \{d(X_k) \le \rho k\}$. It follows from Lemma 1 that

$$\mathbb{P}(A'_{k+1}) \ge \mathbb{P}(A'_k)\left(1 - 3\frac{2^{\rho k-1}}{n} - Ce^{-\alpha k}\right) \ge \prod_{j=L}^{k}\left(1 - 3\frac{2^{\rho j-1}}{n} - Ce^{-\alpha k}\right)(1+o(1))$$

where the term $o(1)$ accounts for (11). Taking the logarithm, we have for large n

$$\begin{aligned}\log \mathbb{P}(A'_{k+1}) &\ge \sum_{j=L}^{k} \log\left(1 - 3\frac{2^{j-1}}{n} - Ce^{-\alpha j}\right) + o(1)\\ &\ge -6\sum_{j=1}^{k}\frac{2^j}{n} - Ce^{-\alpha L} + o(1) \ge -\frac{6}{1-2^{-\rho}} \cdot \frac{2^{k\rho}}{n} + o(1).\end{aligned}$$

We want to take $k = 3(1-\varepsilon)\log_2 n$. By choosing ρ close enough to 1/3 so that $3\rho(1-\varepsilon) < 1$, we have $2^{k\rho}/n = n^{-\alpha}$ with $\alpha > 0$ which proves that $\log \mathbb{P}(A'_k) \to 0$ as $n \to \infty$, or equivalently, $\mathbb{P}(A'_k) \to 1$ as $n \to \infty$. However, since $A'_k \subset A_k$, we deduce immediately that $\mathbb{P}(A_k) \to 1$ as $n \to \infty$. This finishes the proof of Lemma 2. □

3.2 Proof for the supercritical regime

Here we wish to prove that if $k = t\log_2 n$, with $t > 3(1-\varepsilon)$, then $d(X_k) \sim \log_2 n$. As already noted, this is the diameter of G_n so all we have to prove is that once it reaches this distance it stays there. To do this we let

$$L(a,b) := \{2 \le v \le n : d(v) \in [a\log_2 n, b\log_2 n]\}$$

and consider $L(1-\varepsilon, 1-\delta)$.

Intuitively, this strip consists of about $n^{1-\varepsilon}$ trees, each with at most $n^{\varepsilon-\delta}$ vertices. However, there are sideways connections between these trees so we have to be careful in making definitions. Let $v_1, \ldots, v_m$ be the m vertices at level $(1-\varepsilon)\log_2 n$. For $j = 1, \ldots, m$ if $v \in L(1-\varepsilon, 1-\delta)$, we say that $v \in T_j$ if v_j is a closest vertex to v among $v_1, \ldots, v_m$. Define an equivalence relation on

$\{1, \ldots, m\}$ by saying that $j \sim j'$ if and only if $T_j \cap T_{j'} \neq \emptyset$. We will call the equivalence classes for this relation *cluster of trees.*

To estimate the number of sideways connections (or rather, the cardinality of an equivalence class), we use the following lemma.

Lemma 3. *The cardinality of the equivalence class of j (i.e., the number of subtrees that T_j is connected to) is dominated by the total population of a branching process with offspring distribution Binomial($n^{\varepsilon-\delta}, 3n^{-\delta}$).*

Proof. Each tree to which we connect requires a bad connection (i.e., one of the two possible errors in Lemma 1). Suppose we generate the connections sequentially. The upper bound in Lemma 1 holds regardless of what happened earlier in the process, so we get an upper-bound by declaring each vertex at level ℓ bad independently with probability $2i/n$ with $i = 3 \cdot 2^{\ell-1}$, so this probability is at most $n^{-\delta}$. Since there are at most $n^{\varepsilon-\delta}$ vertices in a given subtree, the lemma follows immediately. $\square$

Lemma 4. *If $\delta > \varepsilon/2$ then there exists some $K = K(\varepsilon, \delta) > 0$ such that,*

$$\mathbb{P}(\textit{there is a cluster of trees } T_j \textit{ with more than } K \textit{ bad vertices}) \to 0.$$

Proof. The worst case occurs when each bad connection in a tree leads to a new one. Let

$$X \stackrel{d}{=} \mathrm{Bin}(n^{\varepsilon-\delta}, 3n^{-\delta})$$

be the offspring distribution of the branching process of the previous Lemma. In particular

$$\mathbb{E}(X) = O(n^{\varepsilon-2\delta}) \to 0.$$

Let $c = 3n^{\varepsilon-2\delta}$, and let $N = n^{\varepsilon-\delta}$ be the maximum number of vertices in T_j, so $X \stackrel{d}{=}$ Binomial$(N, c/N)$.

Lemma 4 follows from a simple evaluation of the tail of the total progeny Z of a branching process with offspring distributed as X. To do this, we let $\phi_N(\theta) = \mathbb{E}[\exp(\theta(X-1))]$ be the moment generating function of $X - 1$. Then

$$\begin{aligned}\phi_N(\theta) &= e^{-\theta} \sum_{k=0}^{N} \binom{N}{k} \left(\frac{c}{N}\right)^k \left(1 - \frac{c}{N}\right)^{N-k} e^{\theta(k-1)} \\ &= e^{-\theta} \left(1 - \frac{c}{N} + \frac{c}{N} e^{\theta}\right)^N .\end{aligned}$$

Let $(S_k, k \geq 0)$ be a random walk that takes steps with this distribution and $S_0 = 1$. Then $\tau = \inf\{k : S_k = 0\}$ has the same distribution as Z. Let

$$R_k = \exp(\theta S_k)/\phi_N(\theta)^k.$$

Then $(R_k, k \geq 0)$ is a nonnegative martingale. Stopping at time τ we have

$$e^{\theta} \geq \mathbb{E}(\phi_N(\theta)^{-\tau}).$$

If $\phi_N(\theta) < 1$ it follows that

$$\mathbb{P}(\tau \geq y)\phi_N(\theta)^{-y} \leq \mathbb{E}[\phi_N(\theta)^{-\tau}] \leq e^{\theta}.$$

Using $\phi_N(\theta) \leq e^{-\theta} \exp(c(e^{\theta} - 1))$ now we have

$$\mathbb{P}(\tau \geq y) \leq e^{\theta} \left(e^{-\theta} \exp(c(e^{\theta} - 1))\right)^{y}.$$

To optimize the bound we want to minimize $c(e^{\theta} - 1) - \theta$. Differentiating this means that we want $ce^{\theta} - 1 = 0$ or $\theta = -\log(c)$. Plugging this and recalling that τ and Z have the same distribution we have

$$\mathbb{P}(Z \geq y) \leq \frac{1}{c} \exp(-(c - 1 - \ln c)y).$$

Substituting $c = 3n^{-\alpha}$ with $\alpha = 2\delta - \varepsilon$, we find that

$$\mathbb{P}(Z \geq y) \leq 3n^{\alpha} \exp(y(1 - \alpha \log(n))).$$

Since there are $m \leq n^{1-\varepsilon}$ trees to start with, the probability that one of them has more than y trees in its cluster is smaller than

$$3n^{1-\varepsilon} n^{\alpha} \exp(y(1 - \alpha \log n))$$

so if

$$y > \frac{\alpha + 1 - \varepsilon}{\alpha} := K(\delta, \varepsilon)$$

then the probability than one cluster contains more than y trees tends to 0. This implies that with probability 1 asymptotically, no cluster of trees has more than K bad vertices, since the branching process upper-bound is obtained by counting every bad vertex as a sideways connection. □

With Lemma 4 established the rest is routine. Choose K as in Lemma 4. On the event that no cluster of trees contains more than K bad vertices, there is a region of this cluster of length at least $\geq a \log_2 n$ where $a = (\epsilon - \delta)/(K + 1)$ with no bad vertices.

When in this region, X_n is a biased (2/3,1/3) random walk. It is a well-known fact that if $(S_n, n \geq 0)$ is a (p, q)-biased random walk with $p > 1/2$, then $\mathbb{P}_0(T_{-1} < \infty) = (1 - p)/p$. Hence, the probability that, starting from the top of this region, the random walk will reach the bottom is equal to $(1/2)^{a \log_2 n} = n^{-a}$. Therefore, the probability of ever downcrossing this strip in $n^{a/2}$ attempts tends to 0. Now, if $t \geq 3$ is fixed, by time $\lfloor t \log_2 n \rfloor$, there are at most $\lfloor t \log_2 n \rfloor$ attempts to cross it downward, and so we conclude that with high probability, the random walk cannot cross the strip $L(1-\varepsilon, 1-\delta)$ downward. The critical and supercritical part of the theorem follow.

4 Random adjacent transpositions

Let X_t, which we write for now on without the superscript n, to be the continuous time walk on permutations of $\{0, 1, 2, \ldots, n\}$ in which at rate 1 we pick a random $0 \leq i \leq n - 1$ and exchange the values of $X_t(i)$ and $X_t(i + 1)$. As indicated in (2) the distance from a permutation σ to the identity is $d_{\text{adj}}(\sigma) = \#\{0 \leq i < j \leq n : \ \sigma(i) > \sigma(j)\}$, the number of inversions of σ.

4.1 Small times

The reflecting boundaries at 0 and n are annoying complications, so the first thing we will do is get rid of them. To do this and to prepare for the variance estimate we will show that if $i < j$ are far apart then the probability $X_t(i) > X_t(j)$ is small enough to be ignored. Let $\mathbb{P}^{[a,b]}$ be the probabilities for the stirring process with reflection at a and b, with no superscript meaning no reflection.

Lemma 5. $\mathbb{P}^{[0,n]}(X_{nt}(i) > X_{nt}(j)) \le 8\mathbb{P}(X_{nt}(0) > (j-i)/2)$.

Proof. A simple coupling shows

$$\begin{aligned}\mathbb{P}^{[0,n]}(X_{nt}(i) > X_{nt}(j)) &\le \mathbb{P}^{[0,j-i]}(X_s(0) > X_s(j-i) \text{ for some } s \le nt) \\ &\le 2\mathbb{P}^{[0,\infty)}\left(\max_{0\le s\le nt} X_s(0) > (j-i)/2\right).\end{aligned}$$

Using symmetry and then the reflection principle, the last quantity is

$$\le 4\mathbb{P}\left(\max_{0\le s\le nt} X_s(0) > (j-i)/2\right) \le 8\mathbb{P}(X_{nt}(0) > (j-i)/2)$$

which completes the proof. ☐

Since the random walk on time scale nt moves at rate 2,

$$\mathbb{E}\exp(\theta X_{nt}(0)) = \sum_{k=0}^{\infty} e^{-2t}\frac{(2t)^k}{k!}\left(\frac{e^{\theta}+e^{-\theta}}{2}\right)^k = \exp(-2t + t(e^{\theta}+e^{-\theta})).$$

Using Chebyshev's inequality, if $\theta > 0$

$$\mathbb{P}(X_{nt}(0) > x) \le \exp(-\theta x + t[e^{\theta}+e^{-\theta}-2]). \tag{13}$$

Taking $\theta = 1$

$$\mathbb{P}(X_{nt}(0) > x) \le C_t e^{-x} \quad \text{where} \quad C_t = \exp((e+e^{-1}-2)t).$$

When $x = 3\log n$ the right-hand side is $C_t n^{-3}$, so using Lemma 5, for fixed t it suffices to consider "close pairs" with $0 < j-i < 6\log n$. The number of close pairs with $i \le 3\log n$ or $j > n - 3\log n$ is $\le 36\log^2 n$, so we can ignore these as well, and the large deviations result implies that it is enough to consider random stirring on $\mathbf{Z}$.

We are now ready to prove the first conclusion in Theorem 2: if $t > 0$ then as $n \to \infty$

$$\frac{1}{n}D^n_{nt} \to f(t) = \sum_{x=1}^{\infty}\int_0^t \mathbb{P}[T^x \in ds]p(t-s). \tag{14}$$

Proof of (14). It is clear from the Markov property that if $X_t(i)$ and $X_t(j)$ are moved by stirring on $\mathbf{Z}$ then

$$\mathbb{P}(X_t(i) > X_t(j)) = \int_0^t \mathbb{P}[T^{j-i-1} \in ds]p(t-s).$$

With the large deviations bound in (13) giving us domination we can pass to the limit to conclude

$$\frac{1}{n} \sum_{0 \le i < j \le n} \mathbb{P}(X_t(i) > X_t(j)) \to \sum_{x=0}^{\infty} \int_0^t \mathbb{P}[T^x \in ds] p(t-s).$$

To prove convergence in probability let

$$\xi_{i,j} = 1_{(X_t(i) > X_t(j))} - \mathbb{P}(X_t(i) > X_t(j)).$$

By remarks above it suffices to consider the sum over $1 \le i < j \le n$ with $0 < j - i < 6 \log n$, $i \ge 3 \log n$ and $j \le n - 3 \log n$ which we denote by Σ^*, and if $i' > j + 6 \log n$ then

$$\mathbb{E} \xi_{i,j} \xi_{i'j'} \le 4 C_t n^{-3}$$

since the random variables have $|\xi| \le 1$ and will be independent unless some random walk moves by more than $3 \log n$ in the wrong direction. From this it follows that

$$\mathbb{E} \left(\Sigma^* \xi_{i,j} \right)^2 \le n \cdot (6 \log n)^3 + 4 C_t n^{-3} (n \cdot 6 \log n)^2$$

and the result follows from Chebyshev's inequality. □

The remaining detail is to show that f is smooth and that as $t \to \infty$,

$$\lim_{t \to \infty} f(t)/\sqrt{t} = \frac{1}{2} \mathbb{E} \left(\max_{0 \le s \le 1} B_{4s} \right) \tag{15}$$

where $B.$ is a standard Brownian Motion. The fact that f is infinitely differentiable follows easily from repeated use of Lebesgue's theorem and the fact that both $p(u)$ and $d\mathbb{P}(T^x \in du)/du$ are infinitely differentiable smooth functions. This is itself easily checked: for instance, if q_j is the probability that a simple random walk in discrete time started at 0 hits x in j steps, then $T^x = \sum_{j=x}^{\infty} q_j \text{Gamma}(j, 4)$, so T^x has a smooth density. A similar argument also applies for the function $p(u)$.

Proof of (15). The result follows easily from two simple lemmas.

Lemma 6. *$p(t) \to 1/2$ as $t \to \infty$.*

Proof. Each time there is jump when the particles Y and Y' are adjacent, they have a probability $1/3$ of being exchanged the next step. So, conditionally on the number of such jumps N, the number of actual swaps between Y and Y' is Binomial$(N, 1/3)$. Now, $Y > Y'$ if and only if the number of times they are swapped is odd. Hence the lemma follows from the two observations : (i) As $t \to \infty$, the number of jumps while they are adjacent to each other $\to \infty$, and (ii) as $N \to \infty$, $\mathbb{P}[\text{Binomial}(N, p) \text{ is odd}] \to 1/2$ for any given $0 < p < 1$. For (i), observe that the discrete-time chain derived from $\{|Y_t - Y'_t| - 1, t \ge 0\}$ is a reflecting random walk on $\{0, 1, \dots\}$, and therefore visits 0 infinitely many times. (ii) is an easy fact for Bernoulli random variables. □

Lemma 7.

$$\frac{1}{\sqrt{t}} \sum_{x=1}^{\infty} \mathbb{P}[T^x \in (t - \log t, t)] \to 0.$$

Proof. The random walk can only hit a new point when it jumps so

$$\begin{aligned} t^{-1/2}\mathbb{E}\left(\sum_{x=1}^{\infty}\mathbf{1}_{\{T^x\in(t-\log t,t)\}}\right) &\le t^{-1/2}\mathbb{E}(\#\text{jumps of the random walk in}(t-\log t,t)) \\ &\le t^{-1/2}\cdot(4\log t)\to 0 \end{aligned}$$

since jumps occur at rate 4. □

It is now straightforward to complete the proof. Let $\varepsilon > 0$. Fix T large enough so that $|p(t) - 1/2| \le \varepsilon$ as soon as $t \ge T$. Then by Lemma 7, for $t \ge T' := e^T$, letting W_t be a simple random walk on $\mathbf{Z}$ in continuous time jumping at rate 1,

$$\begin{aligned} t^{-1/2}f(t) &= t^{-1/2}\sum_{x=1}^{\infty}\int_0^{t-\log t}\mathbb{P}[T^x\in ds]p(t-s)+o(1) \\ &\le \left(\frac{1}{2}+\varepsilon\right)t^{-1/2}\sum_{x=1}^{\infty}\mathbb{P}[T^x<t-\log t]+o(1) \\ &\le \left(\frac{1}{2}+\varepsilon\right)t^{-1/2}\sum_{x=1}^{\infty}\mathbb{P}\left(\max_{s\le t-\log t}W_{4s}>x\right)+o(1) \\ &\to \left(\frac{1}{2}+\varepsilon\right)\mathbb{E}\ \max_{s\le 1}B_{4s} \end{aligned}$$

by Donsker's theorem. The other direction $\liminf_{t\to\infty} t^{-1/2}f(t) \ge (1/2-\varepsilon)\mathbb{E}\max_{s\le 1}B_{4s}$ can be proved in the same way. □

4.2 Large times

Our next goal is to prove that if $t > 0$ then

$$\frac{1}{n^2}D^n_{n^3t}\to\int_0^1 du\int_u^1 dv\int_0^1 p_t(u,x)dx\int_0^x p_t(v,y)dy=\mathbb{P}[\bar{B}_1(t)>\bar{B}_2(t)] \qquad (16)$$

in probability and where $\bar{B}_1$ and $\bar{B}_2$ are two reflecting Brownian motions run at speed 2 started uniformly on $0 \le \bar{B}_1(0) < \bar{B}_2(0) \le 1$ and evolving independently.

Proof. We first show that the expected value converges. The first step is to observe that the rescaled random walks $X_{n^3t}(i)/n$, $t \ge 0$ converge to reflecting Brownian Motion on $[0,1]$. Indeed, Durrett and Neuhauser [17, (2.8)] showed that for fixed $i < j$, the rescaled pair of random walks converge to two independent Brownian motions. They did this on $\mathbf{Z}$ but the proof extends in a straightforward way to the current setting. Their proof shows that if $i/n \to x$ and $j/n \to y$ we have

$$\mathbb{P}[X_{n^3t}(i)>X_{n^3t}(j)]\to\mathbb{P}_{x,y}[\bar{B}_1(t)>\bar{B}_2(t)].$$

This implies that the convergence occurs uniformly on the compact set so

$$\frac{1}{n^2}\mathbb{E}D^n_{n^3t}=\frac{1}{n^2}\sum_{i<j}\mathbb{P}[X_{n^3t}(i)>X_{n^3t}(j)]\to\mathbb{P}[\bar{B}_1(t)>\bar{B}_2(t)].$$

To get the convergence in probability, we use second moment estimates. Let $A_{i,j} = \{X_{n^3t}(i) > X_{n^3t}(j)\}$.

$$\mathbb{E}\left(\frac{1}{n^2}D^n_{n^3t}\right)^2 = \frac{1}{n^4}\sum_{i<j}\sum_{k<l}\mathbb{P}[A_{i,j}\cap A_{k,\ell}].$$

The first step is to observe that there are only $O(n^3)$ terms in which two of the indices are equal so these can be ignored. When the four indices are distinct we can again apply Durrett and Neuhauser's [17] results to the 4-tuple of random walks $(X(i), X(j), X(k), X(l))$, to conclude that if $i/n \to x$, $j/n \to y$, $k/n \to x'$ and $l/n \to y'$

$$\mathbb{P}[A_{i,j}\cap A_{k,\ell}] \to \mathbb{P}_{x,y}[B_1(t) > B_2(t)]\mathbb{P}_{x',y'}[B_1(t) > B_2(t)].$$

From this it follows that

$$\mathbb{E}\left(\frac{1}{n^2}D^n_{n^3t}\right)^2 - \left(\mathbb{E}\frac{1}{n^2}D^n_{n^3t}\right)^2 \to 0.$$

In other words, the variance of $n^{-2}D^n_{n^3t}$ is asymptotically 0, and applying Chebyshev's inequality, we get the convergence in probability to the limit of the means. $\square$

4.3 Intermediate regime

The proof of Theorem 4 is a hybrid of the two previous proofs. We first truncate to show that it suffices to consider $i < j$ close together and far from the ends, then we compute second moments. We begin with a large deviations result:

Lemma 8. *For all $x > 0$ and $t > 0$ then*

$$\mathbb{P}(X_{nt}(0) > x) \le \exp(-x^2/8et) + \exp(-x\ln(2) - 2t).$$

Proof. First assume $x \le 4et$. From (13) we have $\mathbb{P}(X_{nt}(0) > x) \le \exp(-\theta x + t[e^\theta + e^{-\theta} - 2])$. When $0 \le \theta \le 1$

$$\begin{aligned} e^\theta + e^{-\theta} - 2 &= 2\left[\frac{\theta^2}{2} + \frac{\theta^4}{4!} + \frac{\theta^6}{6!} + \cdots\right] \\ &\le \theta^2\left[1 + \frac{\theta^2}{2^2} + \frac{\theta^4}{2^4} + \cdots\right] = \frac{\theta^2}{1-\theta^2/4} \le \frac{4\theta^2}{3}. \end{aligned}$$

Taking $\theta = x/4et$ which is ≤ 1 by assumption

$$\mathbb{P}(X_{nt}(0) > x) \le \exp\left(-\frac{x^2}{4et} + t\cdot\frac{4}{3}\left(\frac{x}{4et}\right)^2\right) \le \exp(-x^2/8et). \tag{17}$$

When $x > 4et$, note that $\mathbb{P}(X_{nt}(0) > x)$ is smaller than the probability that a Poisson random variable with mean $2t$ is greater than x. Thus for any $\theta > 0$ this is by Markov's inequality smaller than $\exp(-\theta x + 2t(e^\theta - 1))$. This is optimal when $e^\theta = x/2t$, in which case we find that

$$\mathbb{P}(X_{nt}(0) > x) \le \exp(-x\ln(x/2t) + x - 2t) \le \exp(-x\ln(2) - 2t) \tag{18}$$

since $x \ge 4et$. Equations (17) and (18) give us two bounds valid in different regions, so by summing them we get a bound that is everywhere valid, and this concludes the proof. $\square$

Proof of Theorem 4. By assumption we can pick $K_n \to \infty$ so that $K_n^2\sqrt{s}/n \to 0$. (We can also assume, for convenience, $K_n\sqrt{s} \in \mathbb{N}$). By Lemma 5,

$$\begin{aligned}\frac{1}{n\sqrt{s}} \sum_{i,j>i+K_n\sqrt{s}} \mathbb{P}^{[0,n]}(X_{ns}(i) > X_{ns}(j)) &\leq 8\frac{1}{\sqrt{s}} \sum_{x=K_n\sqrt{s}}^{\infty} \mathbb{P}(X_{ns}(0) > x/2) \\ &= 8\int_{K_n}^{\infty} \mathbb{P}(X_{ns}(0) > \lfloor x\sqrt{s}/2 \rfloor)dx.\end{aligned}$$

Applying Lemma 8 it follows that

$$\frac{1}{n\sqrt{s}} \sum_{i,j>i+K_n\sqrt{s}} \mathbb{P}^{[0,n]}(X_{ns}(i) > X_{ns}(j)) \to 0.$$

Letting $I_{i,j}$ be the indicator of $\{X_{ns}(i) > X_{ns}(j)\}$ it follows that

$$\frac{1}{n\sqrt{s}} \sum_{i,j>i+K_n\sqrt{s}} I_{i,j} \to 0 \quad \text{in probability,}$$

i.e., we can restrict our attention to close pairs. Once we do this, we can eliminate ones near the ends since

$$\frac{1}{n\sqrt{s}} \sum_{i<K_n\sqrt{s}, j\leq i+K_n\sqrt{s}} I_{i,j} \leq \frac{(K_n\sqrt{s})^2}{n\sqrt{s}} \to 0$$

by assumption. In a similar way we can eliminate $j > n - K_n\sqrt{s}$.

It follows that it is enough to consider random stirring on **Z**. The result of Durrett and Neuhauser [17] implies that if $s \to \infty$, $i \geq K_n\sqrt{s}$, $j \leq n - K_n\sqrt{s}$ and $(j-i)/\sqrt{s} \to x$ then

$$\mathbb{E}I_{i,j} \to \frac{1}{2}\mathbb{P}\left(\max_{0\leq t\leq 1} B_{4t} > x\right)$$

where the right-hand side is 0 if $x = \infty$. Writing Σ^* again for the i,j with $i \geq K_n\sqrt{s}$, $j \leq n - K_n\sqrt{s}$ and $0 < j-i < K_n\sqrt{s}$, and using the domination that comes from Lemma 8 it follows that

$$\frac{1}{n\sqrt{s}}\Sigma^*\mathbb{E}I_{i,j} \to \frac{1}{2}\mathbb{E}\max_{0\leq t\leq 1} B_{4t}.$$

The next step is to compute the second moment, which can be expressed, as in the proof of Theorem 3, as a sum of probabilities over four possible starting points, $i < j$ and $k < l$. The number of terms in this quadruple sum such that one index in $i < j$ is equal to one index of $k < l$ (say $i = k$), and such that both pairs are close (i.e., $j - i \leq K_n\sqrt{s}$ and $k - l \leq K_n\sqrt{s}$), is at most $n(K_n\sqrt{s})^2$. Dividing by $(n\sqrt{s})^2$, this tends to 0, so we can ignore this part of the summation. On the other hand, the result of Durrett and Neuhauser [17] implies that terms in which all four indices are different are asymptotically uncorrelated. A dominated convergence argument thus implies:

$$\frac{1}{(n\sqrt{s})^2}\Sigma^*_{i<j}\Sigma^*_{k<\ell}\mathbb{E}(I_{i,j}I_{k,\ell}) \to \left(\frac{1}{2}\mathbb{E}\max_{0\leq t\leq 1} B_{4t}\right)^2. \tag{19}$$

Indeed, using Cauchy-Schwarz's inequality, we get:

$$\mathbb{E}(I_{i,j}I_{k,l}) \le \mathbb{P}^{[0,n]}(X_{ns}(i) > X_{ns}(j)^{1/2}\mathbb{P}^{[0,n]}(X_{ns}(k) > X_{ns}(l)^{1/2}.$$

We may apply Lemma 5, and we note that in Lemma 8 we can also get

$$\mathbb{P}(X_{nt}(0) > x)^{1/2} \le \exp(-x\ln(2)/2 - t) + \exp(-x^2/16et)$$

by summing the square-roots of the two terms in (17) and (18) since only one of them applies in a given region. It is then easy to see that the assumptions of the dominated convergence theorem are satisfied and thus

$$\frac{1}{(n\sqrt{s})^2}\Sigma^*_{i<j}\Sigma^*_{k<\ell}(\mathbb{E}(I_{i,j}I_{k,\ell}) - \mathbb{E}(I_{i,j})\mathbb{E}(I_{k,l})) \to 0.$$

As in Theorem 3, this implies (19) and using Chebychev's inequality, this finishes the proof of Theorem 4. □

5 Riffle Shuffle

To prove Theorem 5, we make extensive use of the well-known trick which consists of studying the inverse random walk $(\sigma'_m, m \ge 1)$ rather than the random walk itself. This is a random walk on $\mathcal{S}_n$ which has the following dynamics. For each $m \ge 1$, σ'_m is obtained from σ'_{m-1} by deciding independently for each card if it goes in the upper or lower packet (each with probability 1/2). The upper packet is then put on top of the lower packet. Bayer and Diaconis call such a shuffle an *inverse 2-shuffle*, because there are two piles. More generally, consider the inverse a-shuffle, where a is an integer greater or equal to 2. This process is defined by the following dynamics: at each time step, every card is independently placed into one of a packets with equal probability. At the end, the a packets (some of them may be empty) are stacked one on top of the other. We shall use the following well-known lemma due to Bayer and Diaconis (see [4, Lemma 1] for a proof).

Lemma 9. *The law of σ_m^{-1} is the same as the result of an inverse 2^m-shuffle.*

Recall the definition of the number of rising sequences $R(\pi)$ of a permutation π in section 5, and formula (7). Since $R(\sigma) = \mathrm{Des}(\sigma^{-1}) + 1$, where

$$\mathrm{Des}(\sigma) = \#\{1 \le i \le n-1 : \sigma(i) > \sigma(i+1)\}$$

is the number of descents of σ, it suffices to study the number of descents after performing an inverse 2^m shuffle. We will use the terminology that σ has a descent at i if $\sigma(i+1) < \sigma(i)$.

Let $m = \lfloor \log_2 n - \log_2(C\log n) \rfloor$, for some $C > 0$ whose value will be determined later. In an inverse 2^m shuffle, each pile contains a Binomial (n, p) number of cards with $p = 1/2^m$. The next elementary lemma shows that with high probability each pile has more than half what we expect.

Lemma 10. *Let A be the event that each pile in an inverse 2^m shuffle contains at least $np/2$ cards where $p = 1/2^m$. Then $\mathbb{P}(A) = 1 - o(n^{-1})$.*

Proof. Let S be a Binomial(n,p) random variable with $p = 1/2^m$. The Laplace transform of S is $\mathbb{E}e^{-\theta S} = (1-p+pe^{-\theta})^n = \phi^n(\theta)$. By Markov's inequality

$$e^{-\theta np/2}\mathbb{P}(S \leq np/2) \leq \mathbb{E}e^{-\theta S}$$

so we have

$$\mathbb{P}(S \leq np/2) \leq \exp(n[\theta p/2 + \log\phi(\theta)]).$$

Taking $\theta = \ln 2$, $\phi(\theta) = 1 - p/2$, and hence $\log(\phi(\theta)) \leq -p/2$ so

$$\mathbb{P}(S \leq np/2) \leq \exp([\log 2 - 1]np/2).$$

Using this result with $p = 1/2^m$ and $m = \log_2(n/C\log n)$ where C is large, then

$$\mathbb{P}(S \leq np/2) \leq n^{-2}.$$

Since there are $2^m = o(n)$ piles, a simple union bound shows that with probability greater than $1 - o(n^{-1})$, all 2^m piles are $\geq np/2$. $\square$

The next step is to show that conditionally on the event A, with high probability each pile creates a descent in the permutation that results by putting the piles on top of one another. A moment of thought shows that the last card of pack i is not a descent for the resulting permutation only if the entire pack i is filled before the first card of pack $i+1$ is dropped, an event that we will call B_i. Whenever a card is dropped in one of the piles i or $i+1$, the probability it will go to pile i is just $1/2$, so if the eventual size of pack i is a and that of pack $i+1$ is b, $\mathbb{P}(B_i) = 1/2^{a+b} \leq 2^{-a}$. But conditionally on A, we know $a \geq np/2$. We conclude that

$$\mathbb{P}(B_i|A) \leq 2^{-np/2} = 2^{-C\log n/2} \leq n^{-2}$$

if C is large enough. Hence with probability $1 - o(n^{-1})$, there are at least $2^m - 1$ descents after a 2^m-inverse shuffle. This implies that $R(\sigma_m) = 2^m$ and that $D(m) = m$ with probability $1 - o(n^{-1})$. To finish the first part of Theorem 5 (the part $t < 1$) it suffices to note that if the distance is m at time $m = \log_2(n/C\log n)$, then it is also true that $D(m') = m'$ for smaller values of m', since the distance can increase by at most 1 at each time step.

To finish the proof, it remains to show that if $m = t\log_2 n$ and $t \geq 1$, $D(m) \sim \log_2 n$. We start by the following lemma.

Lemma 11. *With high probability there are at least $n/4$ non-empty piles.*

Proof. When a card is dropped, and k piles have already been started, the probability that the card is dropped in an empty pile is $1 - k/2^m$. If we focus on the first $n/2$ cards, $k \leq n/2$ necessarily, and since $2^m \geq n^t$, it follows that this probability is greater than $1-(1/2)n^{1-t} \geq 1/2$. Since this is independent for different cards, the total number of non-empty piles is greater than a Bernoulli random variable with parameters n and $1/2$. This is greater than $n/4$ with probability exponentially close to 1 by standard large deviations. $\square$

Ignoring the empty piles, we will argue in a way similar as above that these $n/4$ piles give rise to at least cn descents with high probability (for some $c > 0$) when they are stacked one on top of the other. To see this, note that each non-empty pile gives a new descent when it is stacked with the next non-empty pile with positive probability. Indeed, with positive probability it is the pile on the right which gets filled first and this automatically gives rise to a descent. For disjoint pairs of piles, this event happens independently and so by the law of large numbers there are at least cn descents with high probability. This implies

$$R(\sigma_m) \geq cn$$

with high probability. By (7), $D(m) = \lceil \log_2 R(\sigma_m) \rceil \geq \log_2 n - O(1)$ with high probability. On the other hand the Bayer-Diaconis formula (6) tells us that the diameter of the graph is $\lceil \log_2 n \rceil$ and hence we conclude

$$(\log_2 n)^{-1} D(\lfloor t \log_2 n \rfloor) \to_p 1$$

as claimed for $t \geq 1$.

References

[1] D. Aldous (1983). Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII. Lecture Notes in Math.* 986, 243–297. Springer, New-York.

[2] O. Angel, A. Holroyd, D. Romik. Directed random walk on the permutahedron. In preparation.

[3] O. Angel, A. Holroyd, D. Romik and B. Virag (2007). Random sorting networks. *Adv. in Math.*, 215(2):839–868. MR2355610

[4] D. Bayer and P. Diaconis (1992). Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2, 294-313. MR1161056

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

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

[7] B. Bollobás (1985). *Random graphs.* Academic Press, London. MR0809996

[8] B. Bollobás (1988). The isoperimetric number of a random graph, *European Journal of Combinatorics*, 9, 241-244. MR0947025

[9] B. Bollobás and F. de la Vega (1982). The diameter of random regular graphs. *Combinatorica*, 2, 125-134 MR0685038

[10] F.K. Chung and L. Lu (2001). The diameter of sparse random graphs. *Adv. Appl. Math.* 26, 257-279. MR1826308

[11] P. Diaconis (1988). *Group representation in Probability and Statistics*, Institute of Mathematical Statistics Lecture Notes, Vol. 11. MR0964069

[12] P. Diaconis, and R.L. Graham (1977). Spearman's footrule as a measure of disarray. *J. R. Statist. Soc. B*, 39, 262-268. MR0652736

[13] P. Diaconis, R. L. Graham and J. A. Morrison (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Struct. and Alg.* 1, 51–72. MR1068491

[14] P. Diaconis and M. Shahshahani (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* 57, 159-179. MR0626813

[15] R. Durrett (2003). Shuffling Chromosomes. *J. Theor. Prob.* 16, 725–750. MR2009200

[16] R. Durrett (2005). Genome Rearrangement: Recent Progress and Open Problems. In *Statistical Methods in Molecular Evolution*, edited by R. Nielsen, Springer, 2005. MR2161835

[17] R. Durrett and C. Neuhauser (1994). Particle systems and reaction-diffusion equations. *Ann. Prob.*, Vol. 22, No. 1, 289-333. MR1258879

[18] N. Eriksen (2005). Expected number of inversions after a sequence of random adjacent transpositions - an exact expression. *Discrete Mathematics*, 298, 155–168. MR2163446

[19] H. Eriksson, K. Erikkson, and J. Sjöstrand (2000). Expected number of inversions after k random adjacent transpositions. In D. Krob, A.A. Mikhalev, A.V. Mikhalev, eds. *Proceedings of Formal Power Series and Algebraic Combinatorics*, Springer-Verlag (2000) 677-685

[20] J. Fulman (2005). Stein's method and minimum parsimony distance after shuffles. *Electr. J. Probab.* 10, 901–924. MR2164033

[21] Kendall (1970). *Rank Correlation Methods*, 4th edn. London: Griffin.

[22] D. Knuth (1973). *The Art of Computer Programming*, Vol. 2. reading, Mass.: Addison-Wiley.

[23] L. Saloff-Coste (2003). Random Walks on Finite Groups. In: H. Kesten, ed. *Probability on Discrete Structures*, Encyclopaedia of Mathematical Sciences (110), Springer. MR2023654

[24] N.C. Wormald (2005). Models of random regular graphs (survey). Available at `http://www.ms.unimelb.edu.au/~nick/papers/regsurvey.pdf`