Geometric Aspects and Mixing Times of Coalescent and Fragmentation Processes

Batı Şengül

Hughes Hall

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

November, 2014

Anne ve Babama…

Statement of Originality

I hereby declare that my dissertation entitled "Geometric Aspects and Mixing Times of Coalescent and Fragmentation Processes" is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University. I further state that no part of my dissertation has already been or is concurrently submitted for any such degree of diploma or other qualification.

This dissertation is the result of my own work, done partly in collaboration with my supervisor, Dr. Nathanaël Berestycki.

Acknowledgements

First and foremost I would like to thank my supervisor Nathanaël Berestycki. He has supported and inspired me throughout my PhD and helped me greatly through the difficulties that arose.

My thanks also to many friends in Cambridge who made my stay there an unforgettable part of my life. A big thanks to Julio Brau, who lived with me for the duration of the entire PhD, Kolyan Ray, Damon Civin, Meline Joaris, Marc Briant, Sara Merino, Edward Mottram, Ludovic De La Cesbron, Alan Sola for being a fun, lively and nice group of people to hang out with, and Prof. Chang for his inspirational letters. Special thanks to my office mates: Kostas Papafitsorous, who wrote a paper with me which is sadly not a part of this thesis, and Spencer Hughes, who discussed a lot of geometry with me which eventually lead to [13]. Thanks also to my salsa friends Rafael Sanchez Cid, Martyna Popis, Aga Wabik, Carme Chezleduc and the rest of Dile Que Si for filling my nights with some Latin inspiration. Thanks also to my buddy Mike Bush, I'm very happy that I got to see more of him the past couple of years.

I owe a huge thank you to Marion Hesse for her love, care and support. You helped me in every aspect of my life, and made me very happy even through the difficult times.

Son olarak aileme bana verdikleri destek için teşekkür etmek istiyorum, benim hep hatırımı soran anneanneme, Dilek'e, Eylül'e ve özellikle anneme ve babama: ikiniz bana yıllarca sevgi, mutluluk ve güç verdiğiniz. Bu tezi size ithaf ettim.

Contents

CHAPTER 1

Introduction

A coalescent-fragmentation process is a particle system in which particles can coalesce or fragment as time runs forward. A process for which only coalescence (resp. fragmentation) occurs is called a coalescent (resp. fragmentation process). Coalescent-fragmentation processes have found a large number of applications in physics, chemistry and biology.

This thesis consists of two separate works, both of which study coalescent-fragmentation processes from a geometric point of view. The first work, Şengül [50], presented in Chapter 2, deals with scaling limits of coalescent processes near time zero. We approach this problem by viewing the coalescent process as a metric space and then taking a suitable scaling limit of this metric space. The second work, Berestycki and Şengül [13], presented in Chapter 3, concerns the mixing times of random walks on the permutation group. These random walks can be seen as certain coalescent-fragmentation processes. We investigate their mixing times by viewing the permutation group as a graph and showing bounds on a certain notion of curvature on this graph.

This chapter contains the background for both of these works and some of the ideas behind the proofs.

1.1 Coalescent Theory

1.1.1 Kingman's Coalescent

Coalescent theory is a retrospective model of population genetics. It traces back the lineages of a sample from the current population to find their most recent common ancestor. The inheritance relationships between individuals are typically represented as a phylogenetic tree, see Figure 1.1.

[1]Generated using the Interactive Tree of Life: `http://itol.embl.de/`

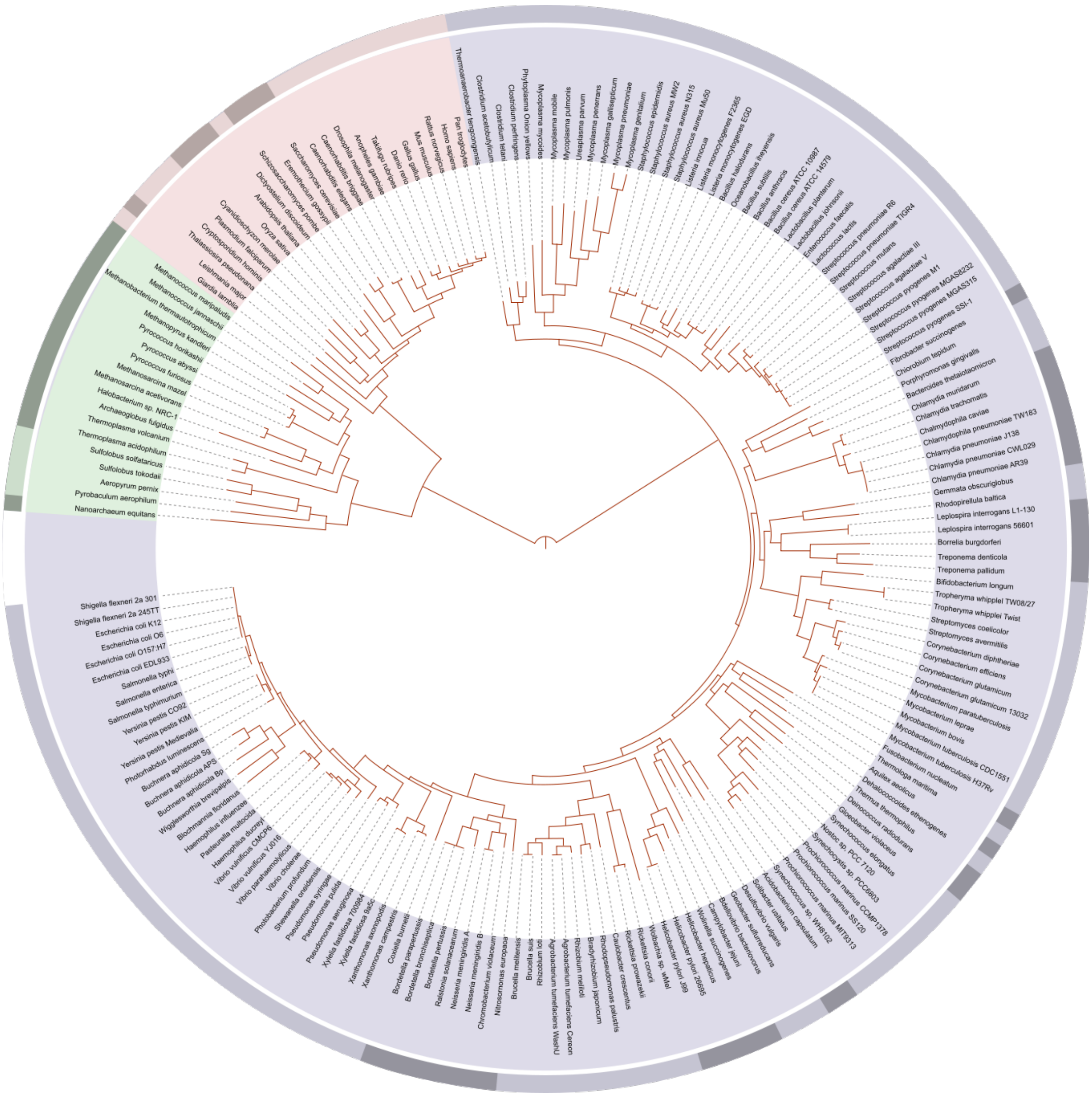


Figure 1.1: A phylogenetic tree showing the ancestral relationship of sampled organisms.[1]

The most natural probabilistic model of a coalescent appeared in the seminal paper of Kingman [33]. It is defined as follows. For $n \in \mathbb{N}$, a coalescent process $\Pi^{(n)} = (\Pi^{(n)}(t) : t \geq 0)$ is a $\mathcal{P}_n$-valued process, where $\mathcal{P}_n$ is the set of partitions of $[n] := \{1, \ldots, n\}$. Suppose that initially we start from the trivial partition consisting of singletons and thereafter each pair of blocks merges at rate 1. One trivial but important property is that this process is consistent, meaning that for every $n \in \mathbb{N}$, the projection of $\Pi^{(n+1)}$ onto $\mathcal{P}_n$ is the same as $\Pi^{(n)}$. This allows us to pass to the limit as $n \to \infty$ using Kolmogorov's extension theorem.

Definition 1.1.1 (Kingman's coalescent)**.** *Let $\mathcal{P}_\infty$ denote the set of partitions of $\mathbb{N}$, then there exists a process $\Pi = (\Pi(t) : t \geq 0)$ such that for any $n \in \mathbb{N}$, the projection onto $\mathcal{P}_n$ has the same law as $\Pi^{(n)}$ as described above. The process Π is called Kingman's coalescent.*

The canonical ordering on the blocks of $\Pi(t)$ is by infimum: for $i < j$ we have that $\inf \Pi_i(t) < \inf \Pi_j(t)$, where $\Pi_i(t)$ is the i-th block of $\Pi(t)$ and by convention $\inf \emptyset = \infty$.

Kingman's coalescent Π describes the ancestral relationships between individuals in the following way. For $t \geq 0$, i and j are in the same block of $\Pi(t)$, denoted by $i \overset{\Pi(t)}{\sim} j$, if the individuals i and j share a common ancestor t years ago. Hence as t increases, the blocks of the partition and thus the lineages coalesce.

An interesting phenomenon, known as coming down from infinity, occurs in Kingman's coalescent whereby the coalescent has finitely many blocks for all times $t > 0$. Let $N = (N(t) : t \geq 0)$ denote the number of blocks of Π. The following result is well known.

Proposition 1.1.2. *Kingman's coalescent comes down from infinity meaning that*

$$\mathbb{P}(N(t) < \infty \text{ for all } t > 0) = 1.$$

Moreover the following convergence holds almost surely and in L^1 as $t \downarrow 0$,

$$tN(t) \to 2.$$

Heuristic. At time $t > 0$ there are $\binom{N(t)}{2}$ many pairs of blocks (this number may in fact be infinite). Each of these pairs of blocks merges at rate 1 and when two blocks merge, N decreases by 1. Hence $N(t)$ roughly solves the differential equation

$$\frac{d}{dt}N(t) \approx -\binom{N(t)}{2} \approx -\frac{N(t)^2}{2}$$

with the initial condition $N(0) = \infty$. The solution to this equation is given by $N(t) \approx 2/t$. □

1.1.2 Λ-coalescents

In this section we describe a generalisation of Kingman's coalescent in order to allow for coalescents with multiple mergers. Here we will detail the account of Pitman, however note that this was discovered independently by Pitman [39], Sagitov [46] and Donnelly and Kurtz [22]. We refer to Bertoin [14] and Berestycki [10] for an overview.

To make this generalisation we must first make precise which type of processes on $\mathcal{P}_\infty$ we are considering. In order to do so we first introduce the notion of exchangeability.

Definition 1.1.3 (Exchangeability). *For a partition $\pi \in \mathcal{P}_\infty$ and a bijection $\sigma : \mathbb{N} \to \mathbb{N}$ with finite support, define $\sigma(\pi) \in \mathcal{P}_\infty$ by $i \overset{\sigma(\pi)}{\sim} j$ if and only if $\sigma(i) \overset{\pi}{\sim} \sigma(j)$. Then a random partition $\pi \in \mathcal{P}_\infty$ is called exchangeable if π and $\sigma(\pi)$ have the same distribution for every bijection $\sigma : \mathbb{N} \to \mathbb{N}$ with finite support.*

Next we define the class of processes that are of interest.

Definition 1.1.4 (Simple Coalescent Process). *A process $\Pi = (\Pi(t) : t \geq 0)$ on $\mathcal{P}_\infty$ is called a simple coalescent process if the following hold:*

(i) the process Π merges blocks as it evolves in time,

(ii) there are no simultaneous mergers,

(iii) for each $t \geq 0$, $\Pi(t)$ is exchangeable,

(iv) for each $n \in \mathbb{N}$ let $\Pi^{(n)}$ denote the restriction of Π to $\{1, \ldots, n\}$, then $\Pi^{(n)}$ has the same law as $\Pi^{(n+1)}$ restricted to $\{1, \ldots, n\}$.

It is not hard to check that Kingman's coalescent is a simple coalescent processes.

The properties (i),(iii) and (iv) in the definition above arise naturally from population genetics. Assumption (ii) can be relaxed to obtain what are known as Ξ-coalescents. We do not discuss this here but refer the interested reader to Schweinsberg [49].

To obtain a simple coalescent process consider the following construction. For $b \in \mathbb{N}$ and $k \geq 2$ let $\lambda_{b,k}$ denote the rate at which k fixed blocks merge when there are b blocks present. Definition 1.1.4 (iii),(iv) imply

$$\lambda_{b,k} = \lambda_{b+1,k} + \lambda_{b+1,k+1}. \tag{1.1}$$

Indeed as there are no simultaneous mergers, k fixed blocks among b blocks may merge in two ways when we reveal an extra block $b + 1$: either the k blocks merge by themselves, without the extra block, or the k blocks together with the extra block merge.

Using (1.1) Pitman was able to show the following result.

Theorem 1.1.5 (Pitman [39, Theorem 1])**.** *There exists a coalescent process* $\Pi = (\Pi(t) : t \geq 0)$ *such that the rate at which* k *fixed blocks merge when there are* b *blocks present is given by* $\lambda_{b,k}$*, if and only if, there exists a finite measure* Λ *on* $[0,1]$ *such that*

$$\lambda_{b,k} = \int_0^1 p^{k-2}(1-p)^{b-k}\Lambda(dp). \tag{1.2}$$

This leads to the following natural definition.

Definition 1.1.6 (Λ-coalescents)**.** *For a finite measure* Λ *on* $[0,1]$*, a* Λ*-coalescent is a simple coalescent process where the rate at which* k *fixed blocks merge when there are* b *blocks present is given by* $\lambda_{b,k}$ *defined in* (1.2)*.*

We give some important and well known Λ-coalescents in the table below.

Name	$\Lambda(dp) =$	$\lambda_{b,k} =$
Kingman's coalescent	$\delta_0(dp)$	$\begin{cases} 1 & \text{if } k=2 \\ 0 & \text{otherwise} \end{cases}$
Bolthausen-Sznitman coalescent	dp	$\dfrac{(k-2)!(b-k)!}{(b-1)!}$
Beta$(2-\alpha,\alpha)$-coalescent, $\alpha \in (0,2)$	$\dfrac{p^{1-\alpha}(1-p)^{\alpha-1}\,dp}{\Gamma(2-\alpha)\Gamma(\alpha)}$	$\dfrac{\sin(\pi\alpha)\Gamma(k-\alpha)\Gamma(b+\alpha-k)}{\pi\Gamma(b)(1-\alpha)}$

Table 1.1: Examples of Λ-coalescents.

The Beta$(2-\alpha,\alpha)$-coalescent for $\alpha = 1$ is just the Bolthausen-Sznitman coalescent and when $\alpha \to 2$ the Beta$(2-\alpha,\alpha)$-coalescent, as a process, converges to Kingman's coalescent.

There is a useful construction of a Λ-coalescent from a Poisson point process in the case when Λ is non-atomic. Let Λ be a finite measure on $[0,1]$ without any atoms. Let $\mathcal{M}$ be a Poisson point process on $[0,\infty) \times [0,1]$ of intensity $dt \otimes \Lambda(dp)p^{-2}$. For each atom $(t,p) \in \mathcal{M}$ of the point process, we perform a p-merger at time t: for each block of $\Pi(t-)$ flip a coin independently with probability p of heads, then merge all the blocks that are marked by heads. Thus if we assume that $\Pi(t-) = b$ then the probability we merge k fixed blocks is precisely given by $\lambda_{b,k}$ in (1.2).

At this point it is natural to investigate the coming down from infinity phenomenon for Λ-coalescents. It is customary to assume that $\Lambda(\{1\}) = 0$ as this avoids the uninteresting case where the coalescent suddenly merges all the blocks present. We will always implicitly assume this. Then there is a dichotomy between staying infinite forever and coming down from infinity.

Proposition 1.1.7 (Pitman [39, Proposition 23]). *Suppose that Λ is a finite measure on $[0,1]$ with no mass at one. Let $\Pi = (\Pi(t) : t \geq 0)$ be the associated Λ-coalescent and let $N = (N(t) : t \geq 0)$ be the number of blocks of Π. Then*

$$\mathbb{P}(\{N(t) = \infty \, \forall t > 0\} \cup \{N(t) < \infty \, \forall t > 0\}) = 1.$$

Sufficient and necessary conditions for a Λ-coalescent to come down from infinity were first given by Schweinsberg [48]. Later Bertoin and Le Gall [15] remark the equivalence of this condition to an integral test.

Theorem 1.1.8 (Schweinsberg [48, Theorem 1] and Bertoin and Le Gall [15, eq. (31)]). *Suppose that Λ is a finite measure on $[0,1]$ with no mass at one. For $b \geq 2$ let $\gamma_b = \sum_{k=2}^{b}(k-1)\binom{b}{k}\lambda_{b,k}$. Then the Λ-coalescent comes down from infinity if and only if*

$$\sum_{b=2}^{\infty} \gamma_b^{-1} < \infty. \tag{1.3}$$

Equivalently this occurs if and only if

$$\int_1^{\infty} \frac{dx}{\psi(x)} < \infty \tag{1.4}$$

where $\psi(x) = \int_0^1 (e^{-xp} - 1 + xp)p^{-2}\Lambda(dp)$.

Heuristics for (1.3). We present the heuristic proof in Berestycki [10, p. 76] which is very similar to the heuristic proof of Proposition 1.1.2. The quantity γ_b gives the rate of decrease when there are b blocks present. Indeed, there are $\binom{b}{k}$ many ways to choose k blocks from b blocks and k chosen blocks merge at rate $\lambda_{b,k}$. Once such a merger has taken place, the number of blocks decreases by $k-1$.

Thus roughly speaking the number of blocks $N(t)$ at time $t > 0$ solves the differential equation

$$\frac{d}{dt}N(t) \approx -\gamma_{N_t}$$

with initial condition $N(0) = \infty$. Hence it follows that for each $t > 0$

$$\int_0^t \frac{\frac{d}{dt}N(s)}{\gamma_{N_s}} ds \approx -t$$

Using the substitution $x = N(s)$ we get that

$$\int_{N_t}^{\infty} \frac{1}{\gamma_x} dx \approx t.$$

hence $N(t) < \infty$ if and only if $\int_1^\infty \frac{1}{\gamma_x}\, dx < \infty$ which is the same as the condition (1.3). $\square$

1.1.3 Behaviour of SRV(α) Coalescent Processes near the Time Zero

The main objective of the paper presented in Chapter 2 is to understand how a Λ-coalescent process comes down from infinity. In order to examine this we first make an assumption about the measure Λ.

Definition 1.1.9 (SRV(α)). *We say that a measure Λ on $[0,1]$ is SRV(α) if Λ is strongly regularly varying with index $\alpha \in (1,2)$. That is when $\Lambda(dp) = f(p)\, dp$ and there exists a constant $A_\Lambda > 0$ such that*

$$f(p) \sim A_\Lambda p^{1-\alpha} \qquad \text{as } p \to 0 \tag{1.5}$$

where the above notation means that the quotient of both sides approaches 1. We say that Λ is SRV(2) when $\Lambda = \delta_{\{0\}}$ and $A_\Lambda = 1$.

The SRV(α) measures are used to model populations in which there is large variability in the offspring distribution (see Berestycki [10, Section 3.2] for example). If Λ is SRV(2) then the Λ-coalescent is Kingman's coalescent. Further one can easily check (see Table 1.1) that for $\alpha \in (1,2)$, Beta$(2-\alpha,\alpha)$-coalescents are covered by this assumption and in this case Λ satisfies (1.5) with $A_\Lambda = (\Gamma(\alpha)\Gamma(2-\alpha))^{-1}$.

It is not hard to check that if Λ is SRV(α) for $\alpha \in (1,2]$, then the corresponding Λ-coalescent comes down from infinity. While it is possible to extend the definition of SRV(α) to $\alpha \in (0,1)$, we do not concern ourselves with this case as for $\alpha \in (0,1)$ the corresponding Λ-coalescent stays infinite forever. We assume henceforth that Λ is an SRV(α) measure with $\alpha \in (1,2]$.

First we present a theorem by Berestycki, Berestycki, and Schweinsberg [9] about the rate at which Π comes down from infinity which extends Proposition 1.1.2.

Theorem 1.1.10 (Berestycki, Berestycki, and Schweinsberg [9, Theorem 1.1]). *Suppose that Λ is a finite SRV(α) measure with $\alpha \in (1,2]$, $\Pi = (\Pi(t) : t \geq 0)$ is a Λ-coalescent and $N = (N(t) : t \geq 0)$ is the process which counts the number of blocks of Π. The following limit holds almost surely as $t \downarrow 0$:*

$$t^{1/(\alpha-1)} N(t) \to \left(\frac{\alpha}{A_\Lambda \Gamma(2-\alpha)}\right)^{1/(\alpha-1)}.$$

There are many frameworks for taking the scaling limit of a coalescent process near time zero. One such approach is to consider the asymptotic frequencies. For $t \geq 0$ and $i \in \mathbb{N}$ we define the asymptotic frequency of $\Pi_i(t)$ as

$$\Phi_i(t) = \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{\{k\in\Pi_i(t)\}}, \tag{1.6}$$

with the convention that $\Phi_i(t) = 0$ for $i > N(t)$. It is not hard to show that the limit in (1.6) exists by using the exchangeability of $\Pi(t)$. As Π comes down from infinity we have that almost surely for all $t > 0$,

$$\sum_{i=1}^{\infty} \Phi_i(t) = 1, \tag{1.7}$$

see Pitman [39, Theorem 8].

Naturally one may set $\Phi(t) = (\Phi_1(t), \Phi_2(t), \dots)$, view $\Phi = (\Phi(t) : t \geq 0)$ as a process on $[0,\infty)^{\mathbb{N}}$ and then take a scaling limit at time zero. Using Theorem 1.1.10, one can guess that the correct scaling factor is $\epsilon^{-1/(\alpha-1)}$ and try to obtain a scaling limit $\Psi = (\Psi(t) : t \geq 0)$ of $(\epsilon^{-1/(\alpha-1)}\Phi(\epsilon t) : t \geq 0)$ in the Skorokhod sense as $\epsilon \to 0$. There is, however, a substantial problem with this viewpoint which we now describe in detail.

Note that a sequence $x^{(n)}$ converges to x in $[0,\infty)^{\mathbb{N}}$ under the product topology if for every $m \in \mathbb{N}$,

$$\sup_{i\leq m} |x_i^{(n)} - x_i| \to 0 \tag{1.8}$$

as $n \to \infty$. Now consider the asymptotic frequencies

$$\epsilon^{-1/(\alpha-1)}\Phi_k(\epsilon t) \text{ such that } t \in [0,1] \text{ and } k \text{ is chosen so that } \frac{N(\epsilon)}{2} \leq k \leq N(\epsilon). \tag{1.9}$$

From (1.8) we see that the frequencies in (1.9) will not be seen in the limit as $\epsilon \to 0$ as they will be "pushed out to infinity". On the other hand many of the frequencies in (1.9) will merge with $\epsilon^{-1/(\alpha-1)}\Phi_1(\epsilon t)$ during the time interval $[0,1]$. Thus the result is that in the limit $(\Psi_1(t) : t \in [0,1])$ will increase its size without merging with other frequencies. Hence we have missed some of what is happening.

Similar problems occur when considering other natural orderings of $(\Phi_1(t), \Phi_2(t), \dots)$ e.g. ordered by size. Any ordering that avoids this problem is too complex to study with the current tools available to us. The only exception is the case when $\alpha = 2$ where one can use the construction in Aldous [4, Section 4.2]. In order to overcome this difficulty we view the coalescent as a metric space.

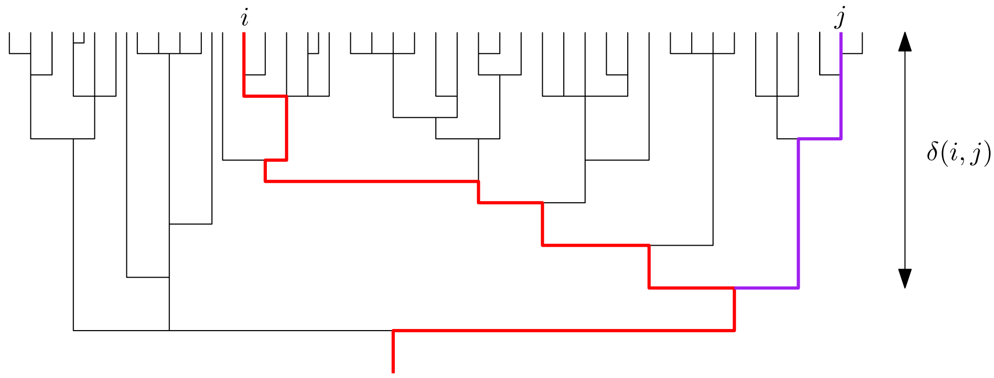


Figure 1.2: Tree view of the Evans space associated to a coalescent. The distance between a pair of points i, j is given by the first time the unique paths from i and j to the root meet.

1.1.4 Evans Metric Space

As before we assume that Π is a Λ-coalescent where Λ satisfies SRV(α) for $\alpha \in (1, 2]$. We wish to represent Π as a metric space. This is done by the so called Evans space (E, δ) associated to the coalescent Π, which is defined as follows. First we define a metric δ on $\mathbb{N}$ by letting it measure the time of the most recent common ancestor:

$$\delta(i, j) = \inf\{t > 0 : i \overset{\Pi(t)}{\sim} j\},$$

see Figure 1.2. Next we let (E, δ) be the completion of $(\mathbb{N}, \delta)$. Clearly (E, δ) contains all the information about the coalescent Π. This space was introduced by Evans [26] in the case of Kingman's coalescent.

We begin by showing compactness for (E, δ).

Proposition 1.1.11. *The space (E, δ) is compact.*

Proof. Recall that a metric space is compact if and only if it is totally bounded (for each $\epsilon > 0$, there are finitely many ϵ-balls that cover the space) and complete. As (E, δ) is complete by definition, we just need to check that it is totally bounded. A closed ball of radius $\epsilon > 0$ around i is given by the closure of $\{j \in \mathbb{N} : j \overset{\Pi(\epsilon)}{\sim} i\}$. Hence every closed ball of radius $\epsilon > 0$ corresponds to a block of $\Pi(\epsilon)$. Thus the number of such closed balls of radius ϵ is precisely $N(\epsilon)$, the number of blocks at time ϵ. The result now follows since Π comes down from infinity. $\square$

Our goal is to find a suitable way to describe the process near time zero. The way we do this geometrically is as follows. Let $\epsilon > 0$ and fix a point $i \in \mathbb{N}$ and a radius $r > 0$. Consider the closed ball $B(i, r\epsilon)$ around i of radius $r\epsilon > 0$. Note that this block contains all $j \in \mathbb{N}$ such that the block containing j merges with the block containing 1 during the time interval $[0, r\epsilon]$. Consider the metric space $(B(i, r\epsilon), \epsilon^{-1}\delta)$, where the scaling on the metric means that the diameter of the space is equal to r for any $\epsilon > 0$. A main result of Chapter 2 is the verification of the following convergence.

Theorem 1.1.12. *There exists a metric space $(\mathbb{S}, d_{\mathbb{S}})$ and a point $o \in \mathbb{S}$ such that when we let $\epsilon \downarrow 0$, the space $(B(i, r\epsilon), \epsilon^{-1}\delta)$ converges in a certain sense to $(B_{\mathbb{S}}(o, r), d_{\mathbb{S}})$.*

Note that the triple $(\mathbb{S}, d_{\mathbb{S}}, o)$ in Theorem 1.1.12 depends on the value of $\alpha \in (1, 2]$.

We will characterise the space $(\mathbb{S}, d_{\mathbb{S}})$ momentarily. First we present a result which is both crucial in the proof of Theorem 1.1.12 and is of independent interest. This result describes the mergers of the block containing 1 at small times and this description will allow us to depict the space $(\mathbb{S}, d_{\mathbb{S}})$. Roughly speaking, this result should be interpreted as a local limit in the spirit of Benjamini and Schramm [6]. More precisely, for $\epsilon > 0$ and $r \in [0, 1)$ let $\mathcal{Z}_\epsilon(r)$ be the number of blocks of $\Pi\left((1-r)\,\epsilon\right)$ that make up $\Pi_1(\epsilon)$, the block containing 1 at time ϵ. Hence there exists $1 = i_1 < \cdots < i_{\mathcal{Z}_\epsilon(r)}$ such that

$$\Pi_1(\epsilon) = \Pi_{i_1}((1-r)\epsilon) \cup \cdots \cup \Pi_{i_{\mathcal{Z}_\epsilon(r)}}((1-r)\epsilon).$$

Theorem 1.1.13. *As $\epsilon \to 0$, $\mathcal{Z}_\epsilon \to \mathcal{Z}$ in the Skorokhod sense on $[0, s]$ for every $s \in [0, 1)$. The process $\mathcal{Z}$ is an inhomogeneous Markov process with generator*

$$L_r(f)(i) = A_\Lambda \sum_{j \geq 1} (j+i) \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} [f(i+j) - f(i)]$$

when $\alpha \in (1, 2)$ and

$$L_r(f)(i) = \frac{(i+1)}{1-r}[f(i+1) - f(i)]$$

when $\alpha = 2$.

One can also write the limiting process $\mathcal{Z}$ as a time-change of a certain Galton-Watson process with immigration (see Chapter 2).

We will now depict how the closed unit ball $B_{\mathbb{S}}(o, 1) \subset (\mathbb{S}, d_{\mathbb{S}})$ is constructed from the process $\mathcal{Z}$. First construct a tree T from the process $\mathcal{Z}$. Start the tree with one particle called the root. Whenever the process $\mathcal{Z}$ makes a jump of size k select a particle uniformly at random; this particle gives birth to k offspring. Thus for each $r \in [0, 1)$, there are precisely $\mathcal{Z}(r)$ particles which are at distance r from the root. The process $\mathcal{Z}(r)$ explodes as $r \to 1$ so we have infinitely many particles at distance one from the root. The space $B_{\mathbb{S}}(o, 1)$ is the set of particles at distance one from the root. For each $v, w \in B_{\mathbb{S}}(o, 1)$ there exists two unique paths from the root ending at the points v, w and these paths deviate at distance $h_{v,w} \geq 0$ from the root. The distance between two points is given by $d_{\mathbb{S}}(v, w) = 1 - h_{v,w}$.

In the case of Kingman's coalescent we have an alternative construction of $(\mathbb{S}, d_{\mathbb{S}})$ from a two-sided Brownian motion $W = (W_t : t \in \mathbb{R})$ as follows. Let $\mathcal{N} := \{t \in \mathbb{R} : W(t) = 0\}$ be the zero set of W. For $x, y \in \mathcal{N}$ with $x \leq y$ define a pseudo-metric on $\mathcal{N}$ by

$$d_{\mathbb{S}}(x, y) := \sup\{W(t) : t \in [x, y]\}.$$

Note that if there are only negative excursions between x and y for $x \neq y$, then $d_{\mathbb{S}}(x, y) = 0$. We then define $\mathbb{S}$ to be the quotient space $\mathcal{N}/\sim$ where $x \sim y$ if and only if $d_{\mathbb{S}}(x, y) = 0$. We will verify that $(\mathbb{S}, d_{\mathbb{S}})$ can be constructed in this way for $\alpha = 2$ in Chapter 2, Theorem 2.1.3.

1.2 Random Walks on the Permutation Group

1.2.1 Mixing Time of Random Walks on the Permutation Group

Let $\mathcal{S}_n$ denote the set of permutations of $\{1, \dots, n\}$. Any permutation $\sigma \in \mathcal{S}_n$ has a unique cycle decomposition which partitions $\{1, \dots, n\}$ into the orbits of σ. For example the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 4 & 3 & 6 & 1 & 2 & 8 & 7 \end{pmatrix}$$

has 4 orbits, $(1\ 5), (3), (2\ 4\ 6), (7\ 8)$. A fixed point i is such that $\sigma(i) = i$ and it is customary to not include the fixed points in the cycle decomposition. Thus this permutation is written as $\sigma = (1\ 5)(2\ 4\ 6)(7\ 8)$.

The cycle structure $(k_2, k_3, \dots)$ of a permutation is a vector of integers such that in the cycle decomposition of σ there are k_2 many 2-cycles (called transpositions), k_3 many 3-cycles and so on. For the example above we have that the cycle structure of σ is $(2, 1, 0, \dots)$ as there are two transpositions and one 3-cycle in the cycle decomposition.

A conjugacy class $\Gamma \subset \mathcal{S}_n$ is a set of permutations having the same given cycle structure. Let $|\Gamma|$ denote the support of Γ which is the number of non-fixed points of any permutation in Γ. Thus by definition we have that $|\Gamma| = \sum_{j \geq 2} j k_j$.

Every permutation can be decomposed into a product of transpositions. Such a decomposition is not necessarily unique but the number of transpositions needed is always either even or odd. We call a permutation σ even if the number of transpositions needed to write down σ is even, otherwise we call σ odd. Every element of a conjugacy class Γ has the same parity: they are either all odd or all even. Hence it makes sense to say if Γ is even or odd. It is well known that if Γ is odd then Γ generates the whole group $\mathcal{S}_n$ while if Γ is even then Γ generates the group $\mathcal{A}_n$ of even permutations.

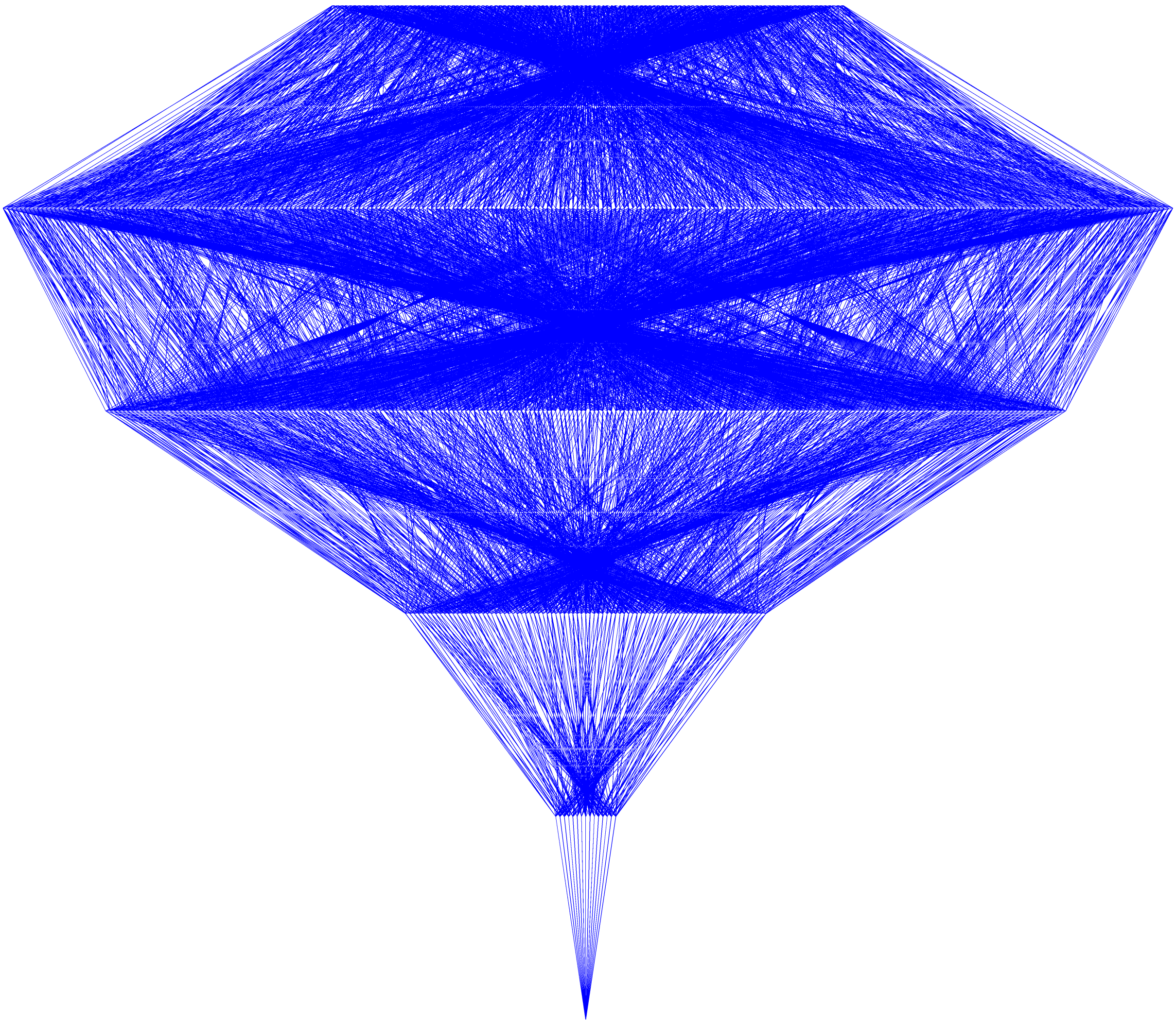


Figure 1.3: The Cayley graph of $\mathcal{S}_6$. An edge between two permutations is present whenever they differ by a transposition.

Let $\Gamma \subset \mathcal{S}_n$ be a conjugacy class. Then the discrete random walk $X = (X_t : t \geq 0)$ associated with Γ is constructed as follows: define $X_0 = \mathrm{id}$ and for $t \geq 1$ define

$$X_t = \gamma_1 \circ \cdots \circ \gamma_{N_t} \tag{1.10}$$

where $\gamma_1, \gamma_2, \ldots$ are i.i.d. random variables which are distributed uniformly in Γ and $N = (N_t : t \geq 0)$ is a rate 1 Poisson process. Then X is a Markov chain which has an invariant measure μ. If Γ is odd then μ is uniformly distributed on $\mathcal{S}_n$ and if Γ is even then μ is uniformly distributed on $\mathcal{A}_n$.

Let $p_t(\cdot)$ denote the law of X_t and define

$$d_{TV}(t) := ||p_t - \mu||_{TV} = \sup_{A \subseteq \mathcal{S}_n} |p_t(A) - \mu(A)| \tag{1.11}$$

to be the total variation distance between p_t and the invariant measure μ. The definition in (1.11) is useful for obtaining lower bounds for $d_{TV}(t)$ but not very useful for upper bounds. An alternative formula (see for example Levin, Peres, and Wilmer [34, Proposition 4.7]) is given by

$$d_{TV}(t) = \inf_{X_t' \sim X_t, X_\infty \sim \mu} \mathbb{P}(X_t' \neq X_\infty)$$

where the infimum is taken over all couplings of X_t' and X_∞ which are distributed X_t and μ respectively.

Let us first observe some basic facts about d_{TV} for large n. Initially we have that $\lim_{n\to\infty} d_{TV}(0) = 1$ as $n \to \infty$. The function $t \mapsto d_{TV}(t)$ is decreasing and intuitively the lower the value of $d_{TV}(t)$ the more the random walk resembles its invariant distribution. Finally $d_{TV}(\infty) = 0$ because the random walk at time t converges towards its invariant distribution as $t \to \infty$. We would like to know if there exists a time $t_{\mathrm{mix}} = t_{\mathrm{mix}}(n)$ such that for each $\epsilon > 0$,

$$\lim_{n\to\infty} d_{TV}((1-\epsilon)t_{\mathrm{mix}}) = 1 \quad \text{and} \quad \lim_{n\to\infty} d_{TV}((1+\epsilon)t_{\mathrm{mix}}) = 0. \tag{1.12}$$

The time t_{mix} is referred to as a *mixing time* and the phenomenon occurring in (1.12) is known as the *cut-off phenomenon*.

Diaconis and Shahshahani [21] were the first to study this phenomenon (at the same time and independently from Aldous [2]) and they arrived at the following result.

Theorem 1.2.1 (Diaconis and Shahshahani [21])**.** *Let Γ be the set of all transpositions and consider the random walk associated to Γ. Define $t_{mix} = (1/2)n\log n$, then for any*

$\epsilon > 0$ *we have that*

$$\lim_{n\to\infty} d_{TV}((1-\epsilon)t_{mix}) = 1 \quad \textit{and} \quad \lim_{n\to\infty} d_{TV}((1+\epsilon)t_{mix}) = 0.$$

Since the work of Diaconis and Shahshahani [21], a long standing open problem has been to show that the cut-off phenomenon holds for any conjugacy class Γ such that $|\Gamma| = o(n)$. Further it has been conjectured that for such conjugacy classes we have that $t_{\mathrm{mix}} = (1/|\Gamma|)n \log n$. We detail the history of this problem in Chapter 3. The following theorem, which is the primary result presented in Chapter 3, verifies this conjecture.

Theorem 1.2.2. *Suppose that $\Gamma \subset \mathcal{S}_n$ is an arbitrary conjugacy class with $|\Gamma| = o(n)$. Define $t_{mix} = (1/|\Gamma|)n \log n$, then for any $\epsilon > 0$ we have that*

$$\lim_{n\to\infty} d_{TV}((1-\epsilon)t_{mix}) = 1 \quad \textit{and} \quad \lim_{n\to\infty} d_{TV}((1+\epsilon)t_{mix}) = 0.$$

The proof of the statement $\lim_{n\to\infty} d_{TV}((1-\epsilon)t_{\mathrm{mix}}) = 1$ follows from a simple coupon collector argument. It is a straightforward adaptation of the argument in Diaconis and Shahshahani [21] for the case of transpositions which we present now.

Proposition 1.2.3. *Suppose that Γ is the set of all transpositions and consider the random walk associated to Γ. Then we have that for any $\epsilon > 0$,*

$$\lim_{n\to\infty} d_{TV}((1-\epsilon)(1/2)n\log n) = 1.$$

Proof. Fix $\epsilon > 0$ and let $t = (1-\epsilon)(1/2)n \log n$. Whenever we apply the transposition $(i\ j)$ we say that we have collected the coupons i and j. Thus the set of coupons is $\{1, \ldots, n\}$. A standard argument shows that at time $t = (1-\epsilon)(1/2)n \log n$ for any $m \in \mathbb{N}$, the probability that we are missing at least m coupons is asymptotically one.

Recall that

$$X_t = \tau_1 \circ \cdots \circ \tau_{N_t}.$$

Then if the coupon i has not been collected then none of the transpositions $\tau_1, \ldots, \tau_{N_t}$ have affected i. This implies that i is a fixed point of X_t. Thus we see that for any $m \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbb{P}(X_t \text{ has at least } m \text{ fixed points}) = 1.$$

On the other hand it is a well known fact that the probability a uniformly random permutation has exactly ℓ fixed points is asymptotically $e^{-1}/\ell!$. Let $A_m \subset \mathcal{S}_n$ be the set of permutations with at least m fixed points. Then from the definition given in (1.11) we have that for any $m \in \mathbb{N}$

$$\liminf_{n\to\infty} d_{TV}((1-\epsilon)(1/2)n\log n) \geq \liminf_{n\to\infty}[\mathbb{P}(X_t \in A_m) - \mu(A_m)] = 1 - \sum_{\ell=m}^{\infty} \frac{e^{-1}}{\ell!}.$$

The proof follows as $m \in \mathbb{N}$ can be taken arbitrarily large. $\square$

The main work in showing Theorem 1.2.2 lies in proving the statement $\lim_{n\to\infty} d_{TV}((1+\epsilon)t_{\text{mix}}) = 0$. The original proof of this in the case of transpositions by Diaconis and Shahshahani [21] makes use of representation theory to obtain bounds on so-called character ratios. The estimates on character ratios become harder as $|\Gamma|$ increases (unless $|\Gamma| > n/2$). A step towards the general case in Theorem 1.2.2 was then made in Berestycki, Schramm, and Zeitouni [12] where the authors use probabilistic methods to show Theorem 1.2.2 in the special case when the conjugacy class Γ is the set of k-cycles with k fixed. Their argument is divided into two parts: in the first part they deal with cycles of small length and in the second part they introduce a coupling to deal with cycles of large length.

We take a conceptually different approach; we bound the total variation distance using a notion of curvature, the so-called coarse Ricci curvature, on $\mathcal{S}_n$. We make use of the coupling in Berestycki, Schramm, and Zeitouni [12] and a result of Schramm [47] to prove bounds on the coarse Ricci curvature. Crucially, our argument enables us to ignore the cycles of small length. This turns out to be a significant advantage as the treatment of the small cycles in [12] is rather delicate.

For the remainder of the introduction we consider the case of random transpositions only as the ideas are cleaner to present in the simpler setting of transposition. We will show that for each $\epsilon > 0$,

$$d_{TV}((1+\epsilon)\,(1/2)n\log n) \to 0 \qquad \text{as } n \to \infty \tag{1.13}$$

by using a notion of Ricci curvature on graphs. We will introduce this notion of Ricci curvature and prove (1.13) in Section 1.2.3. In the next section we briefly recall the notion of Ricci curvature on manifolds.

1.2.2 Ricci Curvature on Manifolds

The notion of curvature, in particular Ricci curvature, plays an important role in analysis on manifolds. We introduce the notion of Ricci curvature in the spirit of Renesse and Sturm [40]. For this we first define a notion of distance between two measures on a metric space.

 16

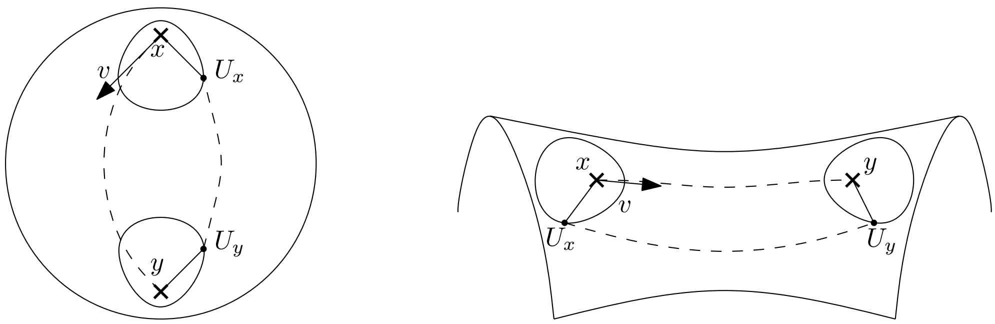


Figure 1.4: On the left is the two dimensional sphere which has positive Ricci curvature. On the right is a saddle like space which has negative Ricci curvature.

Definition 1.2.4 (L^1-Kantorovitch distance)**.** *The L^1-Kantorovitch distance $W_1(\pi, \nu)$ between two measures π and ν defined on a metric space (E, d) is given by*

$$W_1(\pi, \nu) = \inf_{X \sim \pi, Y \sim \nu} \mathbb{E}[d(X, Y)]$$

where the infimum is taken over all couplings of random variables X and Y distributed π and ν respectively.

Remark 1.2.5. *Using duality we can also write*

$$W_1(\pi, \nu) = \sup\{\int f \, d\pi - \int f \, d\nu : f \text{ is Lipschitz with Lipschitz constant } 1\}. \quad (1.14)$$

Now let (M, g) be an N-dimensional manifold and consider the intrinsic metric (in the metric space sense) d on M. Let $x \in M$ be a point on the manifold and let v be a tangent vector at x. Let $y \in M$ be a point which lies in the direction of v from x. Let m_x and m_y be two uniform probability measures over the spheres $S(x, \delta) = \{z \in M : d(x, z) = \delta\}$ and $S(y, \delta) = \{z \in M : d(y, z) = \delta\}$ respectively. The Ricci curvature $\mathrm{Ric}_x(v)$ at x in the direction of v can be defined by (see Ollivier [37, Corollary 10])

$$1 - \frac{W_1(m_x, m_y)}{d(x, y)} = \frac{\delta^2}{2N} \mathrm{Ric}_x(v) + O(\delta^3 + d(x, y)\delta^2). \quad (1.15)$$

Positive Ricci curvature is described in Ollivier [37] as when "small spheres are closer (in transportation distance) than their centers are". We give an example of spaces with positive and negative Ricci curvature in Figure 1.4.

A lower bound on Ricci curvature gives a lot of information on the structure of the manifold such as bounds on the spectral gap, volume growth and the growth of the fundamental group (see Wei [55]).

Ricci curvature is defined using the tangent space and so it is not obvious how to

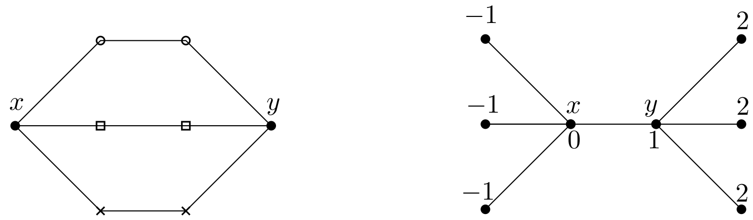


Figure 1.5: Example of coarse Ricci curvature on graphs with m_x uniformly distributed on the neighbours of x. On the left is a graph with $\kappa(x, y) = 2/3$. The optimal coupling is indicated with matching shapes. On the right is a graph with $\kappa(x, y) = -1$. The values of the optimal function in (1.14) is given by the numbers.

define a similar notion on a general metric space. In Lott and Villani [35], Sturm [52] and Sturm [53] a property called displacement convexity is used as the basis for the notion of Ricci curvature in certain metric spaces. However this notion is rather difficult to work with on graphs, even for the simple case of the hypercube (see Ollivier and Villani [38]). Instead we will work with coarse Ricci curvature defined in Ollivier [37] which we introduce in the following section.

1.2.3 Coarse Ricci Curvature

Let us now introduce the notion of coarse Ricci curvature as given by Ollivier [37]. For now we assume that (E, d) is a general metric space but we will later be interested in the case when E is a graph with graph distance d.

Previously in (1.15) we took m_x and m_y to be uniform probability measures on spheres around x and y respectively. Here we do not make this assumption; instead we consider the triple $(E, d, \{m_x\}_{x\in E})$ where $\{m_x\}_{x\in E}$ is a given set of probability measures.

Definition 1.2.6 (Coarse Ricci curvature)**.** *The coarse Ricci curvature of the triple* $(E, d, \{m_x\}_{x\in E})$ *between two points* $x, y \in E$ *with* $x \neq y$ *is given by*

$$\kappa(x, y) := 1 - \frac{W_1(m_x, m_y)}{d(x, y)}. \tag{1.16}$$

The coarse Ricci curvature κ *of the triple* $(E, d, \{m_x\}_{x\in E})$ *is given by*

$$\kappa := \inf_{x\neq y} \kappa(x, y).$$

We give an example of graphs with positive and negative coarse Ricci curvature in Figure 1.5. Compare this to Figure 1.4.

To make the connection to the random walk on $\mathcal{S}_n$ we take $E = \mathcal{S}_n$. The distance $d(\sigma, \sigma')$ between two elements $\sigma, \sigma' \in \mathcal{S}_n$ is given by the minimum number of transpositions one must apply to σ in order to obtain σ'. We take the measure m_x to be the law of a transposition random walk X_t^x started at x and ran for time $t = cn/2$. Note that m_x depends on $t = cn/2$ and hence κ depends on both n and $c > 0$. We suppress the dependence on n and write κ_c for the coarse Ricci curvature of $(\mathcal{S}_n, d, \{m_x\}_{x \in \mathcal{S}_n})$.

Theorem 1.2.7. *If $c \leq 1$,*

$$\lim_{n\to\infty} \kappa_c = 0.$$

On the other hand, for $c > 1$

$$\liminf_{n\to\infty} \kappa_c \geq \theta(c)^4 > 0 \tag{1.17}$$

where $\theta(c)$ is the solution in $(0,1)$ to the equation

$$\theta(c) = 1 - e^{-c\theta(c)}. \tag{1.18}$$

Let us first make some basic observations about κ_c for finite n. If we apply the same transpositions to a random walk started at x and a random walk started at y, this keeps their distance constant and hence $\kappa_c \geq 0$. In fact one can show also that $\kappa_c > 0$ for every *finite* n. However in the limit as $n \to \infty$, it is the case that $\kappa_c \to 0$ when $c \leq 1$ and κ_c remains bounded away from 0 when $c > 1$.

In turns out that in order to compute the value of κ_c it is not necessary to compute $\kappa_c(x, y)$ for all $x, y \in \mathcal{S}_n$.

In addition we show in Chapter 3 that $\limsup_{n\to\infty} \kappa_c \leq \theta(c)^2$ for $c > 1$ and propose the following conjecture.

Conjecture 1.2.8. *We have that for $c \geq 0$,*

$$\lim_{n\to\infty} \kappa_c = \theta(c)^2$$

with the convention that $\theta(c) = 0$ for $c \leq 1$.

An important consequence of (1.17) is that it allows us to obtain bounds on d_{TV} and hence derive the mixing time of the transposition random walk, as seen in the next proposition.

Proposition 1.2.9. *For any $s \geq 0$ we have that*

$$d_{TV}(scn/2) = \sup_{x\in\mathcal{S}_n} \|m_x^{*s} - \mu\|_{TV} \leq diam(\mathcal{S}_n)(1 - \kappa_c)^s$$

where $diam(\mathcal{S}_n) = \sup_{x,y\in\mathcal{S}_n} d(x, y) = n - 1$.

Note that this proposition holds more generally and follows from Ollivier [37, Corollary 21]. Proposition 1.2.9 is in fact a statement about the L^1-Kantorovitch distance bounding the total variation distance. Stated in this form it has been observed before under the name *path coupling* by Bubley and Dyer [16] and Jerrum [31] (see also Levin, Peres, and Wilmer [34, Chapter 14]).

Next we present the proof of the correct mixing time for the transposition walk using Theorem 1.2.7.

Proof of Theorem 1.2.2 in the case of transpositions. Recall that it remains to show (1.13): for each $\epsilon > 0$,

$$\lim_{n\to\infty} d_{TV}((1+\epsilon)t_{\text{mix}}) = 0.$$

Fix $\epsilon > 0$ and let $c > 1$ be arbitrary. Let $s = (1+\epsilon)\log n/c$ so that $scn/2 = (1+\epsilon)t_{\text{mix}}$. Then using Proposition 1.2.9 and Theorem 1.2.7 we have that

$$\limsup_{n\to\infty} d_{TV}((1+\epsilon)t_{\text{mix}}) \le \lim_{n\to\infty} n(1-\theta(c)^4)^s = \lim_{n\to\infty} n^{1+(1+\epsilon)\frac{\log(1-\theta(c)^4)}{c}}. \tag{1.19}$$

Now we claim that we can choose $c > 1$ suitably large so that the right hand side of the above equation is zero. Indeed an easy computation using l'Hopital's rule with (1.18) shows that

$$\lim_{c\uparrow\infty} \frac{\log(1-\theta(c)^4)}{c} = -1.$$

Hence it follows that for $c > 1$ sufficiently large the right hand side of (1.19) is zero. $\square$

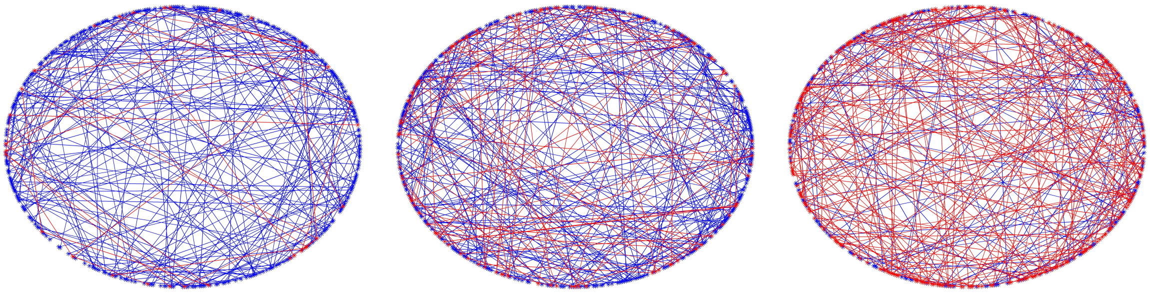
Throughout the rest of the introduction we will investigate why Theorem 1.2.7 holds. It turns out that the lengths of the cycles of X_t at time $t = cn/2$ play an important part in the proof of Theorem 1.2.7. In the next section we study the cycle lengths of X_t.

1.2.4 A Result of Oded Schramm Concerning Cycle Lengths

In this section we will explain a result of Schramm [47] which shows the asymptotic distribution of the cycle lengths of X_t for $t = cn/2$.

First we associate to X a certain random graph process $G = (G_t : t \ge 0)$ defined as follows. For each $t \ge 0$, G_t is a graph on $\{1, \ldots, n\}$ and initially G_0 contains no edges. Suppose that X makes a jump at time t and $X_t = X_{t-} \circ \tau$ where $\tau = (i\ j)$ is a uniform transposition. There are two cases, either the edge $\{i, j\}$ is present in G_{t-1} in which case we set $G_t = G_{t-1}$ or the edge $\{i, j\}$ is not present in G_{t-1} in which case we set G_t to be G_{t-1} together with the edge $\{i, j\}$.

Hence for each $t \ge 0$ any given edge is present independently with probability p_t given



(a) $c < 1$ **(b)** $c = 1$ **(c)** $c > 1$

Figure 1.6: The three phases of an Erdős-Renyi graph. The largest component is highlighted in red.

by

$$p_t = 1 - \exp\left\{ -\frac{t}{\binom{n}{2}} \right\}. \tag{1.20}$$

The graph G_t is what is referred to as an Erdős-Renyi graph $G(n, p_t)$ which has been the focal point of much research in the last 60 years and thus is well understood. First we present a theorem about the largest component of G_t (see also Figure 1.6).

Theorem 1.2.10 (Erdős and Rényi [25])**.** *Suppose that $t = cn/2$ for some $c > 0$ and for $i \geq 1$ let $|L_i(G_t)|$ denote the i-th largest component of G_t. Then*

Regime	*Component sizes*
$c < 1$	$\lvert L_1(G_t)\rvert \sim \frac{3}{(1-c)^2} \log n$
$c = 1$	$\lvert L_1(G_t)\rvert = O(n^{2/3}),\ \lvert L_2(G_t)\rvert = O(n^{2/3}), \ldots$
$c > 1$	$\lvert L_1(G_t)\rvert \sim \theta(c) n$
	$\lvert L_2(G_t)\rvert = O(\log n)$

Here, $\theta(c)$ is the unique solution in $(0, 1)$ to the equation (1.18).

Here, by $F(n) \sim G(n)$ we mean that $F(n)/G(n) \to 1$ in probability as $n \to \infty$. In the case when $c = 1$, the ratio of the largest component to $n^{2/3}$ converges to a non-trivial limit which is described in Aldous [3].

Recall that each transposition that has been applied to X prior to time t is an edge of G_t. Thus we see that every cycle is contained in a component of G_t in the sense that if i, j are in the same cycle of X_t then they are in the same component of G_t (see Figure

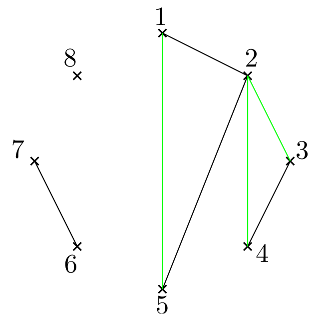


Figure 1.7: The graph G_t associated with $X_t = (1\ 2\ 5)(3\ 4)(6\ 7)$.

1.7). Let $t = cn/2$. Using Theorem 1.2.10 we can immediately deduce that for $c < 1$, every cycle of X_t has length $O(\log n)$ since the length of the longest cycle is less than the size of the largest component of G_t. Similarly, every cycle of X_t has length $O(n^{2/3})$ when $c = 1$. In the case when $c > 1$, Schramm [47] shows that there are cycles of length comparable to n with high probability (a different proof for the general case also appears in Berestycki [11]). Note that this does not immediately follow from the behaviour of G_t, since the giant component of G_t may be made up of many small cycles.

Schramm was also able to identify the limiting distribution as $n \to \infty$ of the renormalised cycle lengths. To present this result we first introduce some necessary notation. For a permutation $\sigma \in \mathcal{S}_n$ let $x = (x_1, \dots, x_\ell)$, $\ell \le n$, be the lengths of the cycles of σ, written in decreasing order. Define $\mathfrak{X}(\sigma) = x/n = (x_1/n, \dots, x_\ell/n, 0, \dots)$ to be the renormalised cycle lengths where we trail at the end by zeros. For example if $\sigma = (1\ 4\ 5)(2\ 3)(6)$, then $\mathfrak{X}(\sigma) = (1/2, 1/3, 1/6, 0, \dots)$. We see that $\mathfrak{X}(\sigma)$ is an element of the set

$$\Omega_\infty := \{(x_1 \ge x_2 \ge \dots) : x_i \ge 0 \text{ for each } i \in \mathbb{N} \text{ and } \sum_{i=1}^{\infty} x_i = 1\}.$$

Suppose again that $t = cn/2$, then we have seen earlier that the cycles of X_t have length $O(n^{2/3})$ when $c \le 1$. It it follows that when $c \le 1$, for every $i \in \mathbb{N}$, $\mathfrak{X}_i(X_t) \to 0$ in probability as $n \to \infty$. The following theorem deals with the case when $c > 1$.

Theorem 1.2.11 (Schramm [47, Theorem 1.1])**.** *Let $c > 1$, $t = cn/2$ and let $\theta(c)$ be the unique solution in $(0, 1)$ to the equation (1.18). Then we have that for each $m \in \mathbb{N}$,*

$$\left(\frac{\mathfrak{X}_1(X_t)}{\theta(c)}, \dots, \frac{\mathfrak{X}_m(X_t)}{\theta(c)}\right) \to (Z_1, \dots, Z_m)$$

in distribution as $n \to \infty$ where $Z = (Z_1, Z_2, \dots)$ is a Poisson–Dirichlet random variable, which we define below.

The law of Z can be derived using the so-called stick breaking construction. Let

$U_1, U_2, \ldots$ be i.i.d. uniform random variables on $[0,1]$. Let $Z_1^* = U_1$ and recursively for $k \geq 2$

$$Z_k^* = \left(1 - \sum_{i=1}^{k-1} Z_i^*\right) U_i.$$

Definition 1.2.12 (Poisson–Drichlet distribution)**.** *A Poisson–Dirichlet random variable Z is the vector $(Z_1^*, Z_2^*, \ldots)$ ordered in decreasing size.*

In the next section we use Theorem 1.2.11 to give a sketch proof of Theorem 1.2.7.

1.2.5 Sketch Proof of the Curvature Theorem

In this section we give a sketch proof of Theorem 1.2.7 for the case when $c > 1$ (see Chapter 3 for a full proof). In fact we will only outline a partial proof: we will show that for $c > 1$

$$\liminf_{n\to\infty} \inf_{x,y} \kappa_c(x,y) \geq \theta(c)^4 \tag{1.21}$$

where the infimum is taken over all $x, y \in \mathcal{S}_n$ with *even* distance.

Fix $c > 1$ and henceforth let $t = cn/2$. Writing out the definition of $\kappa_c(x,y)$, we wish to show

$$\limsup_{n\to\infty} \sup_{x,y} \frac{\mathbb{E}[d(X_t^x, X_t^y)]}{d(x,y)} \leq 1 - \theta(c)^4$$

for some appropriate coupling between X^x and X^y, where the supremum is taken over all x, y with even distance. We make several reductions: first, by vertex transitivity we can assume that $x = \mathrm{id}$. Also, by the triangle inequality (since W_1 is a distance), we can assume that $y = (i\, j) \circ (\ell\, m)$ is the product of two distinct transpositions. There are two cases to consider: either the supports of the transpositions are disjoint, or they overlap on one vertex. We will focus here on the first case where the support of the transpositions are disjoint; that is, i, j, l, m are pairwise distinct. The other case is dealt with similarly.

Let $t = cn/2$. Clearly by symmetry $\mathbb{E}d(X_t^{\mathrm{id}}, X_t^{(i,j)\circ(\ell,m)})$ is independent of i, j, ℓ and m, so long as they are pairwise distinct. Hence it is also equal to $\mathbb{E}d(X_t^{\mathrm{id}}, X_t^{\tau_1\circ\tau_2})$ conditioned on the event A that τ_1, τ_2 having disjoint support, where τ_1 and τ_2 are independent uniform random transpositions. This event has an overwhelming probability for large n, thus it suffices to construct a coupling between X^{id} and $X^{\tau_1\circ\tau_2}$ such that

$$\limsup_{n\to\infty} \mathbb{E}d(X_{cn/2}^{\mathrm{id}}, X_{cn/2}^{\tau_1\circ\tau_2}) \leq 2(1 - \theta(c)^4). \tag{1.22}$$

Indeed, it then immediately follows that the same is true with the expectation replaced by the conditional expectation given A.

23

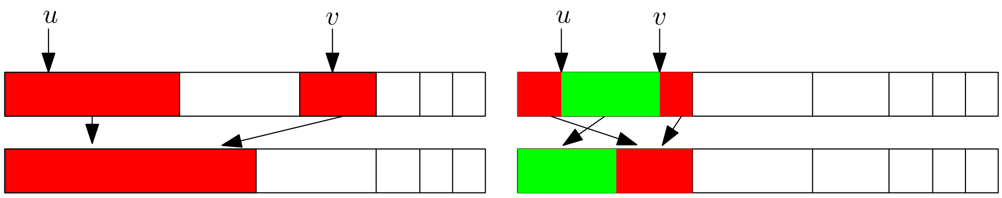


Figure 1.8: The transition of the split–merger walk. On the left is when two pieces coagulate and on the right is when two pieces fragment.

Now we sketch a coupling between X_t^{id} and $X_t^{\tau_1\circ\tau_2}$ which gives the correct bound in (1.22). Although we do not show this here, an important property of our coupling will be that

$$\mathbb{E}[d(X_t^{\mathrm{id}}, X_t^{\tau_1\circ\tau_2})] \approx 2(1 - \mathbb{P}(X_t^{\mathrm{id}} = X_t^{\tau_1\circ\tau_2})). \tag{1.23}$$

Now we sketch the coupling which gives $\liminf_{n\to\infty} \mathbb{P}(X_t^{\mathrm{id}} = X_t^{\tau_1\circ\tau_2}) \geq \theta(c)^4$. Observe the following: given $\mathfrak{X}(X_t^{\mathrm{id}}) = \bar{x}$ and $\mathfrak{X}(X_t^{\tau_1\circ\tau_2}) = \bar{x}'$, the random variables X_t^{id} and $X_t^{\tau_1\circ\tau_2}$ are distributed uniformly on $\{\sigma : \mathfrak{X}(\sigma) = \bar{x}\}$ and $\{\sigma : \mathfrak{X}(\sigma) = \bar{x}'\}$ respectively. Hence it suffices to show that $\mathbb{P}(\mathfrak{X}(X_t^{\mathrm{id}}) = \mathfrak{X}(X_t^{\tau_1\circ\tau_2})) \geq \theta(c)^4$ for large n.

To ease notation, for $s \geq 0$, define $\bar{X}_s = \mathfrak{X}(X_s^{\mathrm{id}})$ and $\bar{Y}_s = \mathfrak{X}(X_s^{\tau_1\circ\tau_2})$. Let us describe the evolution of $\bar{X}$ which is a random walk on Ω_∞, see also Figure 1.8. Suppose that $\bar{X}$ makes a jump at time $s \geq 0$ and $\bar{X}_{s-} = \bar{x} = (x_1, x_2, \dots)$. We pick $u \in \{1/n, \dots, n/n\}$ uniformly and conditionally on u pick $v \in \{1/n, \dots, n/n\}\backslash\{u\}$ uniformly at random. Now imagine the interval $(0, 1]$ tiled using the intervals $(0, x_1], (0, x_2], \dots$ (the specific tiling rule does not matter). If u and v fall within different tiles, say corresponding to the intervals $(0, x_i]$ and $(0, x_j]$, then we merge the intervals $(0, x_i]$ and $(0, x_j]$ into one, and let $\mathfrak{X}(X_t)$ be the ordering of this in decreasing order. Otherwise if u and v fall within the same tile, say corresponding to the interval $(0, x_i]$, then we split $(0, x_i]$ into two intervals. Without loss of generality assume that $u < v$, then the split will result in one interval of length $v - u$ and the other of length $x_i - (v - u)$. Again we let $\bar{X}_s$ be the element of Ω_∞ resulting by ordering in decreasing order.

Fix $\delta > 0$ and set $t_0 = t - \delta^{-5}$. Let $\tau_3, \tau_4, \dots$ be a sequence of i.i.d. uniform transpositions and let $N = (N_s : s \geq 0)$ be a rate 1 Poisson process. Then we define the coupling during the interval $[0, t_0]$ by defining

$$\begin{aligned}\bar{X}_{t_0} &= \mathfrak{X}(\tau_1 \circ \dots \circ \tau_{N_{t_0}}) \\ \bar{Y}_{t_0} &= \mathfrak{X}(\tau_1 \circ \dots \circ \tau_{N_{t_0}} \circ \tau_{N_{t_0}+2})\end{aligned}$$

so that $\bar{Y}_{t_0} = \bar{X}_{t_0+2}$.

For $s \in [t_0, t]$ we create a matching between $\bar{X}_s$ and $\bar{Y}_s$ by matching an entry of $\bar{X}_s$ to exactly one entry of $\bar{Y}_s$ of the same size. Using the fact that $\bar{Y}_{t_0} = \bar{X}_{t_0+2}$, one can check that at time t_0 there are either 6, 4 or 0 unmatched entries. During the time interval $(t_0, t]$ we will match the unmatched entries so that at time t there are no more unmatched entries and thus $\bar{X}_t = \bar{Y}_t$.

We do not present the coupling for the time interval $(t_0, t]$ but note that it has the following properties:

(i) the number of unmatched entries cannot increase at any time,

(ii) the number of unmatched entries is either 6, 4 or 0,

(iii) the processes $\bar{X}$ and $\bar{Y}$ have the same jump times,

(iv) at every jump, the size of the unmatched entry decreases by at most a factor of 2,

(v) at every jump, the probability of decreasing the number of unmatched entries is at least $(x_1 x_2)^2$ where x_1 and x_2 are the smallest and the second smallest unmatched entries respectively.

Now consider the size V_s of the smallest unmatched entry at time $s \in [t_0, t]$. Suppose that $V_s \geq \delta$ for every $s \in [t_0, t]$. Then each time both $\bar{X}$ and $\bar{Y}$ jump, the probability that we decrease the number of unmatched entries bounded below by δ^4. The number jumps are governed by a rate 1 Poisson processes and with high probability, there will jump at least $(t - t_0)/2 = \delta^{-3}/2$ many times during the interval $[t_0, t]$. Thus

$$\mathbb{P}(\bar{X}_t \neq \bar{X}'_t | \{V_s \geq \delta \, \forall s \in [t_0, t]\}) \leq (1 - \delta^2)^{\delta^{-3}/2} = O(\delta).$$

Therefore using (1.23) we have that

$$\lim_{n\to\infty} \mathbb{E}[d(X_t^{\mathrm{id}}, X_t^{\tau_1 \circ \tau_2})] \approx 2(1 - \lim_{n\to\infty} \mathbb{P}(V_s \geq \delta \text{ for all } s \in [t_0, t]) - O(\delta)). \qquad (1.24)$$

Now we estimate the probability on the right hand side of (1.24). We first estimate $\mathbb{P}(V_{t_0} > \sqrt{\delta})$. Recall that $\bar{Y}_{t_0} = \bar{X}_{t_0+2}$ and let (u, v) and (u', v') be the pairs of markers used to obtain $\bar{X}_{t_0+1}$ from $\bar{X}_{t_0}$ and $\bar{X}_{t_0+2}$ from $\bar{X}_{t_0+1}$ respectively. Let $A_1(m)$ be the event that the markers (u, v) fall within the entries $\bar{X}_1(t_0), \ldots, \bar{X}_m(t_0)$. Similarly let $A_2(m)$ be the event that the markers (u', v') fall within the entries $\bar{X}_1(t_0+1), \ldots, \bar{X}_m(t_0+1)$. Then a simple argument using Theorem 1.2.11 shows that

$$\lim_{m\to\infty} \lim_{n\to\infty} \mathbb{P}(A_1(m) \cap A_2(m)) = \theta(c)^4.$$

Using Theorem 1.2.11 once more it follows that

$$\mathbb{P}(V_0 > \sqrt{\delta} | A_1(m) \cap A_2(m)) = 1 - O(\delta).$$

Hence we obtain that

$$\lim_{n\to\infty} \mathbb{P}(V_0 > \sqrt{\delta}) = \theta(c)^4(1 - O(\delta)) = \theta(c)^4 - O(\delta). \tag{1.25}$$

Using Theorem 1.2.11 and a union bound, one can show that with probability $1 - O(\delta^4)$, for every $s \in [t_0, t]$, both $\bar{X}_s$ and $\bar{Y}_s$ will contain an entry of size at least $\delta^{1/4}$. A key idea is the following: suppose an unmatched entry has size in the interval $(\delta, \delta^{1/4})$, then it follows that the unmatched entry is much more likely to merge with an entry of size at least $\delta^{1/4}$ than it is to fragment. Moreover if it does fragment, then the size of the unmatched entry decreases at most by a factor of 2. From this and a birth-death chain argument it follows that

$$\mathbb{P}(V_s \leq \delta \, \forall s \in [t_0, t] | V_{t_0} > \sqrt{\delta}) = O(\delta).$$

Combining this with (1.25) and (1.24) gives

$$\lim_{n\to\infty} \mathbb{E}[d(X_t^{\mathrm{id}}, X_t^{\tau_1 \circ \tau_2})] \approx 1 - \theta(c)^4 - O(\delta).$$

As $\delta > 0$ is arbitrary we obtain (1.22).

CHAPTER 2

Scaling Limits of Coalescent Processes Near Time Zero

Bati Şengül

2.1 Introduction

2.1.1 Statement of the main results

A coalescent process is a particle system in which particles merge into blocks. Coalescent processes have found a variety of applications in physics, chemistry and most notably in genetics where the coalescent process models ancestral relationships as time runs backwards. The work on coalescent theory dates back to the seminal paper of Kingman [33] where Kingman considered coalescent processes with pairwise mergers. In Pitman [39], Sagitov [46] and Donnelly and Kurtz [22] this was extended to the case where multiple mergers are allowed to happen. We refer to Berestycki [10] and Bertoin [14] for an overview of the field.

In this paper we shall consider Λ-coalescents where Λ is a finite strongly regularly varying measure with index $1 < \alpha \leq 2$, see (2.2). These coalescents encompass a large variety of well known examples such as beta coalescents and Kingman's coalescent. Further, these coalescents have the property that they come down from infinity, that is, when starting with infinitely many particles, the process has finitely many blocks for any time $t > 0$. Our goal is to gain precise information about the behaviour near time zero.

One central insight of this work is that the correct framework for taking such scaling limits is to view coalescent processes as geometric objects. What follows is an outline of our approach. To any coalescent process $\Pi = (\Pi(t) : t \geq 0)$, one can associate a certain ultra-metric space (E, δ) which completely characterises the process Π. This was first suggested in the work of Evans [26], who introduced this object in the case of Kingman's coalescent and studied some of its properties.

The construction is simple and we describe it now. Let $\Pi = (\Pi(t) : t \geq 0)$ be a coalescent process and define an ultra-metric on $\mathbb{N}$ by

$$\delta(i, j) := \inf\{t > 0 : i \overset{\Pi(t)}{\sim} j\} \tag{2.1}$$

where $i \overset{\Pi(t)}{\sim} j$ if and only if i, j are in the same block of $\Pi(t)$. The metric space (E, δ) is then the completion of $(\mathbb{N}, \delta)$.

Notice that $\delta(i, j)$ gives the time for the most recent common ancestor of i and j. Moreover it is not hard to check that (E, δ) is compact if and only if the coalescent process Π comes down from infinity, that is for all $t > 0$, $\Pi(t)$ has finitely many (non-empty) blocks. We call the space (E, δ) the Evans space associated to the coalescent Π.

Let Λ be a finite measure on $[0, 1]$. We say that Λ is SRV(α) if Λ is strongly regularly varying with index $\alpha \in (1, 2)$. That is when $\Lambda(dp) = f(p)\, dp$ and there exists a constant $A_\Lambda > 0$ such that

$$f(p) \sim A_\Lambda p^{1-\alpha} \qquad\qquad p \to 0 \tag{2.2}$$

where the above notation means the quotient of both sides approaches 1. We will abuse notation slightly and say that Λ is SRV(2) when $\Lambda = \delta_{\{0\}}$. It is possible to associate with each finite Λ on $[0, 1]$ a coalescent process called the Λ-coalescent and the case when Λ is SRV(2), the Λ-coalescent is Kingman's coalescent. Note that if Λ is a finite SRV(α), then the Λ-coalescent comes down from infinity if and only if $\alpha \in (1, 2]$.

Finite SRV(α) measures encompass a large variety of measures. A prominent example is the Beta$(2 - \alpha, \alpha)$ distribution which has density $B(2 - \alpha, \alpha)^{-1} p^{1-\alpha}(1 - p)^{\alpha-1}\, dp$, where $B(x, y)$ is the beta function. This is a one parameter family which interpolates between the uniform measure $(\alpha = 1)$ and δ_0 $(\alpha \to 2)$ for which the corresponding coalescents are the Bolthausen-Sznitman coalescent and Kingman's coalescent respectively. The importance of the SRV(α) condition stems primarily from population genetics where the models correspond to populations in which there is large variability in the offspring distribution, see Berestycki [10][Section 3.2].

The first theorem of the paper presented below shows convergence of the metric spaces that correspond to the coalescent processes as in (2.1). Let us briefly discuss the pointed Gromov-Hausdorff topology which we use as our notion of convergence (see Section 2.2 for the details). A pointed metric space (S, d, p) is called proper is every closed ball is compact and Polish if it is complete and seperable. A sequence of proper Polish pointed metric spaces (S_n, d_n, p_n) converges to a proper Polish pointed metric space (S, d, p) under the pointed Gromov-Hausdorff topology if for every $r > 0$, the closed ball of radius r around $p_n \in S_n$ converges in the usual Gromov-Hausdorff sense to the closed ball of radius r around $p \in S$. The space of all proper Polish pointed metric spaces can be equipped with a metric, called the pointed Gromov-Hausdorff metric, which is compatible with the notion of convergence described and further this space is itself a Polish space when equipped with the pointed Gromov-Hausdorff metric.

Theorem 2.1.1. *Let Λ be a finite measure satisfying* (2.2) *for some $\alpha \in (1, 2]$ and (E, δ) be the Evans space associated to the corresponding Λ coalescent via* (2.1)*. Then for all $i \in \mathbb{N}$, there exists a random pointed ultra-metric space $(\mathbb{S}, d_{\mathbb{S}}, o)$, which is independent of i, such that*

$$(E, \epsilon^{-1}\delta(\cdot, \cdot), i) \to (\mathbb{S}, d_{\mathbb{S}}, o)$$

weakly under the pointed Gromov-Hausdorff metric as $\epsilon \to 0$.

The limiting spaces in the theorem depend on the value of $\alpha \in (1, 2]$ and A_Λ. We will denote them by $(\mathbb{S}^{(\alpha)}, d_{\mathbb{S}}^{(\alpha)}, o^{(\alpha)})$ if there is a risk of confusion.

Geometrically, the space $(\mathbb{S}, d_{\mathbb{S}}, o)$ in Theorem 2.1.1 is referred to as tangent cone of (E, δ) at the point i. More precisely a tangent cone of a metric space (X, d) at a point $x \in X$ is given by the pointed Gromov-Hausdorff limit of $(X, r_i^{-1}d, x)$ as $i \to \infty$ where $\{r_i\}_{i \geq 1}$ is some sequence such that $r_i \downarrow 0$. Tangent cones are generalisations of tangent spaces on manifolds. Indeed, on a Riemannian manifold the tangent cone at any point exists and is isometric to the tangent space. Tangent cones have appeared in a variety of contexts ranging from geometric measure theory Simon [51] to a recent paper Curien and Le Gall [19] in which the tangent cones of the Brownian map are identified as the Brownian plane. In our case, tangent cones are the correct objects for describing the scaling limits as they allow us to forget about the mass and ordering imposed on the coalescent.

In Hughes [30] the author identifies a homeomorphism between the space of ultra-metric spaces and the space of real trees both equipped with the Gromov-Hausdorff metric. Consequently Theorem 2.1.1 can be stated in terms of the real trees that correspond to the coalescents. The tangent cones are only of interest at the leaves of a real tree as they can be easily identified at any other point as follows. If (T, d) is a coalescent tree and $x \in T$ such that $T \backslash \{x\}$ has exactly two connected components then the tangent cone $\lim_{r \downarrow 0}(T, r^{-1}d, x)$ exists and is isometric to $\mathbb{R}$ with the Euclidian distance. If $T \backslash \{x\}$ has $k \geq 3$ components then the tangent cone around x exists and is isometric to k disjoint copies of $[0, \infty)$ glued together at the point 0, equipped with the intrinsic metric.

The next result (which is both a crucial step in the proof of Theorem 2.1.1, and of

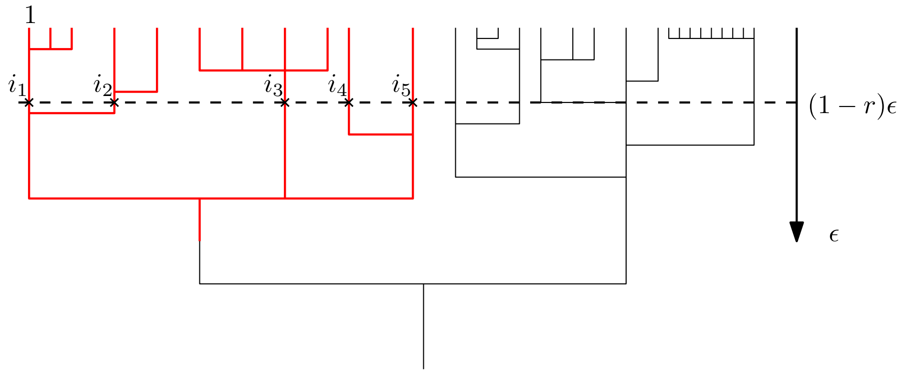


Figure 2.1: Picture illustrating the process $\mathcal{Z}_\epsilon$. The red sub-tree represents the blocks of the coalescent process which eventually merge to form $\Pi_1(\epsilon)$. Here $\mathcal{Z}_\epsilon(r) = 5$.

independent interest) provides a description of the mergers of the block containing 1 at small times. This description will allow us to depict the space $(\mathbb{S}, d_\mathbb{S})$. Loosely speaking, this result should be interpreted as a local limit of the coalescent tree, whereas Theorem 2.1.1 deals with global scaling limits. More precisely for $\epsilon > 0$ and $r \in [0,1)$ let $\mathcal{Z}_\epsilon(r)$ be the number of blocks of $\Pi((1-r)\epsilon)$ that make up $\Pi_1(\epsilon)$, the block containing 1 at time ϵ (see also Figure 2.1). Thus there exists $1 = i_1 < \cdots < i_{\mathcal{Z}_\epsilon(r)}$ such that

$$\Pi_1(\epsilon) = \Pi_{i_1}((1-r)\epsilon) \cup \cdots \cup \Pi_{i_{\mathcal{Z}_\epsilon(r)}}((1-r)\epsilon).$$

Henceforth we shall be considering the càdlàg modification of the process $\mathcal{Z}_\epsilon$.

Theorem 2.1.2. *For $\alpha \in (1,2]$ and Λ a finite SRV(α) measure let $\mathcal{Z}_\epsilon$ be the process constructed above using a Λ-coalescent. Then as $\epsilon \to 0$, $\mathcal{Z}_\epsilon \to \mathcal{Z}$ in the Skorokhod sense on $[0,1)$. The process $\mathcal{Z}$ is an inhomogeneous Markov process with generator*

$$L_r f(i) = A_\Lambda \sum_{j \geq 1} (i+j) \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} [f(i+j) - f(i)]$$

when $\alpha \in (1,2)$ and

$$L_r f(i) = \frac{(i+1)}{1-r} [f(i+1) - f(i)]$$

when $\alpha = 2$.

We will now depict how the closed unit ball $B(o,1) \subset (\mathbb{S}, d_\mathbb{S})$ is constructed. First construct a tree T from a branching process. Start the tree with one particle which does not die and is hence referred to as an immortal particle. The immortal particle produces

$j \geq 1$ offspring at time $r \in [0,1)$ with rate $(j+1)q_{j+1}^{(r)}$ where

$$q_{j+1}^{(r)} = \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)}.$$

For $j \geq 1$, the other particles at time $r \in [0,1)$ die and are replaced by $j+1$ offspring at rate $q_{j+1}^{(r)}$. Thus for each $r \in [0,1)$, the number of particles which are at distance r from the root is distributed $\mathcal{Z}(r)$. The process $\mathcal{Z}(r)$ explodes as $r \to 1$ so we have infinitely many particles at distance one from the root. The space $B(o,1)$ is the set of particles at distance one from the root and o is the immortal particle that is distance one from the root. For each $v, w \in B(o,1)$ there exists two unique paths from the root ending at the points v, w and these paths deviate at distance $h_{v,w} \geq 0$ from the root. The distance between two points is given by $d_{\mathbb{S}}(v,w) = 1 - h_{v,w}$. The multi-type branching process described is the spine decomposition of an inhomogeneous Galton-Watson process for which each particle at time $r \in [0,1)$ dies gives rise to $j+1$ offspring at rate $q_{j+1}^{(r)}$ as introduced in Chauvin and Rouault [18].

In the case $\alpha = 2$ we are able to strengthen the convergence in Theorem 2.1.1 to that of metric measure spaces and explicitly construct the limiting space (see Figure 2.2). To that end construct a measure ν on the space (E, δ) as follows. Let ν be such that the mass it assigns to each closed ball $B(i,t)$ of radius $t > 0$ around i is equal to the asymptotic frequency of the block of $\Pi(t)$ containing i. This extends uniquely to a measure on the whole space by Carathéodory's extension theorem. Our next result shows the tangent cones of the metric space (E, δ) equipped with the measure ν.

Theorem 2.1.3. *In the case when $\alpha = 2$ in Theorem 2.1.1, there exists a locally finite measure μ on the space $(\mathbb{S}, d_{\mathbb{S}})$ such that for all $i \in \mathbb{N}$,*

$$(E, \epsilon^{-1}\delta(\cdot,\cdot), 4\epsilon^{-1}\nu, i) \to (\mathbb{S}, d_{\mathbb{S}}, \mu, o)$$

weakly as $\epsilon \to 0$ under the Gromov-Hausdorff-Prokhorov topology.

The limiting metric measure space $(\mathbb{S}, d_{\mathbb{S}}, \mu, o)$ is independent from i and can be constructed as follows. Let $W = (W(t) : t \in \mathbb{R})$ be a two-sided Brownian motion on $\mathbb{R}$ and let $\mathcal{N} := \{t \in \mathbb{R} : W(t) = 0\}$. For each $x, y \in \mathcal{N}$ with $x \leq y$, define the pseudo-metric

$$d_{\mathbb{S}}(x,y) := \sup\{W(t) : t \in [x,y]\} \vee 0. \tag{2.3}$$

and $\mathbb{S} = \mathcal{N}/\sim$ where $x \sim y$ if and only if $d_{\mathbb{S}}(x,y) = 0$ and $o = 0$. The measure μ is the projection of the local time measure on $\mathcal{N}$.

We delay the exact definition of the Gromov-Hausdorff-Prokhorov metric to Section

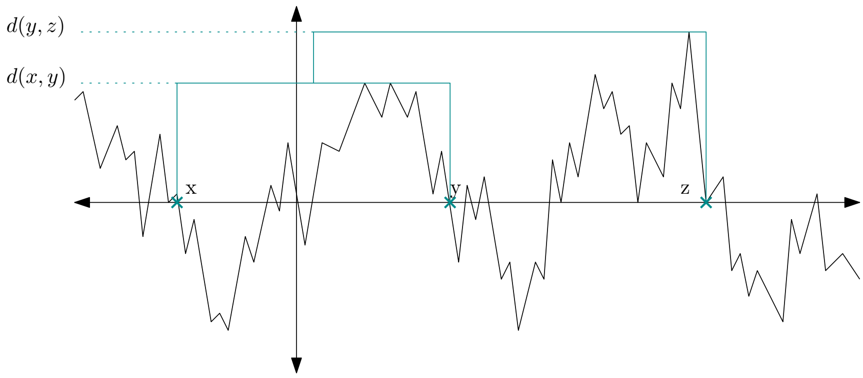


Figure 2.2: Construction of the space $(\mathbb{S}^{(2)}, d_{\mathbb{S}}^{(2)}, o^{(2)})$ from a two-sided Brownian motion.

2.2.

Note that Theorem 2.1.2 in the case $\alpha = 2$ can be obtained from Theorem 2.1.3 through some routine computations. To illustrate the usefulness of the results in the case $\alpha = 2$ we present the following corollary. This is an immediate consequence of Theorem 2.1.3.

Corollary 2.1.4. *Let $F(t)$ be the asymptotic frequency of the block containing 1 in Kingman's coalescent at time $t \geq 0$. Then we have in the sense of weak convergence on the Skorokhod space*

$$(\epsilon^{-1}F(\epsilon t) : t \geq 0) \to (X(t) : t \geq 0)$$

as $\epsilon \to 0$.

The process $X = (X(t) : t \geq 0)$ is characterised by the following.

(i) $X(0) = 0$ and for $t > 0$, $X(t)$ is the sum of two i.i.d. exponential distributions with parameter $1/(2t)$

(ii) X is an inhomogeneous compound Poisson process where at time $t > 0$ the rate of jumps is given by $2/t$ and the jump distribution is exponential with parameter $1/(2t)$.

Note that Corollary 2.1.4 extends Berestycki and Berestycki [7][Corollary 1.3] which shows the above convergence for fixed $t \geq 0$.

2.1.2 Outline of the Paper

In Section 2.2 we introduce some background on metric geometry. We assume the reader is familiar with the basic concepts in coalescent theory and excursion theory. We refer to Berestycki [10], Bertoin [14] and Revuz and Yor [41]. In Section 2.3.1 we prove Theorem

2.1.2 for $\alpha \in (1,2)$ and explain the changes needed for the case $\alpha = 2$. Then in Section 2.3.2 we prove Theorem 2.1.1. In Section 2.4 we shall prove Theorem 2.1.3.

Acknowledgements

I would like to thank my supervisor Nathanaël Berestycki for introducing me to the problem and his continuous assistance throughout. Also my thanks to James Norris and Jean Bertoin who kindly pointed out a mistake in a previous version of this paper.

2.2 Convergence of Metric Spaces

In this section we briefly review some basic notions of convergence of metric spaces. For a detailed treatment of the material refer to Burago, Burago, and Ivanov [17].

Here we introduce the Gromov-Hausdorff metric used in Theorem 2.1.1 and the Gromov-Hausdorff-Prokhorov metric in Theorem 2.1.3. We start by defining a metric on certain metric spaces without measures, called the Gromov-Hausdorff metric. We introduce this by first defining the Gromov-Hausdorff metric on compact metric spaces. Consider two compact metric spaces (X, d_X) and (Y, d_Y). The compact Gromov-Hausdorff distance $d_{cGH}((X, d_X), (Y, d_Y))$ is constructed as follows. Let (Z, d_Z) be a metric space such that there exists isometric embeddings $\phi_X : (X, d_X) \to (Z, d_Z)$, $\phi_Y : (Y, d_Y) \to (Z, d_Z)$, then

$$d_{cGH}((X, d_X), (Y, d_Y)) := \inf\{d_H^Z(\phi_X(X), \phi_Y(Y))\} \tag{2.4}$$

where the infimum is over all metric spaces (Z, d_Z) with the above property and

$$d_H^Z(A, B) := \inf\{\epsilon > 0 : B \subset \{z \in Z : \mathrm{dist}_{d_Z}(z, A) < \epsilon\} \text{ and } A \subset \{z \in Z : \mathrm{dist}_{d_Z}(z, B) < \epsilon\}\}$$

is the Hausdorff distance in (Z, d_Z). In particular, $d_{cGH}((X, d_X), (Y, d_Y)) = 0$ if and only if (X, d_X) and (Y, d_Y) are isometric.

Now we define the Gromov-Hausdorff-Prokhorov metric on compact spaces. Suppose in addition we have two finite measures μ and ν defined on the spaces (X, d_X) and (Y, d_Y) respectively. Again let (Z, d_Z) be a metric space such that there exist isometric embeddings $\phi_X : (X, d_X) \to (Z, d_Z)$, $\phi_Y : (Y, d_Y) \to (Z, d_Z)$. Then $\mu^* = \mu \circ \phi_X^{-1}$ and $\nu^* = \nu \circ \phi_Y^{-1}$ are measures on the space (Z, d_Z). The Prokhorov metric on (Z, d_Z) is given by

$$d_{Pr}^Z(\mu^*, \nu^*) := \inf\{\epsilon > 0 : \mu^*(A) \leq \nu^*(A^\epsilon) + \epsilon \text{ and } \nu^*(A) \leq \mu^*(A^\epsilon) + \epsilon \quad \forall \text{ measurable } A\}$$

where $A^\epsilon := \{z \in Z : \mathrm{dist}_{d_Z}(z, A) < \epsilon\}$. Then the compact Gromov-Hausdorff-Prokhorov distance is given by

$$d_{cGHP}((X, d_X, \mu), (Y, d_Y, \nu)) = \inf\{d_H^Z(\phi_X(X), \phi_Y(Y)) + d_{Pr}^Z(\mu^*, \nu^*)\}$$

where the infimum is over all metric spaces (Z, d_Z) with the above property.

In this paper we work with non-compact spaces and there are several ways to extend the definition above to a certain class of non-compact metric spaces. We now introduce our notion of the Gromov-Hausdorff and Gromov-Hausdorff-Prokhorov distance on non-compact metric spaces satisfying certain properties. Suppose that (X, d_X, μ, p_X) and (Y, d_Y, ν, p_Y) are proper Polish pointed metric spaces, that is they are complete, separable and every closed ball is compact. Then the (pointed) Gromov-Hausdorff distance is given by

$$d_{GH}((X, d_X, p_X), (Y, d_Y, p_Y)) = \sum_{n\geq 1} 2^{-n}(1 \wedge d_{cGH}((B(p_X, n), d_X), (B(p_Y, n), d_Y)))$$

where here and throughout $B(p, r)$ denotes the closed ball of radius r around p.

Denote by $(\mathcal{X}, d_{GH})$ the space of proper Polish spaces with a distinguished point, up to isometry, equipped with the Gromov-Hausdorff metric. The space $(\mathcal{X}, d_{GH})$ is a Polish space (see Evans [27]). Some of the properties of metric spaces are preserved under d_{GH} convergence. One such example is when the metric is an ultra-metric. A metric d is called an ultra-metric if

$$d(x, y) \leq d(x, y) \vee d(y, z) \qquad \forall x, y, z.$$

It is not hard to check all the metric spaces in this paper are in fact ultra-metric spaces and that this property is preserved under d_{GH} convergence.

Suppose that (X, d_X, μ, p_X) and (Y, d_Y, ν, p_Y) are proper Polish pointed metric spaces which come equipped with two measures μ and ν respectively. Suppose further that both measures are finite on compact sets. Then the (pointed) Gromov-Hausdorff-Prokhorov distance is given by

$$d_{GHP}((X, d_X, \mu, p_X), (Y, d_Y, \nu, p_Y)) = \sum_{n\geq 1} 2^{-n}(1 \wedge d_{cGHP}((B(p_X, n), d_X, \mu), (B(p_Y, n), d_Y, \nu))).$$

Denote by $(\mathcal{X}_\mu, d_{GHP})$ the space of proper Polish spaces with a distinguished point, up to measure preserving isometries, equipped with the Gromov-Hausdorff-Prokhorov metric. Then space $(\mathcal{X}_\mu, d_{GHP})$ is a Polish space (see for example Abraham, Delmas, and Hoscheit [1]).

It is not hard to check that the space (E, δ) is compact and the measure ν is finite. Further it can be seen that the limiting space $(\mathbb{S}, d_{\mathbb{S}})$ for $\alpha = 2$ in Theorem 2.1.3 is a proper Polish space and the measure μ is finite on compact sets.

2.3 SRV(α) Case

2.3.1 Proof of Theorem 2.1.2

Throughout this proof we omit the superscript α from the notation and assume that $\alpha \in (1, 2)$. The proof for the case $\alpha = 2$ follows analogously and we remark the only alteration to the proof that is required for $\alpha = 2$.

Let $\Pi = (\Pi(t) : t \geq 0)$ denote a Λ-coalescent such that $\Lambda(dp) = f(p)\, dp$ with $f(p) \sim A_\Lambda p^{1-\alpha}$ as $p \to 0$. Let $N = (N(t) : t > 0)$ denote the number of blocks of process Π. Recall that $\mathcal{Z}_\epsilon(r)$ is the number of blocks at time $(1 - r)\epsilon$ that make up the block containing 1 at time ϵ. We will show the convergence result by showing that $\mathcal{Z}_\epsilon(r)$ almost satisfies a certain martingale problem for small $\epsilon > 0$.

We will simplify the notation further by writing $\Pi^\epsilon(r) := \Pi((1 - r)\epsilon)$ and $N^\epsilon(r) := N((1 - r)\epsilon)$ for $r \in [0, 1)$. We will denote by $\mathcal{F}_r^\epsilon$ the natural filtration of $\mathcal{Z}_\epsilon(r)$.

Let us briefly discuss the outline of the proof. Our technique in showing the convergence $\mathcal{Z}_\epsilon \to \mathcal{Z}$ as $\epsilon \to 0$ is separated in to two: showing that $\{\mathcal{Z}_\epsilon\}_{\epsilon>0}$ is tight in Lemma 2.3.7 and showing that every subsequence of $\mathcal{Z}_\epsilon$ which converges satisfies the martingale problem for L_r, the generator of the limit $\mathcal{Z}$ given in Theorem 2.1.2, which follows from Lemma 2.3.6. To show the latter statement about the martingale problem, for $r \in [0, 1)$ and $j \in \mathbb{N}$ we need to evaluate random variables of the form $\mathbb{P}(\mathcal{Z}_\epsilon(r + \delta) - \mathcal{Z}_\epsilon(r) = j|\mathcal{F}_r^\epsilon)$ for small $\delta > 0$. Unfortunately we cannot show that $\lim_{\delta\to 0} \delta^{-1}\mathbb{P}(\mathcal{Z}_\epsilon(r + \delta) - \mathcal{Z}_\epsilon(r) = j|\mathcal{F}_r^\epsilon)$ exists and so we are forced into computing rather delicate estimates for $\mathbb{P}(A; \mathcal{Z}_\epsilon(r + \delta) - \mathcal{Z}_\epsilon(r) = j)$, where $A \in \mathcal{F}_r^\epsilon$, in Lemma 2.3.2. A key part of Lemma 2.3.2 is that the estimates for $\mathbb{P}(A; \mathcal{Z}_\epsilon(r + \delta) - \mathcal{Z}_\epsilon(r) = j)$ are uniform over all $A \in \mathcal{F}_r^\epsilon$ and $r \in [0, 1)$ in some sense.

Firstly we introduce some notation. For $\ell \leq n$, define

$$\mathcal{S}_{\ell,n} := \{\mathbf{q} = (q_j)_{j=1}^\ell : 1 \leq q_1 < \cdots < q_\ell \leq n\}$$
$$\mathcal{S}_{\ell,n}^1 := \{\mathbf{z} = (z_j)_{j=1}^\ell : 1 = z_1 < \cdots < z_\ell \leq n\}$$

and notice that $|\mathcal{S}_{\ell,n}^1| = \binom{n-1}{\ell-1}$ and $|\mathcal{S}_{\ell,n}| = \binom{n}{\ell}$.

For $\mathbf{z} \in \mathcal{S}^1_{\ell,n}$ and $r \in [0,1)$ consider the event

$$\kappa(r,\mathbf{z}) := \left\{ \Pi^\epsilon_1(0) = \bigcup_{i\in\mathbf{z}} \Pi^\epsilon_i(r) \right\} \cap \{\Pi^\epsilon_i(r) \neq \emptyset,\ \forall i \in \mathbf{z}\} \tag{2.5}$$

and note that $\kappa(r,\mathbf{z}) \in \sigma(\Pi(s) : s \in [(1-r)\epsilon, \epsilon])$. In words, this is the event that the block containing 1 at time ϵ is made up of blocks with labels given by $\mathbf{z}$ at time $(1-r)\epsilon$. In particular

$$\{\mathcal{Z}_\epsilon(r) = \ell\} \cap \{N^\epsilon(r) = n\} = \bigcup_{\mathbf{z}\in\mathcal{S}^1_{\ell,n}} \kappa(r,\mathbf{z}) \cap \{N^\epsilon(r) = n\}. \tag{2.6}$$

For the next lemma let R_n be the map which maps a partition on $\mathbb{N}$ to a partition on $[n]$ by projection.

The next lemma will allow us later to control the effects of a single jump.

Lemma 2.3.1. *For any $\epsilon > 0$, $r \in [0,1)$, $A \in \mathcal{F}^\epsilon_r$ and $j < n$, it holds that*

$$|\mathbb{P}(A|N^\epsilon(r) = n) - \mathbb{P}(A|N^\epsilon(r) = n-j)| \leq j(1-e^{-\epsilon}). \tag{2.7}$$

To illustrate the idea behind the proof consider the event $A = \{\mathcal{Z}_\epsilon(r) = 2\}$. Appealing to the Markov property we have that $\mathbb{P}(A|N^\epsilon(r) = n) = \mathbb{P}(|R_n\Pi_1(r\epsilon)| = 2)$ where $|R_n\Pi_1(r\epsilon)|$ is the size of the block of $R_n\Pi(r\epsilon)$ containing 1. Similarly for $\mathbb{P}(A|N^\epsilon(r) = n-j)$. Suppose now that $k \notin R_n\Pi_1(r\epsilon)$ for every $k \in \{n-j+1,\dots,n\}$, then it follows that $|R_n\Pi_1(r\epsilon)| = |R_{n-j}\Pi_1(r\epsilon)|$. Thus it follows that

$$\begin{aligned}
|\mathbb{P}(A|N^\epsilon(r) = n) - \mathbb{P}(A|N^\epsilon(r) = n-j)| &= \mathbb{P}(\{|R_n\Pi_1(r\epsilon)| = 2\}\Delta\{|R_{n-j}\Pi_1(r\epsilon)| = 2\}) \\
&\leq \mathbb{P}(|R_n\Pi_1(r\epsilon)| = |R_{n-j}\Pi_1(r\epsilon)|) \\
&= \mathbb{P}\left(\bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\} \right).
\end{aligned}$$

It is then easy to show that the probability in the final line is bounded by $j(1-e^{-\epsilon})$.

Proof. Fix $r \in [0,1)$, $n \in \mathbb{N}$ and $j < n$ throughout. Condition on $N^\epsilon(r) = n$, so that $\Pi^\epsilon(r) = (\Pi^\epsilon_1(r),\dots,\Pi^\epsilon_n(r))$. Let $\mathcal{A}$ denote the set of events of the form

$$\bigcup_{\mathbf{z}\in\mathcal{S}^1_{\ell,n}} \kappa(u,\mathbf{z}), \qquad u \leq r, \ell \leq n.$$

Let $\mathcal{A}'$ denote the π-system generated by $\mathcal{A}$. Note that conditioning on $N^\epsilon(r) = n$ implies that for each $\ell \leq n$ and $u \leq r$, $\{\mathcal{Z}_\epsilon(u) = \ell\} = \bigcup_{\mathbf{z}\in\mathcal{S}^1_{\ell,n}} \kappa(u,\mathbf{z})$. Hence $\mathcal{A}'$ generates the

36

σ-algebra $\mathcal{F}_r^\epsilon$. We first check (2.7) for $A \in \mathcal{A}'$. Henceforth let $A \in \mathcal{A}'$ be fixed and denote by $A_1, \dots, A_m \in \mathcal{A}$ the elements such that $A = A_1 \cap \dots \cap A_m$.

For any $u \le r$, $\ell \le n$, applying the Markov property we have

$$\mathbb{P}\left(\bigcup_{\mathbf{z} \in \mathcal{S}^1_{\ell,n}} \kappa(u, \mathbf{z}) \middle| N^\epsilon(r) = n\right)$$
$$= \mathbb{P}\left(\bigcup_{\mathbf{z} \in \mathcal{S}^1_{\ell,n}} \{R_n\Pi_1(r\epsilon) = R_n\Pi_{z_1}((r-u)\epsilon) \cup \dots \cup R_n\Pi_{z_\ell}((r-u)\epsilon)\}\right) \tag{2.8}$$

where $R_n\Pi_i(t)$ is the i-th block of $R_n\Pi$ at time $t > 0$. For each A_i, $i = 1, \dots, m$, we denote by $\tilde{A}_i^n$ the event on the right-handside of (2.8) so that $\tilde{A}_i^n \in \sigma(R_n\Pi(s) : s \le r\epsilon)$ and

$$\mathbb{P}(A_i | N^\epsilon(r) = n) = \mathbb{P}(\tilde{A}_i^n) \qquad i \le m \tag{2.9}$$

We will show that $\mathbb{P}(\cap_{i \le m} \tilde{A}_i^n \Delta \cap_{i \le m} \tilde{A}_i^{n-j}) \le j(1 - e^{-\epsilon})$, which will conclude the lemma.

For any $u \le r$, $\ell \le n$ and $\mathbf{z} \in \mathcal{S}^1_{\ell,n}$,

$$\{R_n\Pi_1(r\epsilon) = R_n\Pi_{z_1}((r-u)\epsilon) \cup \dots \cup R_n\Pi_{z_\ell}((r-u)\epsilon)\} \cap \bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\} \tag{2.10}$$
$$= \{R_{n-j}\Pi_1(r\epsilon) = R_{n-j}\Pi_{z_1}((r-u)\epsilon) \cup \dots \cup R_{n-j}\Pi_{z_\ell}((r-u)\epsilon)\} \cap \bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\}.$$

Indeed, consider the blocks at time $(r-u)\epsilon$ that coalesce by time $r\epsilon$ form the block containing 1. The only way these can differ between the restrictions to $\{1, \dots, n-j\}$ and to $\{1, \dots, n\}$ is if at least one of $\{n-j+1, \dots, n\}$ coalesced with 1 by time $r\epsilon$.

Recall that for each $i \in m$, $\tilde{A}_i^n$ and $\tilde{A}_i^{n-j}$ are given by (2.9). Similar to (2.10),

$$\bigcap_{i=1}^{m} \tilde{A}_i^n \cap \bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\} = \bigcap_{i=1}^{m} \tilde{A}_i^{n-j} \cap \bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\}.$$

The event $\bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\}$ does not depend on u, it suffices to show that this event has probability at least $1 - j(1 - e^{-\epsilon})$ as

$$\left(\bigcap_{i=1}^{m} \tilde{A}_i^n\right) \Delta \left(\bigcap_{i=1}^{m} \tilde{A}_i^{n-j}\right) \subset \left(\bigcap_{k=n-j+1}^{n} \{k \notin R_n\Pi_1(r\epsilon)\}\right)^c.$$

Now notice that$\mathbb{P}(\{k \in R_n\Pi_1(r\epsilon)\}) = 1 - e^{-r\epsilon} \leq 1 - e^{-\epsilon}$ and hence

$$\mathbb{P}\left(\bigcup_{k=n-j+1}^{n} \{k \in R_n\Pi_1(r\epsilon)\}\right) \leq j(1 - e^{-\epsilon}).$$

Thus it follows that (2.7) holds for every $A \in \mathcal{A}'$.

Now let $\mathcal{M}$ be the set of $A \in \mathcal{F}_r^\epsilon$ such that (2.7) holds. Then it is easy to see that $\mathcal{M}$ is a monotone class. Further $\mathcal{A}' \subset \mathcal{M}$ and $\mathcal{A}'$ generates $\mathcal{F}_r^\epsilon$. Hence by the monotone class theorem we have that $\mathcal{M} = \mathcal{F}_r^\epsilon$. □

Let $\lambda_{n,k}$ be the rate at which a collision involving exactly k fixed blocks occurs when there are currently n blocks present. Define

$$\gamma_{n,k} := \binom{n}{k}\lambda_{n,k}$$

which is the total rate of mergers of k blocks when n blocks are present. Let us fix $M > 1$ and define the process $\mathcal{Z}_\epsilon^M$ by

$$\mathcal{Z}_\epsilon^M(r) := \mathcal{Z}_\epsilon(r) \wedge M.$$

Lemma 2.3.2. *Let $K \subset [0, 1)$ be a compact set, then for any $j \leq M$,*

$$\sup_{r\in K}\sup_{A\in\mathcal{F}_r^\epsilon}\limsup_{\delta\to 0}\left|\frac{1}{\delta}\mathbb{P}(A; \mathcal{Z}_\epsilon^M(r+\delta) - \mathcal{Z}_\epsilon^M(r) = j) - \mathbb{E}\left[\epsilon(j + \mathcal{Z}_\epsilon^M(r))\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\mathbb{1}_{A;\mathcal{Z}_\epsilon^M(r)<M-j}\right]\right|$$
$$\leq j(1 - e^{-\epsilon})\frac{3M(M+1)}{2}\mathbb{E}\left[\sup_{r\in K}\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right].$$

Proof. Notice first that for any $u \in [0, 1)$, $k \leq n$, $\mathbf{z} \in \mathcal{S}_{k,n}^1$ and $\pi \in \mathcal{P}_\infty$ with n blocks, we have

$$\mathbb{P}(A; \kappa(u, \mathbf{z})|\Pi^\epsilon(u) = \pi) = C \qquad \forall A \in \mathcal{F}_u^\epsilon$$

for some constant $C = C(A, u, n, k)$. Indeed this follows from (2.8) and the exchangeability of the coalescent. In particular if $\#\pi$ denotes the number of blocks of π we have

$$\mathbb{P}(A; Z_\epsilon(u) = k; N^\epsilon(u) = n) = \sum_{\#\pi=n}\sum_{\mathbf{z}\in\mathcal{S}_{k,n}^1}\mathbb{P}(A; \kappa(u, \mathbf{z})|\Pi^\epsilon(u) = \pi)\mathbb{P}(\Pi^\epsilon(u) = \pi)$$
$$= \binom{n-1}{k-1}C\mathbb{P}(N^\epsilon(u) = n).$$

 38

Thus we have that for any $u \in [0,1)$, $k \le n \in \mathbb{N}$, $\mathbf{z} \in \mathcal{S}^1_{k,n}$ and $\pi \in \mathcal{P}_\infty$ with n blocks,

$$\mathbb{P}(A; \kappa(u, \mathbf{z}) | \Pi^\epsilon(u) = \pi) = \mathbb{P}(A; Z_\epsilon(u) = k | N^\epsilon(u) = n) \binom{n-1}{k-1}^{-1}. \tag{2.11}$$

Fix $n \in \mathbb{N}$, $\ell < n \wedge M$, $j \le (n-\ell) \wedge M$, $r \in [0,1)$ and $A \in \mathcal{F}^\epsilon_r$. Suppose further that $\pi \in \mathcal{P}_\infty$ with n blocks and that $\delta > 0$ is small. For $\mathbf{q} \in \mathcal{S}_{j+1,n}$ let $M_\delta(\mathbf{q})$ denote the event that there is one coalescent event in the interval $((1-r-\delta)\epsilon, (1-r)\epsilon)$ which merges the blocks of $\Pi^\epsilon(r+\delta)$ with labels $\mathbf{q}$.

Then we have that for any $\mathbf{z} \in \mathcal{S}^1_{\ell+j,n}$

$$\begin{aligned} &\mathbb{P}(A; \kappa(r+\delta, \mathbf{z}); \mathcal{Z}_\epsilon(r) = \ell; \Pi^\epsilon(r+\delta) = \pi) \\ &= \sum_{\substack{\mathbf{q} \in \mathcal{S}_{j+1,n} \\ \mathbf{q} \subset \mathbf{z}}} \mathbb{P}(A; M_\delta(\mathbf{q}); \kappa(r, \mathbf{z_q}); \Pi^\epsilon(r+\delta) = \pi) + o(\delta), \end{aligned} \tag{2.12}$$

where $\mathbf{z_q} \in \mathcal{S}^1_{\ell,n-j}$ represents the position of the indices $\mathbf{z}$ after the merger involving indices $\mathbf{q}$ has occurred. Indeed, suppose there is only one coalescent event during the interval $((1-r-\delta)\epsilon, (1-r)\epsilon)$, as more coalescent events are of $o(\delta)$. On the event $\kappa(r+\delta, \mathbf{z}) \cap \{\Pi^\epsilon(r+\delta) = \pi\}$, $\mathcal{Z}_\epsilon(r) = \ell$ if and only if during this coalescent event exactly $j+1$ blocks $\pi_{q_1}, \ldots, \pi_{q_{j+1}}$ merge with $q_i \in \mathbf{z}$ for each $i \le j+1$. We set $\mathbf{q} = \{q_1, \ldots, q_{j+1}\}$. After this merger the blocks of π with labels given by $\mathbf{z}$ now have new labels $\mathbf{z_q} \in \mathcal{S}^1_{\ell,n-j}$. Then we require that the blocks of $\Pi^\epsilon(r)$ with labels $\mathbf{z_q}$ eventually merge to give $\Pi^\epsilon_1(1)$, i.e. $\kappa(r, \mathbf{z_q})$ holds.

For $\mathbf{q} \in \mathcal{S}_{j+1,n}$ let $\pi^{(\mathbf{q})} \in \mathcal{P}_\infty$ be the partition obtained from π by merging the blocks with labels $\mathbf{q}$.

Markov property of the coalescent implies that

$$\mathbb{P}(A; \kappa(r, \mathbf{z_q}) | \Pi^\epsilon(r+\delta) = \pi; M_\delta(\mathbf{q})) = \mathbb{P}(A; \kappa(r, \mathbf{z_q}) | \Pi^\epsilon(r) = \pi^{(\mathbf{q})}),$$

and using (2.12) we have

$$\begin{aligned} &\mathbb{P}(A; \kappa(r+\delta, \mathbf{z}); \mathcal{Z}_\epsilon(r) = \ell; \Pi^\epsilon(r+\delta) = \pi) \\ &= \sum_{\substack{\mathbf{q} \in \mathcal{S}_{j+1,n} \\ \mathbf{q} \subset \mathbf{z}}} \mathbb{P}(M_\delta(\mathbf{q}) | \Pi^\epsilon(r+\delta) = \pi) \mathbb{P}(A; \kappa(r, \mathbf{z_q}) | \Pi^\epsilon(r+\delta) = \pi; M_\delta(\mathbf{q})) \mathbb{P}(\Pi^\epsilon(r+\delta) = \pi) + o(\delta) \\ &= \sum_{\substack{\mathbf{q} \in \mathcal{S}_{j+1,n} \\ \mathbf{q} \subset \mathbf{z}}} \mathbb{P}(M_\delta(\mathbf{q}) | \Pi^\epsilon(r+\delta) = \pi) \mathbb{P}(A; \kappa(r, \mathbf{z_q}) | \Pi^\epsilon(r) = \pi^{(\mathbf{q})}) \mathbb{P}(\Pi^\epsilon(r+\delta) = \pi) + o(\delta). \end{aligned} \tag{2.13}$$

To compute each term inside the sum on the last line of (2.13), notice that

$$\mathbb{P}(M_\delta(\mathbf{q})|\Pi^\epsilon(r+\delta)=\pi)=\delta\epsilon\lambda_{n,j+1}+o(\delta). \tag{2.14}$$

For the second term we can use (2.11) to deduce

$$\mathbb{P}(A;\kappa(r,\mathbf{z_q})|\Pi^\epsilon(r)=\pi^{(\mathbf{q})})=\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\binom{n-j-1}{\ell-1}^{-1}. \tag{2.15}$$

Plugging (2.14) and (2.15) into (2.13) gives

$$\begin{aligned}
&\mathbb{P}(A;\kappa(r+\delta,\mathbf{z});\mathcal{Z}_\epsilon(r)=\ell;\Pi^\epsilon(r+\delta)=\pi)\\
&\quad=\sum_{\substack{\mathbf{q}\in\mathcal{S}_{j+1,n}\\ \mathbf{q}\subset\mathbf{z}}}\delta\epsilon\lambda_{n,j+1}\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\frac{1}{\binom{n-j-1}{\ell-1}}\mathbb{P}(\Pi^\epsilon(r+\delta)=\pi)+o(\delta)\\
&\quad=\delta\epsilon\lambda_{n,j+1}\mathbb{P}(A;\mathcal{Z}_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\frac{\binom{\ell+j}{j+1}}{\binom{n-j-1}{\ell-1}}\mathbb{P}(\Pi^\epsilon(r+\delta)=\pi)+o(\delta).
\end{aligned}$$

Summing the above and using (2.6) we have

$$\begin{aligned}
&\mathbb{P}(A;Z_\epsilon(r+\delta)=\ell+j;\mathcal{Z}_\epsilon(r)=\ell;N^\epsilon(r+\delta)=n)\\
&\quad=\sum_{\#\pi=n}\sum_{\mathbf{z}\in\mathcal{S}^1_{\ell+j,n}}\mathbb{P}(A;\kappa(r+\delta,\mathbf{z});\mathcal{Z}_\epsilon(r)=\ell;\Pi^\epsilon(r+\delta)=\pi)\\
&\quad=\delta\epsilon\lambda_{n,j+1}\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\mathbb{P}(N^\epsilon(r+\delta)=n)\frac{\binom{\ell+j}{j+1}\binom{n-1}{\ell+j-1}}{\binom{n-j-1}{\ell-1}}+o(\delta)\\
&\quad=\delta\epsilon\gamma_{n,j+1}\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\mathbb{P}(N^\epsilon(r+\delta)=n)\frac{j+\ell}{n}+o(\delta)
\end{aligned}$$

where $\gamma_{n,j+1}=\binom{n}{j+1}\lambda_{n,j+1}$ and $\#\pi$ denotes the number of blocks of π.

Notice that $\mathbb{P}(N^\epsilon(r)=n)=\mathbb{P}(N^\epsilon(r+\delta)=n)+o(1)$ thus

$$\begin{aligned}
&\mathbb{P}(A;Z_\epsilon(r+\delta)=\ell+j;\mathcal{Z}_\epsilon(r)=\ell;N^\epsilon(r+\delta)=n)\\
&\qquad=\delta\epsilon\gamma_{n,j+1}\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)\mathbb{P}(N^\epsilon(r)=n)\frac{j+\ell}{n}+o(\delta).
\end{aligned} \tag{2.16}$$

On the other hand Lemma 2.3.1 gives that

$$|\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n-j)-\mathbb{P}(A;Z_\epsilon(r)=\ell|N^\epsilon(r)=n)|\le j(1-e^{-\epsilon}). \tag{2.17}$$

Now we no longer think of n and ℓ as fixed. Notice that

$$\sum_{n\geq j}\sum_{\ell\leq(n-j)\wedge M}\mathbb{P}(A;\mathcal{Z}_\epsilon(r+\delta)=\ell+j;\mathcal{Z}_\epsilon(r)=\ell;N^\epsilon(r+\delta)=n)$$
$$=\mathbb{P}(A;\mathcal{Z}_\epsilon^M(r+\delta)-\mathcal{Z}_\epsilon^M(r)=j)$$

and hence from (2.17) and (2.16),

$$\left|\frac{1}{\delta}\mathbb{P}(A;\mathcal{Z}_\epsilon^M(r+\delta)-\mathcal{Z}_\epsilon^M(r)=j)-\sum_{n\geq j}\sum_{\ell\leq(n-j)\wedge M}\epsilon\gamma_{n,j+1}\frac{j+\ell}{n}\mathbb{P}(A;\mathcal{Z}_\epsilon(r)=\ell;N^\epsilon(r)=n)\right|$$
$$\leq j(1-e^{-\epsilon})\sum_{n\geq j}\sum_{\ell\leq(n-j)\wedge M}\epsilon\gamma_{n,j+1}\frac{j+\ell}{n}\mathbb{P}(N^\epsilon(r)=n)+o(1)$$
$$\leq j(1-e^{-\epsilon})\frac{3M(M+1)}{2}\sum_{n\geq j}\epsilon\frac{\gamma_{n,j+1}}{n}\mathbb{P}(N^\epsilon(r)=n)+o(1).$$

In other words

$$\left|\frac{1}{\delta}\mathbb{P}(A;\mathcal{Z}_\epsilon(r+\delta)-\mathcal{Z}_\epsilon^M(r)=j)-\mathbb{E}\left[\epsilon(j+\mathcal{Z}_\epsilon^M(r))\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\mathbb{1}_{A;\mathcal{Z}_\epsilon^M(r)\leq M-j}\mathbb{1}_{N^\epsilon(r)\geq j}\right]\right|$$
$$\leq j(1-e^{-\epsilon})\frac{3M(M+1)}{2}\mathbb{E}\left[\epsilon(j+\mathcal{Z}_\epsilon(r))\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right]+o(1).$$

Taking limits concludes the proof. □

Now we can identify the limiting behaviour of $\delta^{-1}\mathbb{P}(\mathcal{Z}_\epsilon(r+\delta)-\mathcal{Z}_\epsilon(r)=j|\mathcal{F}_r^\epsilon)$ by using the approximation given in Lemma 2.3.2.

Lemma 2.3.3. *For any $M>1$, $j<M$ and $K\subset[0,1)$ compact,*

$$\mathbb{E}\left[\sup_{r\in K}\left|\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}-A_\Lambda\frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)}\right|\right]\to 0 \tag{2.18}$$

as $\epsilon\to 0$.

Proof. Fix $M>1$, $j<M$ and $K\subset[0,1)$ compact. Firstly from Berestycki, Berestycki, and Limic [8][Theorem 2] we have that there exists a constant $C_\alpha>0$ such that

$$\lim_{\epsilon\to 0}\mathbb{E}\left[\sup_{r\in K}\left|\frac{(1-r)\epsilon}{C_\alpha N^\epsilon(r)^{1-\alpha}}-1\right|^2\right]=0. \tag{2.19}$$

One can show (see Feller [28][XIII.6]) that this constant is given by

$$C_\alpha=\frac{\alpha}{\Gamma(2-\alpha)}.$$

Using the Cauchy-Schwartz inequality we have

$$\mathbb{E}\left[\sup_{r\in K}\left|\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)} - \frac{\alpha}{(1-r)\Gamma(2-\alpha)}\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha}\right|\right]^2 = \mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha}\left(\frac{(1-r)\epsilon\Gamma(2-\alpha)}{\alpha N^\epsilon(r)^{1-\alpha}} - 1\right)\right|\right]^2$$
$$\leq \mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha}\right|^2\right]\mathbb{E}\left[\sup_{r\in K}\left|\frac{(1-r)\epsilon\Gamma(2-\alpha)}{\alpha N^\epsilon(r)^{1-\alpha}} - 1\right|^2\right]. \tag{2.20}$$

The second term on the last line converges to zero by (2.19), thus we focus on the first term in the last line.

Recall that for $n \in \mathbb{N}$ we have that

$$\gamma_{n,j+1} = \binom{n}{j+1}\lambda_{n,j+1} = \frac{\Gamma(n+1)}{\Gamma(j+2)\Gamma(n-j)}\int_0^1 p^{j-1}(1-p)^{n-j-1}\,\Lambda(dp). \tag{2.21}$$

Moreover $\Lambda(dp) = f(p)\,dp$ where $f(p) \sim A_\Lambda p^{1-\alpha}$ as $p \to 0$. Fix $\eta_0 > 0$, then there exist a $p_0 \in (0,1)$ such that whenever $p < p_0$ we have $|f(p) - A_\Lambda p^{1-\alpha}| \leq \eta p^{1-\alpha}$. Thus

$$\left|\int_0^{p_0} p^{j-1}(1-p)^{n-j-1}\,\Lambda(dp) - A_\Lambda\int_0^{p_0} p^{j-\alpha}(1-p)^{n-j-1}\,dp\right|$$
$$\leq \eta A_\Lambda\int_0^{p_0} p^{j-\alpha}(1-p)^{n-j-1}\,dp \leq \eta A_\Lambda\int_0^1 p^{j-\alpha}(1-p)^{n-j-1}\,dp. \tag{2.22}$$

Then combining (2.21) and (2.22) we have

$$\left|\gamma_{n,j+1} - \binom{n}{j+1}A_\Lambda\int_0^1 p^{j-\alpha}(1-p)^{n-j-1}\,dp.\right|$$
$$\leq \eta\binom{n}{j+1}A_\Lambda\int_0^1 p^{j-\alpha}(1-p)^{n-j-1}\,dp + \int_{p_0}^1 p^{j-1}(1-p)^{n-j-1}|A_\Lambda p^{1-\alpha} - f(p)|\,dp$$
$$\leq \eta\binom{n}{j+1}A_\Lambda\int_0^1 p^{j-\alpha}(1-p)^{n-j-1}\,dp + (1-p_0)^{n-j-1}(A_\Lambda p_0^{1-\alpha} + \Lambda[0,1]) \tag{2.23}$$

From the definition of the Beta function we have that

$$\binom{n}{j+1}\int_0^1 p^{j-\alpha}(1-p)^{n-j-1}\,dp = \frac{\Gamma(j-\alpha+1)\Gamma(n+1)}{\Gamma(j+2)\Gamma(n-\alpha+1)}. \tag{2.24}$$

Thus using (2.23) and (2.24) we have,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha} - A_\Lambda\frac{\Gamma(j-\alpha+1)\Gamma(N^\epsilon(r)+1)}{N^\epsilon(r)^\alpha\Gamma(j+2)\Gamma(N^\epsilon(r)-\alpha+1)}\right|^2\right]\\
&\qquad\le \eta A_\Lambda\mathbb{E}\left[\sup_{r\in K}\left|\frac{\Gamma(j-\alpha+1)\Gamma(N^\epsilon(r)+1)}{N^\epsilon(r)^\alpha\Gamma(j+2)\Gamma(N^\epsilon(r)-\alpha+1)}\right|^2\right]\\
&\qquad\quad + (A_\Lambda p_0^{1-\alpha}+\Lambda[0,1])^2\mathbb{E}[\sup_{r\in K}(1-p_0)^{2(N^\epsilon(r)-j-1)}]. \qquad (2.25)
\end{aligned}$$

The final term on the right hand side converges to 0 as $\epsilon\to 0$. For the penultimate term an application of Stirling's formula yields that

$$\mathbb{E}\left[\sup_{r\in K}\left|\frac{\Gamma(j-\alpha+1)\Gamma(N^\epsilon(r)+1)}{N^\epsilon(r)^\alpha\Gamma(j+2)\Gamma(N^\epsilon(r)-\alpha+1)} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)}\right|^2\right]\to 0$$

as $\epsilon\to 0$. Thus the first term on the left hand side of (2.25) goes to zero as $\epsilon\to 0$. Moreover using the triangle inequality we have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha} - A_\Lambda\frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)}\right|^2\right]\\
&\le \mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha} - A_\Lambda\frac{\Gamma(j-\alpha+1)\Gamma(N^\epsilon(r)+1)}{N^\epsilon(r)^\alpha\Gamma(j+2)\Gamma(N^\epsilon(r)-\alpha+1)}\right|^2\right] \qquad (2.26)\\
&\quad + \mathbb{E}\left[\sup_{r\in K}\left|A_\Lambda\frac{\Gamma(j-\alpha+1)\Gamma(N^\epsilon(r)+1)}{N^\epsilon(r)^\alpha\Gamma(j+2)\Gamma(N^\epsilon(r)-\alpha+1)} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)}\right|^2\right]\\
&\to 0
\end{aligned}$$

as $\epsilon\to 0$. A final application of the triangle inequality and using (2.20) gives

$$\begin{aligned}
&\mathbb{E}\left[\sup_{r\in K}\left|\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)} - A_\Lambda\frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)}\right|\right]\\
&\le \mathbb{E}\left[\sup_{r\in K}\left|\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)} - \frac{\alpha}{(1-r)\Gamma(2-\alpha)}\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha}\right|\right] + \mathbb{E}\left[\sup_{r\in K}\left|\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)^\alpha} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)}\right|\right].
\end{aligned}$$

The proof now follows from (2.26) and (2.25).

□

Remark 2.3.4. *In the case when $\alpha = 2$ all the arguments in this section apply apart from Lemma 2.3.3. In its place we have that for $\alpha = 2$ as $\epsilon\to 0$,*

$$\mathbb{E}\left[\sup_{r\in K}\left|\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)} - \frac{1}{1-r}\mathbb{1}_{j=1}\right|\right]\to 0.$$

Indeed, this follows from the fact that for $\alpha = 2$, $\gamma_{n,2} = \binom{n}{2}$ *and* $\gamma_{n,j+1} = 0$ *for* $j > 1$, *together with Berestycki, Berestycki, and Limic [8][Theorem 2].*

The preceding two lemmas directly imply the following.

Lemma 2.3.5. *For each* $M > 1$, $j \le M$ *and* $K \subset [0,1)$, *there exists a constant* $C = C(j, M, K) > 0$ *such that*

$$\limsup_{\delta\to 0}\left|\frac{1}{\delta}\mathbb{P}(A; \mathcal{Z}_\epsilon^M(r+\delta) - \mathcal{Z}_\epsilon^M(r) = j) \; - A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)}\mathbb{E}\left[(j + \mathcal{Z}_\epsilon^M(r))\mathbb{1}_A\right]\right| \le C\epsilon$$

uniformly over all $r \in K$ *and* $A \in \mathcal{F}_r^\epsilon$.

For $f : \mathbb{N} \to \mathbb{R}$ recall that

$$L_r f(i) := A_\Lambda \sum_{j\ge 1}(i+j)\frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)}[f(i+j) - f(i)]. \tag{2.27}$$

Using the last two lemmas we are able to show that $\mathcal{Z}_\epsilon^M$ almost solves a martingale problem. This will enable us to show that the limiting process satisfies the martingale problem.

Lemma 2.3.6. *Let* $u < r \in [0,1)$ *then for any* $f : \mathbb{N} \to \mathbb{R}$ *with support in* $\{1, \ldots, \lfloor M \rfloor\}$,

$$\lim_{\epsilon\to 0}\sup_{A\in\mathcal{F}_u^\epsilon}\left|\mathbb{E}\left[\left(f(\mathcal{Z}_\epsilon^M(r)) - \int_0^r L_s f(\mathcal{Z}_\epsilon^M(s))\,ds\right)\mathbb{1}_A\right] - \mathbb{E}\left[\left(f(\mathcal{Z}_\epsilon^M(u)) - \int_0^u L_s f(\mathcal{Z}_\epsilon^M(s))\,ds\right)\mathbb{1}_A\right]\right| = 0.$$

Proof. Fix $u < r \in [0,1)$, $A \in \mathcal{F}_s^\epsilon$ and $f : \mathbb{N} \to \mathbb{R}$ with support in $\{1, \ldots, \lfloor M \rfloor\}$. Suppose that $\delta > 0$ is small. Suppose that $\{s_\ell\}_{i=0}^m$ is such that $s_0 = u$, $s_m = r$ and $s_\ell - s_{\ell-1} = \delta$ for each $\ell = 1, \ldots, m$. Now

$$\begin{aligned}\mathbb{E}[(f(\mathcal{Z}_\epsilon^M(r)) - f(\mathcal{Z}_\epsilon^M(u)))\mathbb{1}_A] &= \sum_{\ell=1}^m \mathbb{E}[(f(\mathcal{Z}_\epsilon^M(s_\ell)) - f(\mathcal{Z}_\epsilon^M(s_{\ell-1})))\mathbb{1}_A] \\ &= \sum_{\ell=1}^m\sum_{j=1}^M\sum_{i=1}^M [f(i+j) - f(i)]\mathbb{P}(\mathcal{Z}_\epsilon^M(s_\ell) - \mathcal{Z}_\epsilon^M(s_{\ell-1}) = j; \mathcal{Z}_\epsilon^M(s_{\ell-1}) = i; A).\end{aligned} \tag{2.28}$$

Now by Lemma 2.3.5, for small enough $\delta > 0$, there exists a constant $C > 0$ independent of i, j, ℓ and A (but depending on M) such that

$$\left|[f(i+j) - f(i)]\frac{\mathbb{P}(\mathcal{Z}_\epsilon^M(s_\ell) - \mathcal{Z}_\epsilon^M(s_{\ell-1}) = j; \mathcal{Z}_\epsilon^M(s_{\ell-1}) = i; A)}{\delta} - L_{s_{\ell-1}}f(i)\mathbb{P}(\mathcal{Z}_\epsilon^M(s_{\ell-1}) = i; A)\right| \le C\epsilon. \tag{2.29}$$

Suppose also that $\delta > 0$ is small enough so that

$$E\left[\left|\int_u^r L_s f(\mathcal{Z}_\epsilon^M(s))\,ds - \delta\sum_{\ell=1}^m L_{s_{\ell-1}} f(\mathcal{Z}_\epsilon^M(s_{\ell-1}))\right|\right] < \epsilon. \tag{2.30}$$

Thus from (2.28), (2.29) and (2.30) and using the fact that f is bounded,

$$\left|\mathbb{E}\left[\left(f(\mathcal{Z}_\epsilon^M(r)) - f(\mathcal{Z}_\epsilon^M(u)) - \int_u^r L_s f(\mathcal{Z}_\epsilon^M(s))\,ds\right)\mathbb{1}_A\right]\right| \le C'\epsilon$$

for some new constant $C' > 0$. The result follows by taking limits. $\square$

Next we show a tightness result.

Lemma 2.3.7. *For each $M > 1$, the sequence of processes $\{\mathcal{Z}_\epsilon^M\}_{\epsilon>0}$ is tight in the Skorokhod sense.*

Proof. Let $M > 1$ be fixed. To prove this lemma we will verify the conditions of **aldous_tightness** [Corollary 2]. Note that $\mathcal{Z}_\epsilon^M$ is uniformly bounded by M. Hence it suffices to check that for each $s \in [0,1)$, there exists a deterministic constant $\alpha(\epsilon,\delta)$ such that

$$\sup_{r\le s}\mathbb{P}(\mathcal{Z}_\epsilon^M(r+\delta) - \mathcal{Z}_\epsilon^M(r) > 0|\mathcal{F}_r^\epsilon) \le \alpha(\epsilon,\delta) \qquad \text{a.s.} \tag{2.31}$$

and

$$\lim_{\delta\to 0}\limsup_{\epsilon\to 0}\alpha(\epsilon,\delta) = 0. \tag{2.32}$$

Fix $s \in [0,1)$, let $r \le s$ and let $\delta > 0$. From (2.16) we have that for each $A \in \mathcal{F}_r^\epsilon$, $\ell, j \in \{1,\ldots,M\}$ and $n \in \mathbb{N}$,

$$\mathbb{P}(A; \mathcal{Z}_\epsilon(r+\delta) = \ell + j; \mathcal{Z}_\epsilon(r) = \ell; N^\epsilon(r+\delta) = n) \le \delta\epsilon\frac{\gamma_{n,j+1}}{n}\mathbb{P}(N^\epsilon(r) = n)2M + o(\delta).$$

Summing over $\ell, j \in \{1,\ldots,M\}$ and $n \in \mathbb{N}$ we get

$$\mathbb{P}(A; \mathcal{Z}_\epsilon^M(r+\delta) - \mathcal{Z}_\epsilon^M(r) > 0) \le 3M\delta\sum_{j=1}^M \mathbb{E}\left[\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right] + o(\delta) \le \delta C\sum_{j=1}^M \mathbb{E}\left[\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right]$$

for some constant $C > 0$.

Hence we have that (2.31) holds with

$$\alpha(\epsilon,\delta) = \delta C\sup_{r\le s}\sum_{j=1}^M \mathbb{E}\left[\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right].$$

Using Lemma 2.3.3,

$$\begin{aligned}\limsup_{\epsilon\to 0}\alpha(\epsilon,\delta) &= \delta C \limsup_{\epsilon\to 0}\sup_{r\le s}\sum_{j=1}^{M}\mathbb{E}\left[\epsilon\frac{\gamma_{N^\epsilon(r),j+1}}{N^\epsilon(r)}\right]\\ &= \delta C\sum_{j=1}^{M} A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-s)\alpha\Gamma(j+2)}.\end{aligned}$$

Taking limits as $\delta \to 0$ in the above equation we obtain (2.32). □

Now we can show the convergence of $\mathcal{Z}_\epsilon$ in the Skorokhod sense to a Markov process $\mathcal{Z} = (\mathcal{Z}(r) : r \in [0,1))$ with generator given by (2.27).

Proof of Theorem 2.1.2. We are done if we can show that for any $M > 1$ we have that

$$\mathcal{Z}_\epsilon^M \to (\mathcal{Z}(r) \wedge M : r \in [0,1))$$

in the Skorokhod sense as $\epsilon \to 0$.

Fix $M > 1$. The sequence of processes $\{\mathcal{Z}_\epsilon^M\}_{\epsilon>0}$ is tight by Lemma 2.3.7. Suppose now that for some sequence $\epsilon' \to 0$ we have that

$$\mathcal{Z}_{\epsilon'}^M \to \mathcal{Z}^M$$

in the Skorokhod sense as $\epsilon' \to 0$, to some process $\mathcal{Z}^M = (\mathcal{Z}^M(r) : r \in [0,1))$. It is enough to show that $\mathcal{Z}^M$ has the same law as $(\mathcal{Z}(r) \wedge M : r \in [0,1))$. We will show this by showing that $\mathcal{Z}^M$ satisfies a martingale problem. Let $f : \mathbb{N} \to \mathbb{R}$ have support in $\{1, \dots, \lfloor M \rfloor\}$. Then to prove the lemma, it is enough to show that

$$M_r^f := f(\mathcal{Z}^M(r)) - \int_0^r L_s f(\mathcal{Z}^M(s))\, ds \tag{2.33}$$

is a martingale. Let $u \in [0,1)$ be fixed and let $\mathbb{D}([0,u],\mathbb{N})$ denote the Skorokhod space of cádlág functions $g : [0,u] \to \mathbb{N}$. Suppose that $F : \mathbb{D}([0,u],\mathbb{N}) \to \mathbb{R}$ is a continuous and bounded function. We will show (2.33) by showing that for $u < r < 1$,

$$\mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \le u))] = \mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \le u))]. \tag{2.34}$$

Fix $u < r < 1$ and $\eta > 0$. The Skorokhod convergence implies that there exist an

$\epsilon_0 > 0$ such that

$$\left|\mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}\left[\left(f(\mathcal{Z}_{\epsilon_0}^M(r)) - \int_0^r L_s f(\mathcal{Z}_{\epsilon_0}^M(s))\, ds\right) F((\mathcal{Z}_{\epsilon_0}^M(s) : s \leq u))\right]\right| < \eta. \tag{2.35}$$

On the other hand by Lemma 2.3.6, we have that there exists an $\epsilon_1 > 0$ such that

$$\begin{aligned}\Bigg|\mathbb{E}&\left[\left(f(\mathcal{Z}_{\epsilon_1}^M(r)) - \int_0^r L_s f(\mathcal{Z}_{\epsilon_1}^M(s))\, ds\right) F((\mathcal{Z}_{\epsilon_1}^M(s) : s \leq u))\right] \\ &- \mathbb{E}\left[\left(f(\mathcal{Z}_{\epsilon_1}^M(r)) - \int_0^r L_s^{\epsilon_1} f(\mathcal{Z}_{\epsilon_1}^M(s))\, ds\right) F((\mathcal{Z}_{\epsilon_1}^M(s) : s \leq u))\right]\Bigg| < \eta.\end{aligned} \tag{2.36}$$

Applying the Skorokhod convergence once more yields that there exists an $\epsilon_2 > 0$ such that

$$\left|\mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}\left[\left(f(\mathcal{Z}_{\epsilon_2}^M(u)) - \int_0^u L_s f(\mathcal{Z}_{\epsilon_2}^M(s))\, ds\right) F((\mathcal{Z}_{\epsilon_2}^M(s) : s \leq u))\right]\right| < \eta.$$

Combining this with (2.35) and (2.36) gives that

$$|\mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \leq u))]| < 3\eta$$

As $\eta > 0$ is arbitrary this shows (2.34) which concludes the proof. □

2.3.2 Proof of Theorem 2.1.1

The proof of Theorem 2.1.1 follows from Theorem 2.1.2 in a straightforward manner. We explain the main idea. Here and throughout we let $B_\epsilon(i, r)$ denote the closed ball of radius r around i in the space $(E, \epsilon^{-1}\delta)$. Then to show the theorem it suffices to show that $(B_\epsilon(1, 1), \epsilon^{-1}\delta)$ converges weakly. Indeed we may assume that $i = 1$ by exchangeability of the coalescent and the proof for general $r > 0$ follows with more cumbersome notation.

Recall the definition of the process $\mathcal{Z}_\epsilon$ from Theorem 2.1.2. As any two balls in an ultra-metric space are either disjoint or one contains the other, for each $r \in (0, 1)$, the space $(B_\epsilon(1, 1), \epsilon^{-1}\delta)$ can be covered uniquely by disjoint closed balls $B_1, \ldots, B_n \subset (B_\epsilon(1, 1), \epsilon^{-1}\delta)$ of radius $1 - r$. The number of such balls is precisely the number of blocks of Π at time $(1 - r)\epsilon$ that make up the block containing 1 at time ϵ. Thus

$$\mathcal{Z}_\epsilon(r) = \#\{\text{disjoint closed balls of radius } 1 - r \text{ needed to cover } B_\epsilon(1, 1)\}. \tag{2.37}$$

Furthermore we shall see we can discover the exact structure of $(B_\epsilon(1, 1), \epsilon^{-1}\delta)$ using the process $\mathcal{Z}_\epsilon$. Fix $\eta \in (0, 1)$ and let $B_1, \ldots, B_{\mathcal{Z}_\epsilon(1-\eta)}$ denote the disjoint closed balls of radius η that cover the space $(B_\epsilon(1,1), \epsilon^{-1}\delta)$. Now define a metric $r_\epsilon^{(\eta)}$ on $S_\epsilon^{(\eta)} := \{1, \dots, \mathcal{Z}_\epsilon(1-\eta)\}$ by

$$r_\epsilon^{(\eta)}(i,j) = \mathrm{dist}(B_i, B_j). \tag{2.38}$$

Notice that the ordering of the balls $B_1, \dots, B_{\mathcal{Z}_\epsilon(1-\eta)}$ do not matter in the sense that the spaces constructed from two different orderings are isometric. Henceforth suppose that $1 \in B_1$. We will first show the convergence of the metric space $(S_\epsilon^{(\eta)}, r_\epsilon^{(\eta)})$ as $\epsilon \to 0$.

Given the process $\mathcal{Z}_\epsilon$ we can construct the space $(S_\epsilon^{(\eta)}, r_\epsilon^{(\eta)})$ as follows. First construct a tree T from a branching process. Start the tree with one immortal particle which will not die. We call all other particles mortal. Suppose now that for some $i \in \mathbb{N}$ and $j \geq 1$, $\mathcal{Z}_\epsilon(r-) = i$ and $\mathcal{Z}_\epsilon(r+) = i + j$. Then there is a birth at height r of the tree. With probability $(j+1)/(i+j)$, the immortal particle gives birth to j offspring and with probability $(i-1)/(i+j)$ a uniformly chosen mortal individual dies gives birth to $j+1$ offspring. Thus the tree has exactly $\mathcal{Z}_\epsilon(1-\eta)$ leaves and height $1-\eta$. These leaves form the space $S_\epsilon^{(\eta)}$ and the distance r between two leaves is the genealogical distance i.e. half of the length of the unique path between the two leaves. Use the same procedure but with the process $\mathcal{Z}$ to obtain a space $(S^{(\eta)}, r^{(\eta)})$. It is clear that $(S^{(\eta)}, r^{(\eta)})$ is a compact metric space.

Using Theorem 2.1.2 and Skorokhod's representation theorem suppose henceforth that $\mathcal{Z}_\epsilon \to \mathcal{Z}$ as $\epsilon \to 0$ almost surely under the Skorokhod topology. Let $J_1^\epsilon, \dots, J_n^\epsilon$ be the jumps of the process $\mathcal{Z}_\epsilon$ before time $1-\eta$ and similarly let $J_1, \dots, J_m$ be the jumps of the process $\mathcal{Z}$ before time $1-\eta$. Then from the almost sure convergence of the process $\mathcal{Z}_\epsilon$ to $\mathcal{Z}$ we can conclude the following for $\epsilon > 0$ small enough. Firstly $m = n$ and $Z_\epsilon(J_i^\epsilon) = Z(J_i)$ for each $i \leq n = m$. Second, $\max_{i \leq n} |J_i^\epsilon - J_i|$ is small.

Thus for $\epsilon > 0$ small enough this gives a coupling between the spaces $(S_\epsilon^{(\eta)}, r_\epsilon^{(\eta)})$ and $(S^{(\eta)}, r^{(\eta)})$ such that $S^{(\eta)} = S_\epsilon^{(\eta)}$ and further

$$\max_{i,j \in S^{(\eta)}} |r_\epsilon^{(\eta)}(i,j) - r^{(\eta)}(i,j)| \leq 2\mathcal{Z}_\epsilon(1-\eta) \max_{i \leq n} |J_i^\epsilon - J_i|$$

which is small. Hence $(S_\epsilon^{(\eta)}, r_\epsilon^{(\eta)}) \to (S^{(\eta)}, r^{(\eta)})$ almost surely under the compact Gromov-Hausdorff topology as $\epsilon \to 0$.

Notice that $\{(S^{(\eta)}, r^{(\eta)})\}_{\eta \in (0,1)}$ is a Cauchy sequence and by completeness we have that $((S^{(\eta)}, r^{(\eta)}) \to (S, r)$ to some compact space (S, r) almost surely under the compact Gromov-Hausdorff topology as $\eta \downarrow 0$. On the other hand we have that the compact Gromov-Hausdorff distance between $(S_\epsilon^{(\eta)}, r_\epsilon^{(\eta)})$ and $(B_\epsilon(1,1), \epsilon^{-1}\delta)$ is a most η. Thus it follows that $(B_\epsilon(1,1), \epsilon^{-1}\delta) \to (S, r)$ almost surely under the compact Gromov-Hausdorff topology as $\epsilon \to 0$, which finishes the proof.

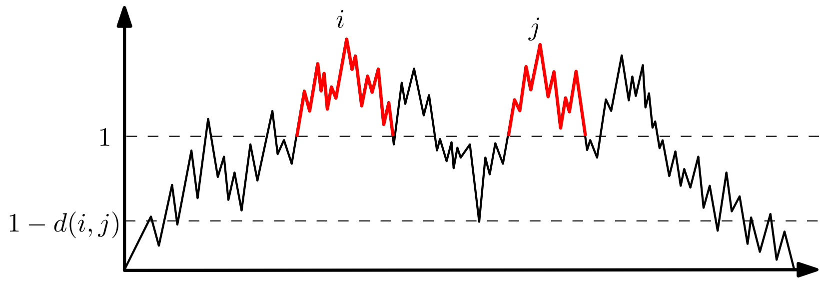


Figure 2.3: Visual interpretation of the construction of the metric space (S, d).

2.4 Kingman Case

In this section we will prove Theorem 2.1.3. As before we write $B_\epsilon(i, r)$ to mean the closed ball of radius r around i in the space $(E, \epsilon^{-1}\delta)$. Recall the construction of the space $(\mathbb{S}, d_\mathbb{S}, \mu, 0)$ given in the statement of Theorem 2.1.3. Again we will only show that

$$(B_\epsilon(1,1), \epsilon^{-1}\delta, 4\epsilon^{-1}\nu) \to (B(0,1), d_\mathbb{S}, \mu)$$

weakly under the compact Gromov-Hausdorff metric as $\epsilon \to 0$, where $B(0,1) \subset (\mathbb{S}, d_\mathbb{S})$ is the closed ball of radius 1 around 0.

We first show how to metric measure spaces using excursions. We term this the Evans metric space associated to an excursion due to the similarities of the Evans metric space associated to coalescent processes. We describe this process in generality and then use it to construct the spaces $(B_\epsilon(1,1), \epsilon^{-1}\delta, 4\epsilon^{-1}\nu)$, $(B(0,1), d_\mathbb{S}, \mu)$ as well as an auxiliary space.

Constructing Evans metric measure space from an excursion

Let $f = (f(t) : 0 \le t \le \zeta(f))$ be an excursion that has height greater than 1 meaning $f : [0, \zeta(f)] \to [0, \infty)$ is a continuous path such that f hits 1 and further $f(t) = 0$ if and only if $t \in \{0, \zeta(f)\}$. For $t \in [0, \zeta(f)]$ and $x \in \mathbb{R}$ define the local time $L(t, x)$ at level x at time t by

$$L(t,x) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t \mathbb{1}_{f(s) \in (x-\epsilon, x+\epsilon)\}} ds.$$

We will suppose that such limit exists and moreover that $(L(t,x) : t \in [0, \zeta(f)], x \in \mathbb{R})$ is jointly continuous. In fact in all the application we shall consider, the excursion f will be Brownian and we refer to Revuz and Yor [41][Chapter VI] for definition and basic properties of local times in this case. We now show how to obtain a metric measure space (S, d, π) from the excursion f. Let $\{\epsilon_i\}_{i=1}^\infty$ be the positive excursions of the excursion f

above level 1 and we order them as follows. Let $U_1, U_2 \dots$ be i.i.d. random variables with the law given by the normalised local time at level 1, $dL(\cdot, 1)/Z_1$ where $Z_1 = L(\zeta(f), 1)$ is the total local time spent at level 1. For each $k \in \mathbb{N}$, U_k is the local time corresponding to a unique excursion at level 1 which may be positive or negative. Let $U_{k_1}, U_{k_2}, \dots$ denote the local times corresponding to positive excursions, then $\{\epsilon_i\}_{i=1}^{\infty}$ is ordered such that for each $i \in \mathbb{N}$, ϵ_i is the excursion which starts at local time U_{k_i}.

For $i, j \in \mathbb{N}$ with $i \neq j$ we define $1 - d(i,j)$ to be the first height at which ϵ_i and ϵ_j are a part of the same excursion (see Figure 2.3). In other words let $t(\epsilon_i)$ and $t(\epsilon_j)$ denote the start time of the excursions ϵ_i and ϵ_j and suppose without a loss of generality that $t(\epsilon_i) < t(\epsilon_j)$. Then

$$d(i,j) = 1 - \inf\{f(t) : t(\epsilon_i) \leq t \leq t(\epsilon_j)\}. \tag{2.39}$$

By definition, the space (S, d) is the completion of $(\mathbb{N}, d)$. We also define a measure π on (S, d) as follows. For each $i \in \mathbb{N}$ and $r \in (0, 1]$, every closed ball $B(i, r) \subset (S, d)$ corresponds to an excursion e of f above level $1 - r$ that hits level 1. We define $\pi(B(i,r))$ to be the total local time $\ell_1(e)$ the excursion e attains at level 1. Note that this defines the measure uniquely by Carathédeory's extension theorem and thus we obtain a metric measure space (S, d, π). We remark that the total mass $\pi(S) = Z_1$ is the total local time spent at level 1 by the excursion f and thus is finite. Lastly in all of our applications the excursion f will have unique local minima which is enough to conclude that (S, d) is compact.

Definition 2.4.1. *The metric measure space (S, d, π) is called the Evans metric measure space associated to the excursion f.*

Remark 2.4.2. *For the reader who is familiar with continuum real trees, the Evans metric measure space associated to the excursion f can be thought of as follows. Let (T, d_T) be the rooted real tree that is encoded by the excursion f. Delete every branch of T which fails to reach distance 1 from the root. Next, delete every point which is of distance greater than 1 from the root. Let $(\tilde{T}, d_{\tilde{T}})$ denote the rooted real tree which is the result of T after these operations. Then the space S is the set of points which is distance 1 from the root of $\tilde{T}$. The distance d is given by $(1/2)d_{\tilde{T}}$ and the measure π is the uniform measure on $\tilde{T}$.*

We now use this construction to give an alternative construction to the limiting space $(B(0,1), d_{\mathbb{S}}, \mu)$. Let $W = (W_t : t \in \mathbb{R})$ be a two-sided Brownian motion and let $Y = (Y_t : t \in [0, \zeta(Y)])$ be the excursion of W above level -1 straddling the origin. That is, let $\tau_+ = \inf\{t > 0 : W_t = -1\}$ and $\tau_- = \sup\{t < 0 : W_t = -1\}$, then $Y_t = W_{t+\tau_-} + 1$ for $t \leq \tau_+ - \tau_-$. It is not hard to check that the space $(B(0,1), d_{\mathbb{S}}, \mu)$ can be constructed as the Evans metric measure space associated to the excursion Y.

Next we recall an alternative construction of the Evans space (E, δ, ν) associated to Kingman's coalescent. Let $X = (X_t : 0 \leq t \leq \zeta(X))$ be a Brownian excursion conditioned to hit level 1. Let $(\tilde{E}, \tilde{\delta}, \tilde{\nu})$ be the Evans metric measure space associated to the excursion X. For $x \geq 0$ let Z_x be the total local time attained at level $x \geq 0$ by the process X. For $t \in [0, 1]$ define

$$V(t) := \int_{1-t}^{1} \frac{4}{Z_v} dv. \tag{2.40}$$

Then from Berestycki and Berestycki [7][Theorem 1.1] we can construct the space (E, δ, ν) as follows. Firstly $E = \tilde{E}$. Next for $x, y \in E$ we set $\delta(x, y) = V(\tilde{\delta}(x, y))$. Lastly we let ν be the renormalisation of $\tilde{\nu}$, that is $\nu(\cdot) = \tilde{\nu}(\cdot)/Z_1$ where Z_1 is the total local time that the excursion X attains at level 1.

Define

$$T_\epsilon := \frac{4}{\epsilon Z_{1-\sqrt{\epsilon}}} \vee \frac{1}{\sqrt{\epsilon}} \tag{2.41}$$

and let $\tilde{B}_\epsilon(1, 1) \subset (\tilde{E}, T_\epsilon \tilde{\delta})$ be the closed ball of radius 1 around 1. Our proof will go by showing that the spaces $(B(0, 1), d_{\mathbb{S}}, \mu)$ and $(\tilde{B}_\epsilon(1, 1), T_\epsilon \tilde{\delta}, T_\epsilon \tilde{\nu})$ are close (see Lemma 2.4.5). Then we will see in (2.49) that the spaces $(\tilde{B}_\epsilon(1, 1), T_\epsilon \tilde{\delta}, T_\epsilon \tilde{\nu})$ and $(B_\epsilon(1, 1), \epsilon^{-1}\delta, 4\epsilon^{-1}\nu)$ are close.

We start by presenting an alternative construction of the space $(\tilde{B}_\epsilon(1, 1), T_\epsilon \tilde{\delta}, T_\epsilon \tilde{\nu})$. Define

$$X_\epsilon(t) := 1 + [X(tT_\epsilon^{-2}) - 1]T_\epsilon \qquad 0 \leq t \leq \zeta(X)T_\epsilon^2. \tag{2.42}$$

The Evans metric measure space associated with the excursion $(X_\epsilon(t) + T_\epsilon - 1 : 0 \leq t \leq \zeta(X)T_\epsilon^2)$ is precisely $(\tilde{E}, T_\epsilon \tilde{\delta}, T_\epsilon \tilde{\nu})$. We wish to construct the subspace $(\tilde{B}_\epsilon(1, 1), T_\epsilon \tilde{\delta}, T_\epsilon \tilde{\nu})$ directly from an excursion Y_ϵ of X_ϵ which we describe now.

Let $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$ be the excursions of X_ϵ above level 0 that reach level 1 where $M(\epsilon)$ is the total number of such excursions. Though we have obtained the process X_ϵ by performing diffusive scaling on X, the scaling factor T_ϵ is random so it is not obvious that $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$ are themselves Brownian excursions conditioned to reach level 1. We will see that this is nevertheless the case in Lemma 2.4.3.

Given $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$, the excursion Y_ϵ of X that we are after is selected from $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$ where we select biased on how much local time at 1 each of the excursions $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$ accumulate. Precisely, define Y_ϵ as follows

$$\mathbb{P}(Y_\epsilon = e_i^{(\epsilon)} | e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}) = \frac{\ell_1(e_i^{(\epsilon)})}{\sum_{j=1}^{M(\epsilon)} \ell_1(e_j^{(\epsilon)})} \qquad 1 \leq i \leq M(\epsilon) \tag{2.43}$$

where $\ell_1(e_i^{(\epsilon)})$ is the total local time at 1 that the excursion $e_i^{(\epsilon)}$ accumulates. Then the space $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu})$ can be constructed as the Evans metric measure space associated to the excursion Y_ϵ.

We will use this to show that the limiting space $(B(0,1), d_{\mathbb{S}}, \mu)$ and the space $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu})$ are close. We begin with the following lemma.

Lemma 2.4.3. *Let $e_1^{(\epsilon)}, \ldots, e_{M(\epsilon)}^{(\epsilon)}$ be the excursions of X_ϵ above level 0 that reach level 1. Then conditionally on $M(\epsilon)$, $e_1^{(\epsilon)}, \ldots, e_{M(\epsilon)}^{(\epsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1.*

Proof. Fix $\epsilon > 0$. Observe that the excursions $e_1^{(\epsilon)}, \ldots, e_{M(\epsilon)}^{(\epsilon)}$ correspond to excursions of X above level $u := 1 - 1/T_\epsilon$ that hit level 1. Note that since $T_\epsilon \geq 1/\sqrt{\epsilon}$ we have that $u \geq u_0 := 1 - \sqrt{\epsilon}$. Define the σ-algebra $\mathcal{H} = \sigma(X_{\alpha(s)} : s \geq 0)$ where

$$\alpha(s) := \inf\left\{t \geq 0 : \int_0^t \mathbb{1}_{\{X(v) \leq u_0\}} dv > s\right\}.$$

In words $\mathcal{H}$ contains all the information about the excursions of X below level $u_0 = 1 - \sqrt{\epsilon}$. The total local time Z_{u_0} of the process X at level u_0 satisfies (see Revuz and Yor [41][Chapter VI, Corollary (1.9)])

$$Z_{u_0} = \lim_{\eta \to 0} \frac{1}{\eta} \int_0^{\zeta(X)} \mathbb{1}_{\{X_s \in (u_0 - \eta, u_0]\}} ds.$$

Thus Z_{u_0} is measurable with respect to $\mathcal{H}$ and consequently so is T_ϵ and $u = 1 - 1/T_\epsilon$.

It is well known that after hitting level u_0, the law of X is that of a Brownian motion started at level u_0, killed the first time it hits 0 and conditioned to reach level 1 before hitting level 0. Itô's description of Brownian motion (Revuz and Yor [41][Chapter XII, Theorem (2.4)]) tells us that conditionally on $Z_{u_0} = z$ the excursions of the process X above level u_0 form a Poisson point process on the local time interval $[0, z]$, conditioned to have at least one excursion of height grater than $\sqrt{\epsilon}$. Further, the excursions above level $1 - \sqrt{\epsilon}$ are independent of the excursions below level $1 - \sqrt{\epsilon}$, and hence independent of the σ-algebra $\mathcal{H}$.

On the other hand Z_{u_0} and u are measurable with respect to $\mathcal{H}$. Thus conditionally on $\mathcal{H}$ the excursions of the process X above level $u = 1 - 1/T_\epsilon$ that hit level 1 are i.i.d. Brownian excursions conditioned to have height greater than $1/T_\epsilon$. By Brownian scaling and (2.42) it follows that conditionally on $\mathcal{H}$ and $M(\epsilon) = m$, $e_1^{(\epsilon)}, \ldots, e_m^{(\epsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1. In other words let e be a Brownian excursion conditioned to reach level 1 and $\phi_1, \ldots, \phi_m$ be continuous bounded functions mapping the set of excursions to $\mathbb{R}$, then we have just shown that

$$\mathbb{E}[\phi_1(e_1^{(\epsilon)})\dots\phi_m(e_m^{(\epsilon)})|\mathcal{H}, M(\epsilon)=m] = \prod_{i=1}^m \mathbb{E}[\phi_i(e)].$$

Taking expectations conditionally on $M(\epsilon)=m$ on both sides above finishes the proof. □

The next lemma shows the convergence of the local time at 1 of Y_ϵ. Recall that Y is the excursion of two-sided Brownian motion straddling the origin which is used in the construction of the space $(B(0,1), d_{\mathbb{S}}, \mu)$.

Lemma 2.4.4. *We have that*

$$\ell_1(Y_\epsilon) \to \ell_1(Y)$$

in distribution as $\epsilon \to 0$, where $\ell_1(\cdot)$ denotes the total local time attained by the excursion at level 1.

Proof. Let $e_1^{(\epsilon)}, \dots, e_{M(\epsilon)}^{(\epsilon)}$ be the excursions of X_ϵ above level 0 that reach level 1. Condition on $M(\epsilon)=m$. Then from Lemma 2.4.3 it follows that $e_1^{(\epsilon)}, \dots, e_m^{(\epsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1. Thus it follows that $\ell_1(e_1^{(\epsilon)}), \dots, \ell_1(e_m^{(\epsilon)})$ are i.i.d. exponential random variables with parameter 1/2 (see Revuz and Yor [41][Chapter VI, Proposition (4.6)]). Let $E_1, E_2, \dots$ be i.i.d. exponential random variables with parameter 1/2, then by (2.43) it follows that

$$\mathbb{P}(\ell_1(Y_\epsilon) \in \cdot | M(\epsilon)=m) = \mathbb{E}\left[\frac{mE_1}{\sum_{i=1}^m E_i} \mathbb{1}_{\{E_1 \in \cdot\}}\right].$$

On the other hand $\ell_1(Y)$ has the same law as E_1+E_2, the sum of two independent exponential random variables. This is a sized biased exponential random variable and by bounded convergence and the law of large numbers we have

$$\mathbb{P}(\ell_1(Y) \in \cdot) = \lim_{n\to\infty} \mathbb{E}\left[\frac{nE_1}{\sum_{i=1}^n E_i} \mathbb{1}_{\{E_1 \in \cdot\}}\right].$$

The lemma now follows from the fact that $M(\epsilon) \to \infty$ in probability as $\epsilon \to 0$. □

From Lemma 2.4.3 for Y_ϵ and the definition of Y we can deduce the following. Conditionally on $\ell_1(Y_\epsilon) = \ell_\epsilon$ and $\ell_1(Y) = \ell$, the excursions Y_ϵ and Y are both Brownian excursions conditioned on attaining total local time ℓ_ϵ and ℓ, respectively, at level 1. Thus at this point it is immediate that $Y_\epsilon \to Y$ in the Skorokhod sense as $\epsilon \to 0$. Unfortunately, this is not enough to show the weak convergence of the space $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu})$, under the Gromov-Hausdorff-Prokhorov metric, to the space $(B(0,1), d_{\mathbb{S}}, \mu)$. Instead, in

 53

the next lemma we construct a coupling between Y_ϵ and Y under which the paths of the two processes agree up to a time. This will in turn enable us to show that the spaces $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu})$ and $(B(0,1), d_{\mathbb{S}}, \mu)$ are close to each other.

Lemma 2.4.5. *We have that*

$$(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu}) \to (B(0,1), d_{\mathbb{S}}, \mu)$$

weakly as $\epsilon \to 0$ under the Gromov-Hausdorff-Prokhorov topology.

Proof. We first present a coupling between Y_ϵ and Y. By Lemma 2.4.4 and Skorokhod representation theorem we can suppose that $\ell_1(Y_\epsilon)$ and $\ell_1(Y)$ are coupled such that $\ell_1(Y_\epsilon) \to \ell_1(Y)$ almost surely as $\epsilon \to 0$. Fix $\epsilon > 0$ and condition on $\ell_1(Y_\epsilon) = \ell_\epsilon$ and $\ell_1(Y) = \ell$. Suppose further that $\ell_\epsilon \leq \ell$ (the other case is similar). Note that both Y_ϵ and Y are Brownian excursions conditioned to have ℓ_ϵ and ℓ total local time at level 1. Hence we can couple Y_ϵ and Y such that they have the same path until their local time at level 1 reaches ℓ_ϵ.

Excursions of the process Y_ϵ above level 1 correspond to the dense subset $\mathbb{N}$ of of the space $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta})$. Similarly for Y and $(B(0,1), d_{\mathbb{S}})$. Thus the coupling of the processes Y_ϵ and Y gives us a coupling of the spaces such that $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}) \subset (B(0,1), d_{\mathbb{S}})$. Under this coupling of the spaces it is immediate that

$$d_{Pr}(T_\epsilon\tilde{\nu}, \mu) = |\ell - \ell_\epsilon|. \tag{2.44}$$

where d_{Pr} denotes the Prokhorov distance.

Fix $\eta \in (0,1)$ and recall that $d_H((\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}), (B(0,1), d_{\mathbb{S}})) < \eta$ if and only if the η-enlargement of the space $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta})$ contains $(B(0,1), d_{\mathbb{S}})$. Clearly it suffices to show the latter condition for the dense subspaces of $(\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta})$ and $(B(0,1), d_{\mathbb{S}})$ corresponding to the excursions above level 1 of the processes Y_ϵ and Y respectively. Consider the excursions of Y above level 1 which appear before local time ℓ_ϵ. Call these excursions matched; these are also excursions of Y_ϵ above level 1. Then $d_H((\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}), (B(0,1), d_{\mathbb{S}})) < \eta$ if and only if every excursion of Y above level $1-\eta$ that hits level 1 contains a matched excursion. This is the same as the event that every excursion of Y below 1 with local time in the interval $[\ell_\epsilon, \ell)$ has infimum greater than $-\eta$. Thus from standard excursion theory (see Revuz and Yor [41][Chapter XII, Exercise (2.10)]) we have that

$$\mathbb{P}(d_H((\bar{B}_{T_\epsilon}(1,1), T_\epsilon\tilde{\delta}), (\bar{B}(0,1), d_{\mathbb{S}})) \leq \eta) = e^{-\frac{|\ell-\ell_\epsilon|}{\eta}}. \tag{2.45}$$

The equations (2.44) and (2.45) hold by the same argument when $\ell < \ell_\epsilon$. Hence in conclusion we have constructed a coupling where

$$\begin{aligned}
&\mathbb{P}(d_H((\tilde{B}_\epsilon(1,1),T_\epsilon\tilde{\delta}),(B(0,1),d_{\mathbb{S}}))+d_{Pr}(T_\epsilon\tilde{\nu},\mu)>2\eta)\\
&\leq \mathbb{P}(d_H((\tilde{B}_\epsilon(1,1),T_\epsilon\tilde{\delta}),(B(0,1),d_{\mathbb{S}}))>\eta)+\mathbb{P}(d_{Pr}(T_\epsilon\tilde{\nu},\mu)>\eta)\\
&=1-\mathbb{E}\left[\exp\left(-\frac{|\ell_1(Y_\epsilon)-\ell_1(Y)|}{\eta}\right)\right]+\mathbb{P}(|\ell_1(Y_\epsilon)-\ell_1(Y)|>\eta).
\end{aligned}$$

Taking the limit as $\epsilon\to 0$ above and using bounded convergence finishes the proof. □

Using the previous lemma we can now prove Theorem 2.1.3.

Proof of Theorem 2.1.3. For $t\in[0,1]$ let $U(t)$ denote the inverse of $V(t)$ in (2.40), that is

$$U(t)=V^{-1}(t)=\inf\left\{s>0:\int_{1-s}^{1}\frac{4}{Z_v}\,dv>t\right\}.$$

Fix $\eta\in(0,1)$ and define

$$\begin{aligned}
A_\epsilon^\eta &:= \{(1-\eta)Z_{1-\sqrt{\epsilon}}\leq Z_s\leq(1+\eta)Z_{1-\sqrt{\epsilon}}\,\forall s\in[1-U(\epsilon),1]\}\\
B_\epsilon &:= \left\{T_\epsilon=\frac{4}{\epsilon Z_{1-\sqrt{\epsilon}}}\right\}=\left\{Z_{1-\sqrt{\epsilon}}\leq\epsilon^{-1/2}\right\}\\
\mathcal{E}_\epsilon^\eta &:= A_\epsilon^\eta\cap B_\epsilon.
\end{aligned}$$

We claim that $\mathbb{P}(\mathcal{E}_\epsilon^\eta)\to 1$ as $\epsilon\to 0$. Indeed, $x\mapsto Z_x$ is uniformly continuous (this follows from the Ray-Knight theorems, see Revuz and Yor [41][Chapter XI, Theorem (2.2)]). Further it is elementary to check that $\lim_{\epsilon\to 0}U(\epsilon)=0$. This shows that $\mathbb{P}(A_\epsilon^\eta)\to 1$ as $\epsilon\to 0$. The convergence of $\mathbb{P}(B_\epsilon)$ follows from the fact that $Z_{1-\sqrt{\epsilon}}$ is distributed exponentially with parameter $1/(2-2\sqrt{\epsilon})$ (see Revuz and Yor [41][Chapter VI, Proposition (4.6)]). Thus $\mathbb{P}(\mathcal{E}_\epsilon^\eta)\to 1$ as $\epsilon\to 0$.

On the event $\mathcal{E}_\epsilon^\eta$ we have that for each $t\leq\epsilon$,

$$\frac{U(t)\epsilon T_\epsilon}{1+\eta}=\frac{4U(t)}{(1+\eta)Z_{1-\sqrt{\epsilon}}}\leq t\leq\frac{4U(t)}{(1-\eta)Z_{1-\sqrt{\epsilon}}}=\frac{U(t)\epsilon T_\epsilon}{1-\eta}. \tag{2.46}$$

Recall that $E=\tilde{E}$ and that for each $x,y\in E$, $\tilde{\delta}(x,y)=U(\delta(x,y))$. Hence from (2.46) it follows that on the event $\mathcal{E}_\epsilon^\eta$ for any $x,y\in E$ such that $\delta(x,y)\leq\epsilon$ we have

$$\frac{1}{1+\eta}T_\epsilon\tilde{\delta}(x,y)\leq\epsilon^{-1}\delta(x,y)\leq\frac{1}{1-\eta}T_\epsilon\tilde{\delta}(x,y). \tag{2.47}$$

55

A brief computation shows that on the event $\mathcal{E}_\epsilon^\eta$, for each $x \in E$ and $r \in (0,1]$

$$4\epsilon^{-1}\nu(B_\epsilon(x,r)) = \frac{T_\epsilon Z_{1-\sqrt{\epsilon}}}{Z_1}\tilde{\nu}\left(\tilde{B}_\epsilon\left(x, T_\epsilon U(r)\epsilon\right)\right)$$

where $\tilde{B}_\epsilon(p,r) \subset (\tilde{E}, T_\epsilon\tilde{\delta})$ is the closed ball of radius $r > 0$ around p. Thus it follows that on the event $\mathcal{E}_\epsilon^\eta$ for each $x \in E$,

$$\frac{1}{1+\eta}T_\epsilon\tilde{\nu}(\tilde{B}_\epsilon(x,(1-\eta)r)) \leq 4\epsilon^{-1}\nu(\bar{B}_\epsilon(x,r)) \leq \frac{1}{1-\eta}T_\epsilon\tilde{\nu}(\tilde{B}_\epsilon(x,(1+\eta)r)). \tag{2.48}$$

Using the fact that $\mathbb{P}(\mathcal{E}_\epsilon^\eta) \to 1$ as $\epsilon \to 0$, Lemma 2.4.5 and that $\eta > 0$ is arbitrary, an easy pinching argument using (2.47) and (2.48) shows that

$$\lim_{\epsilon\to 0} d_{GHP}((B_\epsilon(1,1), \epsilon^{-1}\delta, 4\epsilon^{-1}\nu), (\tilde{B}_\epsilon(1,1), T_\epsilon\tilde{\delta}, T_\epsilon\tilde{\nu})) = 0 \tag{2.49}$$

almost surely. The theorem now follows from Lemma 2.4.5. $\square$

CHAPTER 3

Mixing times and Ricci curvature on the permutation group

Nathanaël Berestycki and Bati Şengül

3.1 Introduction

3.1.1 Main results

Let $\mathcal{S}_n$ denote the multiplicative group of permutations of $\{1, \ldots, n\}$. Let $\Gamma \subset \mathcal{S}_n$ be a fixed conjugacy class in $\mathcal{S}_n$, i.e., $\Gamma = \{g\tau g^{-1} : g \in \mathcal{S}_n\}$ for some fixed permutation $\tau \in \mathcal{S}_n$. Alternatively, Γ is the set of permutation in $\mathcal{S}_n$ having the same cycle structure as σ. Let $X^\sigma = (X_0, X_1, \ldots)$ be discrete-time random walk on $\mathcal{S}_n$ induced by Γ, started in the permutation $\sigma \in \mathcal{S}_n$, and let Y^σ be the associated continuous time random walk. These are the processes defined by

$$\begin{aligned} X_t^\sigma &= \sigma \circ \gamma_1 \circ \cdots \circ \gamma_t; \qquad && t = 0, 1, \ldots \\ Y_t^\sigma &= X_{N_t}^\sigma; && t \in [0, \infty) \end{aligned} \tag{3.1}$$

where $\gamma_1, \gamma_2, \ldots$ are i.i.d. random variables which are distributed uniformly in Γ; and $(N_t, t \geq 0)$ is an independent Poisson process with rate 1. Then Y is a Markov chain which converges to an invariant measure μ as $t \to \infty$. If $\Gamma \subset \mathcal{A}_n$ (where $\mathcal{A}_n$ denotes the alternating group) then μ is uniformly distributed on $\mathcal{A}_n$ and otherwise μ is uniformly distributed on $\mathcal{S}_n$. The simplest and most well known example of a conjugacy class is the set T of all transpositions, or more generally of all cyclic permutations of length $k \geq 2$. This set will play an important role in the rest of the paper. Note that Γ depends on n but we do not indicate this dependence in our notation.

The main goal of this paper is to study the cut-off phenomenon for the random walk X. More precisely, recall that the total variation distance $\|X - Y\|_{TV}$ between two random variables X, Y taking values in a set S is given by

$$\|X - Y\|_{TV} = \sup_{A \subset S} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|. \tag{3.2}$$

For $0 < \delta < 1$, the mixing time $t_{\text{mix}}(\delta)$ is by definition given by

$$t_{\text{mix}}(\delta) = \inf\{t \geq 0 : d_{TV}(t) \leq \delta\}$$

where

$$d_{TV}(t) = \sup_{\sigma} \|Y_t^\sigma - \mu\|_{TV} \tag{3.3}$$

and μ is the invariant measure defined above.

In the case where $\Gamma = T$ is the set of transpositions, a famous result of Diaconis and Shahshahani [21] is that the cut-off phenomenon takes place at time $(1/2)n \log n$ asymptotically as $n \to \infty$. That is, $t_{\text{mix}}(\delta)$ is asymptotic to $(1/2)n \log n$ for any fixed value of $0 < \delta < 1$. It has long been conjectured that for a general conjugacy class such that $|\Gamma| = o(n)$ (where here and in the rest of the paper, $|\Gamma|$ denotes the number of non fixed points of any permutation $\gamma \in \Gamma$), a similar result should hold at a time $(1/|\Gamma|)n \log n$. This has been verified for k-cycles with a fixed $k \geq 2$ by Berestycki, Schramm, and Zeitouni [12]. This is a problem with a substantial history which will be detailed below.

The primary purpose of this paper is to verify this conjecture. Hence our main result is as follows.

Theorem 3.1.1. *Let $\Gamma = \Gamma(n) \subset \mathcal{S}_n$ be a conjugacy class, which allowed to vary with n, and suppose that $|\Gamma| = o(n)$. Define*

$$t_{mix} := \frac{1}{|\Gamma|} n \log n. \tag{3.4}$$

Then for any $\epsilon > 0$,

$$\lim_{n \to \infty} d_{TV}((1 - \epsilon)t_{mix}) = 1 \quad \textit{and} \quad \lim_{n \to \infty} d_{TV}((1 + \epsilon)t_{mix}) = 0. \tag{3.5}$$

Our main tool for this result is the notion of discrete Ricci curvature as introduced by Ollivier [37], for which we obtain results of independent interest. We briefly discuss this notion here; however we point out that this turns out to be equivalent to the more well-known *path coupling method* and transportation metric introduced by Bubley and Dyer [16] and Jerrum [31] (see for instance Chapter 14 of the book Levin, Peres, and Wilmer [34] for an overview). However we will cast our results in the language of Ricci curvature because we find it more intuitive. Recall first that the definition of the L^1-Kantorovitch distance (sometimes also called Wasserstein or transportation metric) between two random variables X, Y taking values in a metric space (S, d) is given by

$$W_1(X, Y) := \inf \mathbb{E}[d(\hat{X}, \hat{Y})] \tag{3.6}$$

where the infimum is taken over all couplings $(\hat{X}, \hat{Y})$ which are distributed marginally as X and Y respectively. Ollivier's definition of Ricci curvature of a Markov chain $(X_t, t \geq 0)$ on a metric space (S, d) is as follows:

Definition 3.1.2. *Let $t > 0$. The curvature between two points $x, x' \in S$ with $x \neq x'$ is given by*

$$\kappa_t(x, x') := 1 - \frac{W_1(X_t^x, X_t^{x'})}{d(x, x')} \tag{3.7}$$

where X_t^x and $X_t^{x'}$ denote Markov chains started from x and x' respectively. The curvature of X is by definition equal to

$$\kappa_t := \inf_{x \neq x'} \kappa_t(x, x').$$

In the terminology of Ollivier [37], this is in fact the curvature of the discrete-time random walk whose transition kernel is given by $m_x(\cdot) = \mathbb{P}(X_t = \cdot | X_0 = x)$. We refer the reader to Ollivier [37] for an account of the elegant theory which can be developed using this notion of curvature, and point out that a number of classical properties of curvature generalise to this discrete setup.

For our results it will turn out to be convenient to view the symmetric group as a metric space equipped where the metric d is the word metric induced by the set T of transpositions (we will do so even when the random walk is not induced by T but by a general conjugacy class Γ). That is, the distance $d(\sigma, \sigma')$ between $\sigma, \sigma' \in \mathcal{S}_n$ is the minimal number of transpositions one must apply to get from one element to the other (one can check that this number is independent of whether right-multiplications or left-multiplications are used).

For simplicity we focus in this introduction on the case where the random walk is induced by the transpositions T. (A more general result will be stated later on the paper). Let $c > 0$, and let

$$\kappa_c(\sigma, \sigma') = 1 - \frac{W_1(X^{\sigma}_{\lfloor cn/2 \rfloor /}, X^{\sigma'}_{\lfloor cn/2 \rfloor})}{d(\sigma, \sigma')} \tag{3.8}$$

and define $\kappa_c(\sigma, \sigma) = 1$. That is, $\kappa_c(\sigma, \sigma') = \kappa_{\lfloor cn/2 \rfloor}(\sigma, \sigma')$ with our notation from (3.7).

In particular, κ_c depends on n but this dependency does not appear explicitly in the notation. It is not hard to see that $\kappa_c(\sigma, \sigma') \geq 0$ (apply the same transposition to both walks X^σ and $X^{\sigma'}$). For parity reasons it is obvious that that $\kappa_c(\sigma, \sigma') = 0$ if $d(\sigma, \sigma') = 1$. Thus we only consider the curvature between elements of even distance. For $c > 0$ define

$$\kappa_c = \inf \kappa_c(\sigma, \sigma'),$$

where the infimum is taken over all $\sigma, \sigma' \in \mathcal{S}_n$ such that $d(\sigma, \sigma')$ is even. Our main result states that κ_c experiences a phase transition at $c = 1$. More precisely, the curvature κ_c is asymptotically zero for $c \leq 1$ but for $c > 1$ the curvature is strictly positive asymptotically. In order to state our result, we introduce the quantity $\theta(c)$, which is the largest solution in $[0, 1]$ to the equation

$$\theta(c) = 1 - e^{-c\theta(c)}. \tag{3.9}$$

It is easy to see that $\theta(c) = 0$ for $c \leq 1$ and $\theta(c) > 0$ for $c > 1$. In fact, $\theta(c)$ is nothing else but the survival probability of a Galton-Watson tree with Poisson offspring distribution with mean c.

Theorem 3.1.3. *Consider the case when Γ is the set of transpositions T. If $c \leq 1$,*

$$\lim_{n\to\infty} \kappa_c = 0 \tag{3.10}$$

On the other hand, for $c > 1$

$$\liminf_{n\to\infty} \kappa_c \geq \theta(c)^4 \tag{3.11}$$

and

$$\limsup_{n\to\infty} \kappa_c \leq \theta(c)^2 \tag{3.12}$$

A more general version of this theorem will be presented later on, which gives results for the curvature of a random walk induced by a general conjugacy class Γ. This will be stated as Theorem 3.2.2.

We believe that the upper bound is the sharp one here, and thus make the following conjecture.

Conjecture 3.1.4. *For $c > 0$,*

$$\lim_{n\to\infty} \kappa_c = \theta(c)^2.$$

Of course the conjecture is already established for $c \leq 1$ and so is only interesting for $c > 1$.

3.1.2 Relation to previous works and organisation of the paper

Mixing times of Makov chains were initiated independently by Aldous [2] and by Diaconis and Shahshahani [21]. In particular, as already mentioned, Diaconis and Shahshahani [21] proved Theorem 3.1.1 in the case where Γ is the set T of transpositions. Their proof relies on some deep connections with the representation theory of $\mathcal{S}_n$ and bounds on so-called character ratios. The conjecture about the general case appears to have first been made formally in print by Roichman [42] but it has no doubt been asked privately before then. We shall see that that the lower bound $t_{\text{mix}}(\delta) \geq (1/|\Gamma|)n \log n$ is fairly straightforward; the difficult part is the corresponding upper bound.

Flatto, Odlyzko, and Wales [29] built on the earlier work of Vershik and Kerov [54] to obtain that $t_{\text{mix}}(\delta) \leq (1/2)n \log n$ when $|\Gamma|$ is bounded (as is noted in Diaconis [20, p.44-45]). This was done using character ratios and this method was extended further by Roichman [42, 43] to show an upper bound on $t_{\text{mix}}(\delta)$ which is sharp up to a constant when $|\Gamma| = o(n)$ (and in fact, more generally when $|\Gamma|$ is allowed to grow to infinity as fast as $(1-\delta)n$ for any $\delta \in (0,1)$). Again using character ratios Lulov and Pak [36] shows the cut-off phenomenon as well as $t_{\text{mix}} = (1/|\Gamma|)n \log n$ in the case when $|\Gamma| \geq n/2$. Roussel [44, 45] shows the correct mixing time as well as the cut-off phenomenon for the case when $|\Gamma| \leq 6$. Finally, in a more recent article Berestycki, Schramm, and Zeitouni [12], it is shown using coupling arguments that the cut-off phenomenon occurs and $t_{\text{mix}} = (1/k)n \log n$ in the case when Γ consists only of cycles of length k for any $k \geq 2$ fixed.

The authors in Berestycki, Schramm, and Zeitouni [12] remark that their proof can be extended to cover the case when Γ is a fixed conjugacy class and indicate that their methods can probably be pushed to cover the case when $|\Gamma| = o(\sqrt{n})$. Their argument uses very delicate estimates about the mixing time of small cycles, together with a variant of a coupling due to Schramm [47] to deal with large cycles. The most technical part of the argument is to analyse the distribution of small cycles. While our approach in this paper bears some similarities with the paper Berestycki, Schramm, and Zeitouni [12], we shall see that our use of the L^1-Kantorovitch distance (Ricci curvature) allows us to completely bypass the difficulty of ever working with small cycles. This is quite surprising given that the small cycles (in particular, the fixed points) are responsible for the occurrence of the cut-off at time t_{mix}.

 61

3.2 Curvature and mixing

3.2.1 Curvature theorem

We now start the proof of the main results of this paper. We will show how our bounds on coarse Ricci curvature imply the desired result for the upper bound on $t_{\mathrm{mix}}(\delta)$. We first state the more general version of Theorem 3.1.3 discussed in the introduction. To begin, we define the cycle structure $(k_2, k_3, \dots)$ of Γ to be a vector such that for each $j \geq 2$, there are k_j cycles of length j in the cycle decomposition of any $\sigma \in \Gamma$ (note that this is the same for any $\sigma \in \Gamma$). Then $k_j = 0$ for all $j > n$ and we have that $|\Gamma| = \sum_{j=2}^{\infty} jk_j$.

In the case for the transposition random walk the quantity $\theta(c)$ which appears in the bounds is the survival probability of a Galton-Watson process with offspring distribution given by a Poisson random variable with mean c. Our first task is to generalise $\theta(c)$. We do so via a fixed point equation, which is more complex here (and we point out that the interpretation in terms of survival probability of a certain Galton-Watson process does not hold in general). Firstly notice that for each $j \geq 2$ we have that $jk_j/|\Gamma| \leq 1$. Thus $(jk_j/|\Gamma|)_{j\geq 2}$ is compact in the product topology (the topology of pointwise convergence). Hence by considering the mixing time along convergent subsequences we can assume that without loss of generality that

$$\left(\frac{2k_2}{|\Gamma|}, \frac{3k_3}{|\Gamma|}, \dots\right) \to (k_2', k_3', \dots) \tag{3.13}$$

pointwise as $n \to \infty$. It follows that for each $j \geq 2$, $k_j' \in [0,1]$ and $\sum_{j=2}^{\infty} k_j' \leq 1$ by Fatou's lemma. For $x \in [0,1]$ and $c > 0$ define

$$\Psi(x,c) = \exp\left\{-c\left(1 - \sum_{j=2}^{\infty} k_j'(1-x)^{j-1}\right)\right\}. \tag{3.14}$$

Note that for each $c > 0$, $x \mapsto \Psi(x,c)$ is convex on $[0,1]$. In the case when $\sum_{j\geq 2} k_j' < 1$, the function $\Psi(\cdot, c)$ is not a generating function of a random variable for any $c > 0$. On the other hand if $\sum_{j\geq 2} k_j' = 1$ then for any $c > 0$ it is possible to write $\Psi(\cdot, c)$ as the generating function of a random variable.

Lemma 3.2.1. *Define*

$$c_\Gamma := \begin{cases} \left(\sum_{j=2}^{\infty}(j-1)k_j'\right)^{-1} & \text{if } \sum_{j=2}^{\infty} k_j' = 1 \\ 0 & \text{if } \sum_{j=2}^{\infty} k_j' < 1. \end{cases} \tag{3.15}$$

Then for $c > c_\Gamma$ *there exists a unique* $\theta(c) \in (0,1)$ *such that*

$$\theta(c) = 1 - \Psi(\theta(c), c).$$

For $c > c_\Gamma$, $c \mapsto \theta(c)$ *is increasing, continuous and differentiable. Further* $\lim_{c \downarrow c_\Gamma} \theta(c) = 0$ *and* $\lim_{c \uparrow \infty} \theta(c) = 1$.

Proof. For $x \in [0,1]$ and $c > 0$ define $f_c(x) := 1 - \Psi(x,c) - x$. There are two cases to consider. First suppose that $z = \sum_{j=2}^{\infty} k'_j < 1$. Then we have that

$$f_c(0) = 1 - e^{-c(1-z)} > 0 \qquad \text{and} \qquad f_c(1) = -e^{-c} < 0.$$

As $x \mapsto f_c(x)$ is convex on $[0,1]$ it follows that there exists a unique $\theta(c) \in (0,1)$ such that $f_c(\theta(c)) = 0$.

Next suppose that $\sum_{j=2}^{\infty} k'_j = 1$, then

$$f_c(0) = 0 \qquad \text{and} \qquad f_c(1) = -e^{-c} < 0$$

Moreover we have that

$$\frac{d}{dx} f_c(x)|_{x=0} = c \sum_{j=2}^{\infty} (j-1) k'_j - 1.$$

Hence for $c > c_\Gamma$ we have that $\frac{d}{dx} f_c(x)|_{x=0} > 0$ and again by convexity it follows that there exists a unique $\theta(c) \in (0,1)$ such that $f_c(\theta(c)) = 0$.

For the rest of the statements suppose that $c > c_\Gamma$. The fact that $c \mapsto \theta(c)$ is increasing follows from the definition of $\Psi(x,c)$ and the fact that $\theta(c) = \Psi(\theta(c), c)$.

Next we show continuity and differentiability. Define for $x \in [0,1]$ define $g_{c,1}(x) = 1 - \Psi(x,c)$ and for $n \geq 2$ define recursively $g_{c,n}(x) = 1 - \Psi((f_{c,n-1}(x), c)$. Then a simple argument (see Athreya and Ney [5, I.3 Lemma 2] for instance) shows that for any $x \in (0,1)$ we have that $g_{c,n}(x) \to \theta(c)$ as $n \to \infty$.

Let $\delta > 0$, then it follows that for any $x \in (0,1)$:

$$\theta(c+\delta) - \theta(c) = \lim_{n \to \infty} [g_{n,c+\delta}(x) - g_{n,c}(x)].$$

On the other hand we have that uniformly in $x \in [0,1]$,

$$\Psi(x,c) - \Psi(x,c+\delta) \leq 1 - e^{-\delta}$$

and hence it follows that

$$\theta(c+\delta) - \theta(c) \leq 1 - e^{-\delta}$$

and from this it follows that $c \mapsto \theta(c)$ is continuous and differentiable on (c_Γ, ∞).

Notice that $\theta(c) \in [0,1]$, hence $\theta(c)$ has convergent subsequences as $c \downarrow c_\Gamma$. Let L denote a subsequential limit of $\theta(c)$ as $c \downarrow c_\Gamma$. Then it follows that L solves the equation $L = 1 - \Psi(L, c_\Gamma)$. This equation has only a zero solution and thus $L = 0$ and hence $\lim_{c \downarrow c_\Gamma} \theta(c) = 0$. The limit as $c \uparrow \infty$ follows from the same argument. $\square$

In the case when $\Gamma = T$ is the set of transpositions we have that $k_2' = 1$ and $k_j' = 0$ for $j \geq 3$, hence $\Psi(x, c) = e^{-cx}$ and thus the definition of $\theta(c)$ above agrees with the definition given in the introduction.

Having introduced the bound $\theta(c)$ we now introduce the notion of Ricci curvature we will use in the general case. Let $c > 0$, and let

$$\kappa_c(\sigma, \sigma') = 1 - \frac{W_1(X^{\sigma}_{\lfloor cn/k \rfloor /}, X^{\sigma'}_{\lfloor cn/k \rfloor})}{d(\sigma, \sigma')} \tag{3.16}$$

where $k = |\Gamma|$ and define $\kappa_c(\sigma, \sigma) = 1$. Then let

$$\kappa_c = \inf \kappa_c(\sigma, \sigma'),$$

where the infimum is taken over all $\sigma, \sigma' \in \mathcal{S}_n$ such that $d(\sigma, \sigma')$ is even. That is, $\kappa_c(\sigma, \sigma') = \kappa_{\lfloor cn/k \rfloor}(\sigma, \sigma')$ with our notation from (3.7). Notice that κ_c depends on Γ, which itself depends on n. We have suppressed this dependence in our notation.

We now state a more general form of Theorem 3.1.3 which in particular covers the case of Theorem 3.1.3.

Theorem 3.2.2. *Let $\Gamma \subset \mathcal{S}_n$ be a conjugacy class and recall the definition of c_Γ from* (3.15)*. Then for $c \leq c_\Gamma$,*

$$\lim_{n \to \infty} \kappa_c = 0. \tag{3.17}$$

On the other hand, for $c > c_\Gamma$

$$\liminf_{n \to \infty} \kappa_c \geq \theta(c)^4 > 0 \tag{3.18}$$

and

$$\limsup_{n \to \infty} \kappa_c \leq \theta(c)^2 \tag{3.19}$$

where $\theta(c)$ is the maximal solution in $[0,1]$ of

$$\theta(c) = 1 - \Psi(\theta(c), c). \tag{3.20}$$

where Ψ is given by (3.14).

3.2.2 Curvature implies mixing

We now show how Theorem 3.2.2 implies Theorem 3.1.1. Again fix $\epsilon > 0$ and define $t = (1+2\epsilon)(1/k)n \log n$ and let $t' = \lfloor (1+\epsilon)(1/k)n \log n \rfloor$ where $k = |\Gamma|$. The lower bound is given in an appendix hence we are left to prove that $d_{TV}(t) \to 0$ as $n \to \infty$. For $s \geq 0$ let

$$\bar{d}_{TV}(s) := \sup_{\sigma,\sigma'} \|X_s^\sigma - X_s^{\sigma'}\|_{TV},$$

where the sup is taken over all permutations at even distances. We first claim that to show $d_{TV}(t) \to 0$ as $n \to \infty$, it suffices to prove that

$$\bar{d}_{TV}(t') \to 0 \text{ as } n \to \infty. \tag{3.21}$$

Indeed, assume that $\bar{d}_{TV}(t') \to 0$ as $n \to \infty$. Then there are two cases to consider. Assume first Γ is an even conjugacy class (meaning that $\Gamma \subset \mathcal{A}_n$). Then $X_s \in \mathcal{A}_n$ for all $s \geq 1$ and μ is uniform on $\mathcal{A}_n$. Then by Lemma 4.11 in Levin, Peres, and Wilmer [34],

$$\sup_{\sigma \in \mathcal{A}_n} \|X_{t'}^\sigma - \mu\|_{TV} \leq 2\bar{d}_{TV}(t').$$

Hence Theorem 3.1.1 follows from (3.21) in this case. In the second case, $\Gamma \subset \mathcal{A}_n^c$. In this case $X_s \in \mathcal{A}_n$ for s even, and $X_s \in \mathcal{A}_n^c$ for s odd. Using the same lemma, we deduce that if $s \geq t'$ is even,

$$\|X_s^{\mathrm{id}} - U_1\|_{TV} \leq 2\bar{d}_{TV}(s)$$

where U_1 is uniform on $\mathcal{A}_n$. However, if $s \geq t'$ is odd,

$$\|X_s^{\mathrm{id}} - U_2\|_{TV} \leq 2\bar{d}_{TV}(s)$$

where this time U_2 is uniform on $\mathcal{A}_n^c$. Let $N = (N_s : s \geq 0)$ be the Poisson clock of the random walk Y. Then $\mathbb{P}(N_s \text{ even}) \to 1/2$ as $s \to \infty$, $\mu = (1/2)(U_1 + U_2)$, and $\mathbb{P}(N_t \geq t') \to 1$ as $n \to \infty$. Thus we deduce that

$$\|Y_t^{\mathrm{id}} - \mu\|_{TV} \to 0.$$

Again, Theorem 3.1.1 follows. Hence it suffices to prove (3.21).

Note that for any two random variables X, Y on a metric space (S, d) we have the obvious inequality $\|X - Y\|_{TV} \leq W_1(X, Y)$ provided that $x \neq y$ implies $d(x, y) \geq 1$ on S. This is in particular the case when $S = \mathcal{S}_n$ and d is the word metric induced by the set T of transpositions. In other words it suffices to prove mixing in the L^1-Kantorovitch distance.

By Corollary 21 in Ollivier [37] we have that for each $s \geq 1$,

$$\sup_{\sigma,\sigma'} W_1(X^{\sigma}_{s\lfloor cn/k\rfloor}, X^{\sigma'}_{s\lfloor cn/k\rfloor}) \leq (1-\kappa_c)^s \sup_{\sigma,\sigma'} d(\sigma,\sigma') \leq n(1-\kappa_c)^s \tag{3.22}$$

since the diameter of $\mathcal{S}_n$ is equal to $n-1$. Solving

$$n(1-\kappa_c)^s \leq \delta$$

we get that

$$s \geq \frac{\log n - \log \delta}{-\log(1-\kappa_c)} \tag{3.23}$$

Thus if $u = scn/k \geq s\lfloor cn/k\rfloor$, it suffices that

$$u \geq \frac{1}{k}\frac{c}{-\log(1-\kappa_c)} n(\log n - \log \delta). \tag{3.24}$$

Now, Theorem 3.1.3 gives

$$\liminf_{n\to\infty} -\log(1-\kappa_c) \geq -\log(1-\theta(c)^4).$$

Lemma 3.2.3. *We have that*

$$\lim_{c\to\infty} \frac{c}{\log(1-\theta(c)^4)} = -1.$$

Proof. Using L'Hopital's rule twice we have that

$$\lim_{\theta\uparrow 1} \frac{\log(1-\theta)}{\log(1-\theta^4)} = \lim_{\theta\uparrow 1} \frac{1-\theta^4}{(1-\theta)4\theta^3} = 1.$$

Next we have that $\lim_{c\to\infty} \theta(c) = 1$ and hence

$$\begin{aligned}\lim_{c\to\infty} \frac{c}{\log(1-\theta(c)^4)} &= \lim_{c\to\infty} \frac{c}{\log(1-\theta(c))} = \lim_{c\to\infty} \frac{c}{\Psi(\theta(c),c)} \\ &= \lim_{c\to\infty} -\frac{1}{1-\sum_{j=2}^{\infty} k'_j(1-\theta(c))^{j-1}} \\ &= -1.\end{aligned}$$

□

Consequently we have that for $u \geq t' = \lfloor (1+\epsilon)(1/k)n\log n\rfloor$ u satisfies (3.24) for some sufficiently large $c > c_\Gamma$. Hence $\limsup_{n\to\infty} \bar{d}_{TV}(t') \to 0$, which finishes the proof.

3.2.3 Stochastic commutativity

To conclude this section on curvature, we state a simple but useful lemma. Roughly, this says that the random walk is “stochastically commutative". This can be used to show that the L^1-Kantorovitch distance is decreasing under the application of the heat kernel. In other words, initial discrepancies for the Kantorovitch metric between two permutations are only smoothed out by the application of random walk.

Lemma 3.2.4. *Let σ be a random permutation with distribution invariant by conjugacy. Let σ_0 be a fixed permutation. Then $\sigma_0 \circ \sigma$ has the same distribution as $\sigma \circ \sigma_0$.*

Proof. Define $\sigma' = \sigma_0 \circ \sigma \circ \sigma_0^{-1}$. Then since σ is invariant under conjugacy, the law of σ' is the same as the law of σ. Furthermore, we have $\sigma_0\sigma = \sigma'\sigma_0$ so the result is proved. □

This lemma will be used repeatedly in our proof, as it allows us to concentrate on events of high probability for our coupling.

3.3 Preliminaries on random hypergraphs

For the proof of Theorem 3.1.1 we rely on properties of certain random hypergraph processes. The reader who is only interested in a first instance in the case of random transpositions, and is familiar with Erdős–Renyi random graphs and with the result of Schramm [47] may safely skip this section.

3.3.1 Hypergraphs

In this section we present some preliminaries which will be used in the proof of Theorem 3.2.2. Throughout we let $\Gamma \subset \mathcal{S}_n$ be a conjugacy class and let $(k_2, k_3, \dots)$ denote the cycle structure of Γ. Thus Γ consists of permutations such that in their cycle decomposition they have k_2 many transpositions, k_3 many 3-cycles and so on. We assume that (3.13) is satisfied so that for each $j \geq 2$, $jk_j/|\Gamma| \to k'_j$ as $n \to \infty$. We also let $k = |\Gamma|$ so that $k = \sum_{j\geq 2} jk_j$.

We will need some results which are generalise those of Schramm [47]. The framework which we will use is that of random hypergraphs.

Definition 3.3.1. *A hypergraph $H = (V, E)$ is given by a set V of vertices and $E \subset \mathcal{P}(V)$ of edges, where $\mathcal{P}(V)$ denotes the set of all subsets of V. An element $e \in E$ is called a hyperedge and we call it a j-hyperedge if $|e| = j$.*

Consider the random walk $X = (X_t, t = 0, 1 \ldots)$ on $\mathcal{S}_n$ where $X_t = X_t^{\mathrm{id}}$ with our notations from the introduction. Hence

$$X_t = \gamma_1 \circ \ldots \circ \gamma_t$$

where the sequence $(\gamma_i)_{i \geq 1}$ is i.i.d. uniform on Γ. A given step of the random walk, say γ_s, can be broken down into cycles, say $\gamma_{s,1} \circ \ldots \gamma_{s,k-1}$ where $r = \sum_j k_j$. We will say that a given cyclic permutation γ has been applied to X before time t if $\gamma = \gamma_{s,j}$ for some $s \leq t$ and $1 \leq j \leq r$.

To X we associate a certain hypergraph process $H = (H_t : t = 0, 1, \ldots)$ defined as follows. For $t = 0, 1, \ldots$, H_t is a hypergraph on $\{1, \ldots, n\}$ where a hyperedge $\{x_1, \ldots, x_j\}$ is present if and only if the cyclic permutation $(x_1, \ldots, x_j)$ has been applied to the random walk X prior to time t. For instance, H_1 has exactly k_j many j-hyperedges for $j \geq 2$. Note that the presence of hyperedges are not independent.

3.3.2 Giant component of the hypergraph

In the case $\Gamma = T$, the set of transpositions, the hypergraph H_s is a realisation of an Erdős-Renyi graph. Analogous to Erdős-Renyi graphs, we first present a result about the size of the components of the hypergraph process $H = (H_t : t = 0, 1, \ldots)$ (where by size, we mean the number of vertices in this component). For the next lemma recall the definition of $\Psi(x, c)$ in (3.14). Recall that for $c > c_\Gamma$, where c_Γ is given by (3.15), there exists a unique root $\theta(c) \in (0, 1)$ of the equation $\theta(c) = 1 - \Psi(\theta(c), c)$.

Theorem 3.3.2. *Consider the random hyper graph H_s and suppose that $s = s(n)$ is such that $sk/n \to c$ as $n \to \infty$ for some $c > c_\Gamma$. Then there is a constant $\beta > 0$, depending only on c, such that with probability tending to one all connected components but the largest have size at most $\beta \log(n)$. Further the size of the largest connected component, normalised by n, converges to $\theta(c)$ in probability as $n \to \infty$.*

Of course, this is the standard Erdős–Renyi theorem in the case where $\Gamma = T$ is the set of transpositions. See for instance Durrett [24], in particular Theorem 2.3.2 for a proof. In the case of k-cycles with k fixed and finite, this is the case of random regular hyper graphs analysed by Karoński and Łuczak [32]. For the slightly more general case of bounded conjugacy classes, this was proved by Berestycki [11].

Remark 3.3.3. *Note that the behaviour of H_s in Theorem 3.3.2 can deviate markedly from that of Erdős–Renyi graphs. The most obvious difference is that H_s can contain mesoscopic components, something which has of course negligible probability for Erdős-Renyi graphs. For example, suppose Γ consists of $n^{1/2}$ transpositions and one cycle of*

 68

length $n^{1/3}$. Then the giant component appears at time $n^{1/2}/2$ with a phase transition. Yet even at the first step there is a component of size $n^{1/3}$. (However it will follow from the proof that, in the supercritical phase $c > c_\Gamma$, such a dichotomy still holds). From a technical point of view this has nontrivial consequences, as proofs of the existence of a giant component are usually based on the dichotomy between microscopic components and giant components. Furthermore, when the conjugacy class is large and consists of many small or mesoscopic cycles, the hyperedges have a strong dependence, which makes the proof very delicate.

Proof of Theorem 3.3.2. Suppose that $s = s(n)$ is such that $sk/n \to c$ for some $c > c_\Gamma$ as $n \to \infty$ for some $c \geq 0$. We reveal the vertices of the component containing a fixed vertex $v \in \{1, \ldots, n\}$ using breadth-first search exploration, as follows. There are three states that each vertex can be: unexplored, removed or active. Initially v is active and all the other vertices are unexplored. At each step of the iteration we select an active vertex w according to some prescribed rule among the active vertices at this stage (say with the smallest label). The vertex w becomes removed and every unexplored vertex which is joined to w by a hyperedge becomes active. We repeat this exploration procedure until there are no more active vertices.

At stage $i = 0, 1, \ldots$ of this exploration process, we let A_i, R_i and U_i denote the set of active, removed and unexplored vertices respectively. Thus initially $A_0 = \{v\}$, $U_0 = \{1, \ldots, n\} \backslash \{v\}$ and $R_0 = \emptyset$.

We will need to keep track of the hyperedges we reveal and where they came from, in order to deal with dependencies mentioned in Remark 3.3.3. For $t = 1, \ldots, s$ we call the hyperedges which are in H_t but not in H_{t-1} the t-th packet. Note that each packet consists of k_j hyperedges of size j, $j \geq 2$, which are sampled uniformly at random without replacement from $\{1, \ldots, n\}$. However, crucially, hyperedges from different packets are independent. For $t = 1, \ldots, s$ and $j \geq 2$ let $Y_j^{(t)}(i)$ be the number of j-hyperedges in the t-th packet that were revealed in the exploration process, prior to step i. Let $i \geq 0$ and let $\mathcal{H}_i$ denote the filtration generated by the exploration process up to stage i, including the information of which edge came from which packet:

$$\mathcal{H}_i = \sigma(A_1, \ldots, A_i, Y_j^{(t)}(1), \ldots, Y_j^{(t)}(i) : 1 \leq t \leq s, j \geq 2).$$

Our goal will be to give uniform stochastic bounds on the distribution of $|A_{i+1} \setminus A_i|$, so long as i is not too large. We will thus fix i and in order to ease notations we will often suppress the dependence on i, in $Y_j^{(t)}(i)$: we will thus simply write $Y_j^{(t)}$. Note that by definition, for each $t = 1, \dots, s$ and $j \geq 2$, $Y_j^{(t)} \leq k_j$ and

$$\sum_{t=1}^{s} \sum_{j \geq 2} Y_j^{(t)} = n - |U_i| = |A_i| + i. \tag{3.25}$$

Let w be the vertex being explored for stage $i+1$. For $t = 1, \dots, s$ let M_t be the indicator that w is part of a hyperedge in the t-th packet. Thus, $(M_t)_{1 \leq t \leq s}$ are independent conditionally given $\mathcal{H}_i$, and

$$\mathbb{P}(M_t = 1 | \mathcal{H}_i) = \sum_{j \geq 2} \frac{j(k_j - Y_j^{(t)})}{|U_i|} \tag{3.26}$$

If w is part of a hyperedge in the t-th packet, let V_t be the size of the (unique) hyperedge of that packet containing it. Then

$$\mathbb{P}(V_t = j | \mathcal{H}_i, M_t = 1) = \frac{j(k_j - Y_j^{(t)})}{\sum_{m \geq 2} m(k_m - Y_m^{(t)})} \tag{3.27}$$

Note that when $M_t = 1$ it implies that the denominator above is non-zero and thus (3.27) is well defined. When $M_t = 0$ we simply put $V_t = 1$ by convention. Then we have the following almost sure inequality:

$$|A_{i+1} \setminus A_i| \leq -1 + \sum_{t=1}^{s} M_t(V_t - 1). \tag{3.28}$$

This would be an equality if it were not for possible self-intersections, as hyperedges connected to w coming from different packets may share several vertices in common. In order to get a bound in the other direction, we simply truncate the $|A_{i+1} \setminus A_i|$ at $n^{1/4}$. Let I_i be the indicator that among the first $n^{1/4}$ vertices, no such self-intersection occurs. Note that $\mathbb{E}(I_i) \geq p_n = 1 - n^{-1/2}$, by straightforward bounds on the birthday problem. We then have

$$|A_{i+1} \setminus A_i| \wedge n^{1/4} \geq -1 + I_i \left(\sum_{t=1}^{s} M_t(V_t - 1) \wedge n^{1/4} \right). \tag{3.29}$$

We will stop the exploration process once we have discovered enough vertices, or if the active set dies out, whichever comes first. Therefore we define

$$T^{\uparrow} := \inf\{\ell \geq 1 : |A_\ell| + \ell > 2n^{2/3}\}$$
$$T^{\downarrow} := \inf\{\ell \geq 1 : |A_\ell| = 0\}$$

and we set $T = T^{\uparrow} \wedge T^{\downarrow}$. The following lemma shows that the distribution of $|A_{i+1} \setminus A_i|$ converges to a limit in distribution, uniformly for $i < T$. (Note however that the limit is improper if $\sum_j k'_j < 1$.)

Lemma 3.3.4. *There exists some deterministic function $w : \mathbb{N} \to \mathbb{R}$ such that $w(n) \to 0$ as $n \to \infty$ with the following property. For each $x \in (0, 1)$,*

$$\sup_{i \geq 1} \left| \mathbb{E}[x^{|A_{i+1} \setminus A_i|} | \mathcal{H}_i] - \frac{\Psi(1-x, c)}{x} \right| 1_{\{T > i\}} \leq w(n)$$

almost surely.

Proof. Suppose $T > i$. In particular, from the definition of $T^{\uparrow}$ and (3.25) we have that

$$\sum_{t=1}^{s} \sum_{j \geq 2} j Y_j^{(t)} \leq 2n^{2/3} \tag{3.30}$$

almost surely. From (3.28) we have that

$$x\mathbb{E}[x^{|A_{i+1} \setminus A_i|} | \mathcal{H}_i] \geq \mathbb{E}[x^{\sum_{t=1}^{s} M_t(V_t - 1)} | \mathcal{H}_i] = \prod_{t=1}^{s} \mathbb{E}(x^{M_t(V_t-1)} | \mathcal{H}_i)$$
$$= \prod_{t=1}^{s} \left[1 - \mathbb{P}(M_t = 1 | \mathcal{H}_i)(1 - \mathbb{E}(x^{V_t - 1} | \mathcal{H}_i, M_t = 1))\right].$$

Recall from (3.27) that

$$\mathbb{E}(x^{V_t - 1} | \mathcal{H}_i, M_t = 1) = \sum_{j \geq 2} x^{j-1} \frac{j(k_j - Y_j^{(t)})}{\sum_{m \geq 2} m(k_m - Y_m^{(t)})} \geq \sum_{j \geq 2} x^{j-1} \frac{j(k_j - Y_j^{(t)})}{k}$$

and from (3.26) that

$$\mathbb{P}(M_t = 1 | \mathcal{H}_i) = \sum_{m \geq 2} \frac{m(k_m - Y_m^{(t)})}{|U_i|} \leq \sum_{m \geq 2} \frac{m k_m}{n - 2n^{2/3}} \leq \frac{k}{n}(1 + 3n^{-1/3})$$

by (3.30). Therefore, using $1-x \geq e^{-x-x^2}$ for all x sufficiently small,

$$\begin{aligned}
x\mathbb{E}[x^{|A_{i+1}\setminus A_i|}|\mathcal{H}_i] &\geq \prod_{t=1}^{s}\left[1-\frac{k}{n}(1+3n^{-1/3})\left(1-\sum_{j\geq 2}x^{j-1}\frac{j(k_j-Y_j^{(t)})}{k}\right)\right] \\
&\geq \prod_{t=1}^{s}\left\{1-\frac{k}{n}\left(1+3n^{-1/3}-\sum_{j\geq 2}x^{j-1}\frac{jk_j}{k}\right)-\frac{k}{n}\sum_{j\geq 2}x^{j-1}\frac{jY_j^{(t)}}{k}\right\} \\
&\geq \prod_{t=1}^{s}\left\{1-\frac{k}{n}\left(1-\sum_{j\geq 2}x^{j-1}\frac{jk_j}{k}\right)-3n^{-1/3}\frac{k}{n}-\frac{1}{n}\sum_{j\geq 2}jY_j^{(t)}\right\} \\
&\geq \exp\left\{-s\frac{k}{n}\left(1-\sum_{j\geq 2}x^{j-1}\frac{jk_j}{k}\right)-O(n^{-1/3})-O(s\frac{k^2}{n^2})\right\}.
\end{aligned}$$

Hence

$$x\mathbb{E}[x^{|A_{i+1}\setminus A_i|}|\mathcal{H}_i] \geq \psi(1-x,c)(1+w_1(n))$$

where $w_1(n)$ vanishes at infinity. For the last inequality we have used that

$$\exp(-c\sum_j x^{j-1}\frac{jk_j}{k}) \to \exp(-c\sum_j x^{j-1}k'_j) \tag{3.31}$$

which follows from the dominated convergence theorem, as jk_j/k is uniformly bounded by 1. Note that the above estimate is uniform in $i \geq 1$.

For the upper bound, we use (3.29). Let $\epsilon_n \to 0$ sufficiently slowly that $\epsilon_n n^{1/3} \to \infty$. For concreteness take $\epsilon_n = n^{-1/6}$. Define

$$G := \left\{t \in \{1,\dots,s\} : \sum_{m\geq 2} mY_m^{(t)} \leq \epsilon_n k\right\},$$

and let $I = G^c$. Packets $t \in I$ are the bad packets for which a significant fraction of the mass (at least ϵ_n) was already discovered. In the case where the conjugacy class contains only one type of cycles, say k-cycles, then I coincides with the set of hyperedges already revealed. At the other end of the spectrum, when the conjugacy class Γ is broken down into many small cycles, then I is likely to be empty. But in all cases, $|I|$ satisfies the trivial bound

$$|I| \leq \frac{n^{2/3}}{\epsilon_n k}$$

by (3.30), and in particular

$$\frac{k|I|}{n} \leq \frac{1}{\epsilon_n n^{1/3}} \leq n^{-1/6} \to 0. \tag{3.32}$$

This turns out to be enough for our purposes.

Note that $\mathbb{E}(x^{\sum_{t=1}^s M_t(V_t-1)})$ and $\mathbb{E}(x^{n^{1/4}\wedge\sum_{t=1}^s M_t(V_t-1)})$ can only differ by at most $x^{n^{1/4}}$, which is exponentially small, so we can neglect this difference. Then we may write, counting only hyper edges from good packets, using the fact that $1-x \le e^{-x}$ for all $x \in \mathbb{R}$, and (3.32):

$$\begin{aligned}
&x\mathbb{E}[x^{|A_{i+1}\setminus A_i|}|\mathcal{H}_i] \\
&\le 1-\mathbb{E}(I_i)+\mathbb{E}(I_i)\left(x^{n^{1/4}}+\prod_{t=1}^{s}\left[1-\frac{k-\sum_{m\ge 2} mY_m^{(t)}}{n}\left(1-\sum_{j\ge 2}x^{j-1}\frac{j(k_j-Y_j^{(t)})}{k-\sum\limits_{m\ge 2} mY_m^{(t)}}\right)\right]\right) \\
&\le 2n^{-1/2}+\prod_{t\in G}\left[1-\frac{k}{n}(1-\epsilon_n)\left(1-\sum_{j\ge 2}x^{j-1}\frac{jk_j}{k(1-\epsilon_n)}\right)\right] \\
&\le 2n^{-1/2}+\exp\left\{-s\frac{k}{n}+(1-\epsilon_n)\frac{k}{n}|I|+\epsilon_n\frac{sk}{n}+s\frac{k}{n}\sum_{j\ge 2}x^{j-1}\frac{jk_j}{k}\right\} \\
&= 2n^{-1/2}+\exp\left\{-s\frac{k}{n}+\frac{sk}{n}\sum_{j\ge 2}x^{j-1}\frac{jk_j}{k}\right\}(1+2c\epsilon_n+2n^{-1/6}) \qquad (3.33) \\
&\le \Psi(1-x,c)(1+w_2(n))
\end{aligned}$$

where the function $w_2:\mathbb{N}\to\mathbb{R}$ vanish at infinity, invoking again (3.31). The proof is complete. □

We will need the following lemma which tells us the number of vertices in logarithmically large components, among other things.

Lemma 3.3.5. *For any $\beta > 0$, We have that*

$$\lim_{n\to\infty}\mathbb{P}(T^{\downarrow} > \beta\log n) = \theta(c). \qquad (3.34)$$

Moreover, letting for $v \in \{1,\dots,n\}$ let $\mathcal{C}_v$ denote the size of the component containing v

$$\frac{1}{n}|\{v : |\mathcal{C}_v| \ge \beta\log n\}| \to \theta(c) \qquad (3.35)$$

in probability as $n\to\infty$.

Proof. We start with the lower bound of (3.34). For simplicity write $\theta = \theta(c)$. Let $x = x_n$

be the solution of the equation

$$\exp\left\{-s\frac{k}{n}+\frac{sk}{n}\sum_{j\geq 2}x^{j-1}\frac{jk_j}{k}\right\}=x.$$

It is easy to check that $x_n \in (0,1)$ is well-defined and $x_n \to 1-\theta$ as $n\to\infty$. From (3.33) we see that

$$\mathbb{E}[x_n^{|A_{i+1}\setminus A_i|}|\mathcal{H}_i]\leq 1+\tilde{w}(n),$$

uniformly in $i \leq T$ for some deterministic function $\tilde{w}:\mathbb{N}\to\mathbb{R}$ such that $|\tilde{w}(n)| = O(n^{-1/6})$. Consequently,

$$M_i = x_n^{|A_i|}(1+\tilde{w}(n))^i, i=1,\dots,T,$$

forms a supermartingale. Let $T_r = \inf\{i\geq 0 : |A_i|\geq r\}$. Note that if $T_r < T^{\downarrow}$ then $T^{\downarrow} > r$. Thus if $T^{\downarrow} < r$, then $T^{\downarrow} < T_r$. We apply the optional stopping theorem at time $r\wedge T_r\wedge T^{\downarrow}$, and we bound from below M by considering only its value on the event $\{T^{\downarrow} < r\}$, in which case also $T_r > T^{\downarrow}$, and hence $r\wedge T_r\wedge T^{\downarrow} = T^{\downarrow}$. Therefore,

$$\begin{aligned} M_{r\wedge T_r\wedge T^{\downarrow}} &\geq M_{T^{\downarrow}}1_{\{T^{\downarrow}<r\}} \\ &\geq (1+\tilde{w}(n)))^r 1_{\{T^{\downarrow}<r\}}. \end{aligned}$$

Hence

$$\mathbb{P}(T^{\downarrow}<r)\leq (1+\tilde{w}(n))^r\mathbb{E}(M_{r\wedge T_r\wedge T^{\downarrow}})\leq \mathbb{E}(M_0)=x_n.$$

Taking $r=\beta\log n$, and recalling that $x_n\to 1-\theta$, we deduce that

$$\limsup_{n\to\infty}\mathbb{P}(T^{\downarrow}<\beta\log n)\leq 1-\theta$$

from which the lower bound of (3.34) follows. For the upper bound of (3.34), we make the following observation. Let $m\geq 1$, be finite arbitrary (eventually chosen to be large), and observe $\mathbb{P}(T^{\downarrow}>\beta\log n)\leq \mathbb{P}(T^{\downarrow}>m)$. Now, let $X_{i+1}=|A_{i+1}\setminus A_i|$. It follows from Lemma 3.3.4 that $(X_1,\dots,X_m)$ converge to i.i.d. random variables $(Y_1,\dots,Y_m)$ (which are possibly improper, if $\sum k_j'<1$) having as generating function $\mathbb{E}(x^Y)=\psi(1-x,c)/x$. Formally,

$$Y=\left(\sum_j (j-1)\text{ Poisson }(ck_j')\right)+\left(\infty\cdot\text{ Poisson }(c(1-\sum_j k_j'))\right)$$

where the Poisson random variables are independent. Let $S_i=\sum_{j\leq i}Y_i$ and $H=\inf\{i\geq$

$0 : S_i = 0\}$. Then clearly for all m,

$$\lim_{m\to\infty} \mathbb{P}(T^{\downarrow} \geq m) = \mathbb{P}(H \geq m)$$

and thus

$$\limsup_{n\to\infty} \mathbb{P}(T^{\downarrow} > \beta \log n) \leq \limsup_{m\to\infty} \mathbb{P}(H > m).$$

On the other hand the right hand side is easily shown, by standard random walk theory, to equal θ. Thus the upper bound of (3.34) follows. We now turn to (3.35). Observe that $|\mathcal{C}_v| \geq \beta \log n$ precisely if $T^{\downarrow} \geq \beta \log n$, hence if $Z = \sum_{v=1}^{n} 1_{\{|\mathcal{C}_v| \geq \beta \log n\}}$, we have that $\mathbb{E}(Z)/n \to \theta$ by (3.34). Hence if we show that $Var(Z) = o(n^2)$ then (3.35) follows by Chebyshev's inequality. In particular, it suffices to show that for $v \neq w \in \{1, \ldots, n\}$,

$$\limsup_{n\to\infty} \mathrm{Cov}(1_{\{|\mathcal{C}_v| \geq \beta \log n\}}, 1_{\{|\mathcal{C}_w| \geq \beta \log n\}}) \leq 0$$

or equivalently,

$$\limsup_{n\to\infty} \mathbb{P}(|\mathcal{C}_v| \geq \beta \log n, |\mathcal{C}_w| \geq \beta \log n) \leq \theta(c)^2 \tag{3.36}$$

given that we already from (3.34) that $\mathbb{P}(|\mathcal{C}_v| \geq \beta \log n) \to \theta(c)$. On the other hand, (3.36) can be proved in exactly the same way as the upper bound of (3.34) above. Details are left to the reader. □

We claim that we can choose $x \in (0, 1)$ such that $\Psi(1 - x, c)/x < 1$. There are two cases to consider. If $\sum_{j\geq 2} k'_j < 1$

$$\psi(1-x, c) =: z(x) = \exp\left(-c + c\sum_{j\geq 2} x^{j-1} k'_j\right) \leq z(1) = \exp\left(-c + c\sum_{j\geq 2} k'_j\right) < 1$$

so the result is trivial. Otherwise, $z(1) = 1$ and it is not hard to argue that

$$\frac{d}{dx} z(x)|_{x=1} = c\sum_{j\geq 2}(j-1)k'_j > 1$$

by definition of c_Γ and since $c > c_\Gamma$. By Taylor's theorem it follows that $z(x)/x < 1$ for some $0 < x < 1$ sufficiently close to 1.

Hence, using Lemma 3.3.4, for x fixed as above, we can suppose that n is large enough so that

$$\mathbb{E}[x^{|A_{i+1} \setminus A_i|} | \mathcal{H}_i] \leq (1 + \epsilon)^{-1} \tag{3.37}$$

almost surely on $\{T > i\}$ for some fixed $\epsilon > 0$.

Step 1. We now fix x as in (3.37) and let $r = \beta \log n$ for some constant $\beta > 0$ to be chosen later. Our first task is to show that if $|A_r| > 0$, then it follows that $T^{\uparrow} < T^{\downarrow}$ with high probability. If $T^{\uparrow} < r$ there is nothing to do so we may assume that $T \geq r$. To do this, let $C > 3\log(1/x)$, we first show that if $T \geq r$ and $|A_r| > 0$, then $|A_r| > C \log n$ with high probability.

$$\begin{aligned}
\mathbb{P}(T > r, 0 < |A_r| \leq C \log n) &\leq \mathbb{P}(x^{|A_r|} \geq x^{C \log n}; T \geq r) \\
&\leq \frac{\mathbb{E}[x^{|A_r|} \mathbb{1}_{\{T \geq r\}}]}{x^{C \log n}} \\
&= x^{-C \log n + 1} \mathbb{E}\left[\prod_{i=0}^{r-1} \mathbb{E}[x^{|A_{i+1} \setminus A_i|} \mathbb{1}_{\{T > i\}} | \mathcal{H}_i] \right] \\
&\leq x^{-C \log n + 1} (1+\epsilon)^{-\beta \log n}.
\end{aligned}$$

Thus we can choose $\beta > 0$ suitably large so that

$$\mathbb{P}(T > r, 0 < |A_r| \leq C \log n) \leq n^{-3}.$$

Step 2. We now show that, under the assumption $T > r$, $T^{\downarrow}$ is unlikely to occur before $T^{\uparrow}$. For $i \geq r$, let $M_i := x^{|A_{i \wedge T}|}(1+\epsilon)^{i \wedge T - r}$. Then it is not hard to check that $M = (M_i : i = r, \dots, T)$ is a supermartingale in the filtration $(\mathcal{H}_r, \mathcal{H}_{r+1}, \dots)$. Observe that necessarily $T \leq 2n^{2/3}$ so M is bounded. Suppose that $T > r$. Note that on the event $\{T = T^{\downarrow}\}$,

$$M_T = (1+\epsilon)^{T^{\downarrow} - r} \geq 1_{\{T = T^{\downarrow}\}}$$

hence by the optional stopping theorem (since M is bounded), on the event $\{T > r\}$

$$\begin{aligned}
\mathbb{P}(T = T^{\downarrow} | \mathcal{H}_r) &\leq \mathbb{E}(M_T 1_{\{T = T^{\downarrow}\}} | \mathcal{H}_r) \\
&\leq M_r = x^{|A_r|}.
\end{aligned}$$

We deduce that

$$\mathbb{P}\left(T = T^{\downarrow}; T > r\right) \leq \mathbb{E}(x^{C \log n}; |A_r| > 0) + \mathbb{P}(T > r; 0 < |A_r| < C \log n)$$

and hence

$$\mathbb{P}\left(T = T^{\downarrow}; T > r\right) \leq x^{C \log n} + n^{-3} \leq 2n^{-3}. \tag{3.38}$$

Step 3. Note that if $T^{\uparrow} > T^{\downarrow}$ we have necessarily that $|A_{n^{2/3}}| > 0$ (indeed, recall that $|A_i| + i$ is monotone as the total number of vertices discovered by stage i). In our third step we show that if $|A_{n^{2/3}}| > 0$, then with high probability $|A_{n^{2/3}}| \geq K n^{2/3}$ for some constant $K > 0$. There are two cases to consider: either $T^{\uparrow} \leq n^{2/3}$ or $T^{\uparrow} > n^{2/3}$. In the first case we have that

$$|A_{n^{2/3}}| + n^{2/3} \geq |A_T^{\uparrow}| + T^{\uparrow} \geq 2n^{2/3}$$

since $|A_i| + i$ is the number of vertices discovered by stage i and is thus monotone, and the second inequality is the definition of $T^{\uparrow}$. Therefore,

$$|A_{n^{2/3}}| \geq n^{2/3}$$

and so the claim is satisfied with $K = 1$. Thus consider the second case $T^{\uparrow} < n^{2/3}$. Since we are also assuming that $|A_{n^{2/3}}| > 0$, we may thus assume that $T > n^{2/3}$. Now

$$\begin{aligned}
\mathbb{P}(|A_{n^{2/3}}| \leq Kn^{2/3}; T > n^{2/3}) &\leq \frac{\mathbb{E}[x^{|A_{n^{2/3}}|}\mathbb{1}_{\{T > n^{2/3}\}}]}{x^{Kn^{2/3}}} \\
&= x^{-Kn^{2/3}}\mathbb{E}\left[\prod_{i=0}^{n^{2/3}-1} \mathbb{E}[x^{|A_{i+1} \setminus A_i|}\mathbb{1}_{\{T>i\}}|\mathcal{H}_i]\right] \\
&\leq (1+\epsilon)^{-n^{2/3}} x^{-Kn^{2/3}}.
\end{aligned}$$

Let $K > 0$ be chosen small enough that $x^{-K} < 1 + \epsilon$, so that the above quantity decays exponentially in $n^{2/3}$, and in particular is smaller than n^{-3} for n sufficiently large. In either case we see that

$$\mathbb{P}\left(|A_{n^{2/3}}| \leq Kn^{2/3}; |A_{n^{2/3}}| > 0\right) \leq n^{-3}.$$

Step 4. Combining this with (3.38) we get

$$\begin{aligned}
\mathbb{P}\left(|A_{n^{2/3}}| \leq Kn^{2/3}; T > r\right) &\leq \mathbb{P}(|A_{n^{2/3}}| = 0; T > r) + \mathbb{P}(|A_{n^{2/3}}| \leq Kn^{2/3}; T > n^{2/3}) \\
&\leq \mathbb{P}(T = T^{\downarrow}; T > r) + n^{-3} \\
&\leq 3n^{-3}.
\end{aligned}$$

In particular,

$$\mathbb{P}(|A_{T\wedge n^{2/3}}| \leq Kn^{2/3}; T^{\downarrow} > r) \leq 3n^{-3}. \tag{3.39}$$

Suppose v is a vertex and that $|\mathcal{C}_v| > \beta \log n = r$. Then observe that $T^{\downarrow} > r$. Accordingly, it is likely that $|A_{T\wedge n^{2/3}}| \geq Kn^{2/3}$ by (3.39). If v' is another vertex and we assume that $|\mathcal{C}_{v'}| > r$, we may likewise explore its component. We seek to show that v and v' are likely to be connected. As we explore $\mathcal{C}_{v'}$ we may find a connection from $\mathcal{C}_{v'}$ to $\mathcal{C}_v$ before time $T \wedge n^{2/3}$ (in the exploration of $\mathcal{C}_{v'}$) in which case we are done. Else, we can repeat the argument above and show that it is likely that the active vertex set of $\mathcal{C}_{v'}$ also reaches $Kn^{2/3}$, at a time $T' \wedge n^{2/3}$, with obvious notations. Let $A = A_{T\wedge n^{2/3}}$

(resp. $A' = A'_{T' \wedge n^{2/3}}$) denote the active vertex set of $\mathcal{C}_v$ (resp. $\mathcal{C}_{v'}$) at time $T \wedge n^{2/3}$ (resp. $T' \wedge n^{2/3}$). Hence we may assume that $A \cap A' = \emptyset$ and $|A|, |A'| \geq Kn^{2/3}$. We now show that A and A' are likely to be connected by making use of the sprinkling technique. That is, suppose we add s' packets, with

$$s' = \left\lceil \frac{Dn^{2/3} \log n}{k} \right\rceil$$

for some $D > 0$ to be chosen later on. Note that $s'k/n \to 0$ so that $(s+s')k/n \to c$. Since $s = s(n)$ is an arbitrary sequence such that $sk/n \to c$ it suffices to show that v and v' are then connected at time $s + s'$. In fact we will check that A and A' are connected using smaller edges that the hyperedges making each packet, as follows. For each hyperedge of size j we will only reveal a subset of $\lfloor j/2 \rfloor$ edges with disjoint support. This gives us a total of at least $k/2$ edges for each packet which are sampled uniformly at random without replacement from $\{1, \ldots, n\}$. We will check that a connection occurs between A and A' within these $s'k/2$ edges.

Let us say that A (resp. A') is left half-vacant by a given (sub)packet if the intersection of the edges of the pack with A (resp. A') don't contain more than $Kn^{2/3}/2$ vertices, and call A or A' half-full otherwise. It is obvious that if A is half-full then the probability for an edge to join A to A' tends to one exponentially fast in $n^{2/3}$, so we restrict to the case where a given (sub)packet leaves both A and A' half-vacant. In this case, each edge from subpacket connects A to A' with probability at least

$$\frac{(Kn^{2/3}/2)^2}{(n-k)^2} \geq \frac{K^2}{8n^{2/3}},$$

independently for each edge within a given (sub)packet, and hence in particular independently for all the $s'k/2$ edges we are adding in total. Consequently, the probability that no connection occurs during these $s'k/2$ trials is at most

$$\left(1 - \frac{K^2}{8n^{2/3}}\right)^{s'k/2} \leq \exp\left(-\frac{K^2}{8n^{2/3}} \frac{Dn^{2/3} \log n}{4}\right) = \exp\left(-\frac{DK^2}{32} \log n\right).$$

For $D > 0$ sufficiently large this is less than n^{-3}.

Step 5. We are now ready to conclude that vertices are either in small component at time s or connected at time $s+s'$. For $v \in \{1, \ldots, n\}$ let $\mathcal{C}_v(s)$ denote the the component containing v at time s and for $v, v' \in \{1, \ldots, n\}$. Write $v \leftrightarrow v'$ to indicate that v is connected to v' and define the good event

$$G(v, v') := \{v \leftrightarrow v' \text{ at time } s + s'\} \cup \{|\mathcal{C}_v(s)| \leq \beta \log n\} \cup \{|\mathcal{C}_{v'}(s)| \leq \beta \log n\}.$$

Altogether we have just shown that $\mathbb{P}(G(v,v')^c) \le 4n^{-3}$, since if . Hence we see that by a union bound

$$\mathbb{P}\left(\bigcap_{v,v' \in \{1,\dots,n\}} G(v,v')\right) \ge 1 - 4n^{-1}.$$

Let $V' = \{v : |\mathcal{C}_v(s)| \ge \beta \log n\}$. Then by (3.35), we know that $|V'|/n \to \theta(c)$ in probability as $n \to \infty$. Moreover, we see that all the vertices of V' are connected with probability $1 - o(1)$ at time $s + s'$. Theorem 3.3.2 follows. □

3.3.3 Poisson–Dirichlet structure

The renormalised cycle lengths $\mathfrak{X}(\sigma)$ of a permutation $\sigma \in \mathcal{S}_n$ is the cycle lengths of σ divided by n, written in decreasing order. In particular we have that $\mathfrak{X}(\sigma)$ takes values in

$$\Omega_\infty := \{(x_1 \ge x_2 \ge \dots) : x_i \in [0,1] \text{ for each } i \ge 1 \text{ and } \sum_{i=1}^{\infty} x_i = 1\}. \tag{3.40}$$

We equip Ω_∞ with the topology of pointwise convergence. If σ_n is uniformly distributed in $\mathcal{S}_n$ then $\mathfrak{X}(\sigma_n) \to Z$ in distribution as $n \to \infty$ where Z is known as a Poisson–Dirichlet random variable. It can be constructed as follows. Let $U_1, U_2, \dots$ be i.i.d. uniform random variables on $[0,1]$. Let $Z_1^* = U_1$ and inductively for $i \ge 2$ set $Z_i^* = U_i(1 - \sum_{j=1}^i Z_j^*)$. Then $(Z_1^*, Z_2^*, \dots)$ can be ordered in decreasing size and the random variable Z has the same law as $(Z_1^*, Z_2^*, \dots)$ ordered by decreasing size.

The next result is a generalisation of Theorem 1.1 in Schramm [47] to the case of general conjugacy classes. The proof is a simple adaptation of the proof of Schramm and we provide the details in an appendix.

Theorem 3.3.6. *Suppose $s = s(n)$ is such that $sk/n \to c$ as $n \to \infty$ for some $c > c_\Gamma$. Then we have that for any $m \in \mathbb{N}$*

$$\left(\frac{\mathfrak{X}_1(X_s)}{\theta(c)}, \dots, \frac{\mathfrak{X}_m(X_s)}{\theta(c)}\right) \to (Z_1, \dots, Z_m)$$

in distribution as $n \to \infty$ where $Z = (Z_1, Z_2, \dots)$ is a Poisson–Dirichlet random variable.

3.4 Proof of curvature theorem

3.4.1 Proof of the upper bound on curvature

We claim that it is enough to show the upper bound for $c > c_\Gamma$ in (3.19). Indeed, notice that $c \mapsto \kappa_c$ is increasing. Let $c \le c_\Gamma$ and assume that $\limsup_{n\to\infty} \kappa_{c'} \le \theta(c')^2$ holds for all $c' > c_\Gamma$. Then we have that $\limsup_{n\to\infty} \kappa_c \le \theta(c')^2$ for each $c' > c_\Gamma$. Taking $c' \downarrow c_\Gamma$ and using the fact that $\lim_{c\downarrow c_\Gamma} \theta(c) = 0$ shows that $\lim_{n\to\infty} \kappa_c = 0$.

Fix $c > c_\Gamma$ and let $t := \lfloor cn/|\Gamma| \rfloor$. We are left to show (3.19). In other words, we wish to prove that for some $\sigma, \sigma' \in \mathcal{S}_n$

$$\liminf_{n\to\infty} \frac{W_1(X_t^\sigma, X_t^{\sigma'})}{d(\sigma,\sigma')} \ge 1 - \theta(c)^2.$$

We will choose $\sigma = \mathrm{id}$ and $\sigma' = \tau_1 \circ \tau_2$, where τ_1, τ_2 are independent uniformly chosen transpositions. To prove the lower bound on the Kantorovitch distance we use the dual representation of the distance $W_1(X,Y)$ between two random variables X, Y:

$$W_1(X,Y) = \sup\{\mathbb{E}[f(X)] - \mathbb{E}[f(Y)] : f \text{ is Lipschitz with Lipschitz constant } 1\}. \quad (3.41)$$

Let $f(\sigma) = d(\mathrm{id}, \sigma)$ be the distance to the identity (using only transpositions, as usual). Then observe that f is 1-Lipschitz. It suffices to show

$$\liminf_{n\to\infty} \mathbb{E}[f(X_t^{\tau_1\circ\tau_2})] - \mathbb{E}[f(X_t^{\mathrm{id}})] \ge 2(1-\theta(c)^2). \quad (3.42)$$

We will now show (3.42) by a coupling argument. Construct the two walks $X^{\tau_1\circ\tau_2}$ and X^{id} as follows. Let $\gamma_1, \gamma_2, \ldots$ be a sequence of i.i.d. random variables uniformly distributed on Γ, independent of (τ_1, τ_2). Using Lemma 3.2.4 with $\sigma_0 = \tau_1 \circ \tau_2$, which is independent of X^{id}, we can construct $X_t^{\tau_1\circ\tau_2}$ as

$$X_t^{\tau_1\circ\tau_2} = \gamma_1 \circ \cdots \circ \gamma_t \circ \tau_1 \circ \tau_2.$$

Next we couple X_t^{id} by constructing it as

$$X_t^{\mathrm{id}} = \gamma_1 \circ \cdots \circ \gamma_t.$$

Thus under this coupling we have that $X_t^{\tau_1\circ\tau_2} = X_t^{\mathrm{id}} \circ \tau_1 \circ \tau_2$. Let $X = X^{\mathrm{id}}$, then from (3.42) the problem reduces to showing

$$\liminf_{n\to\infty} \mathbb{E}[d(\mathrm{id}, X_t \circ \tau_1 \circ \tau_2) - d(\mathrm{id}, X_t)] \ge 2(1-\theta(c)^2). \quad (3.43)$$

Either τ_1 fragments a cycle of X_t or τ_1 coagulates two cycles of X_t. In the first case, $d(\mathrm{id}, X_t \circ \tau_1) = d(\mathrm{id}, X_t \circ \tau_1) - 1$, and in the second case we have $d(\mathrm{id}, X_t \circ \tau_1) = d(\mathrm{id}, X_t \circ \tau_1) + 1$. Let F denote the event that τ_1 causes a fragmentation. Then

$$\mathbb{E}[d(\mathrm{id}, X_t \circ \tau_1) - d(\mathrm{id}, X_t)] = 1 - 2\mathbb{P}(F).$$

Using the Poisson–Dirichlet structure described in Theorem 3.3.6 it is not hard to show that $\mathbb{P}(F) \to \theta(c)^2/2$ (see, e.g., Lemma 8 in Berestycki [10]). Applying the same reasoning to $X_t \circ \tau_1 \circ \tau_2$ and $X_t \circ \tau_1$ we deduce that

$$\lim_{n\to\infty} \mathbb{E}[d(\mathrm{id}, X_t \circ \tau_1 \circ \tau_2) - d(\mathrm{id}, X_t)] = 2(1-\theta(c)^2)$$

from which the lower bound (3.43) and in turn (3.12) follow readily.

3.4.2 Proof of lower bound on curvature.

We now assume that $c > c_\Gamma$ and turn out attention to the lower bound on the Ricci curvature, which is the heart of the proof. Throughout we let $k = |\Gamma|$ and $t = \lfloor cn/k \rfloor$. With this notation in mind we wish to prove that

$$\limsup_{n\to\infty} \sup_{\sigma,\sigma'} \frac{\mathbb{E}d(X_t^\sigma, X_t^{\sigma'})}{d(\sigma,\sigma')} \le \alpha := 1 - \theta(c)^4$$

for some appropriate coupling of X^σ and $X^{\sigma'}$, where the supremum is taken over all σ, σ' with even distance. Note that we can make several reductions: first, by vertex transitivity we can assume $\sigma = \mathrm{id}$ is the identity permutation. Also, by the triangle inequality (since W_1 is a distance), we can assume that $\sigma' = (i,j) \circ (\ell, m)$ is the product of two distinct transpositions. There are two cases to consider: either the supports of the transpositions are disjoint, or they overlap on one vertex. We will focus in this proof on the first case where the support of the transpositions are disjoint; that is, i, j, l, m are pairwise distinct. The other case is dealt with very much in the same way (and is in fact a bit easier).

Clearly by symmetry $\mathbb{E}d(X_t^{\mathrm{id}}, X_t^{(i,j)\circ(\ell,m)})$ is independent of i, j, ℓ and m, so long as they are pairwise distinct. Hence it is also equal to $\mathbb{E}d(X_t^{\mathrm{id}}, X_t^{\tau_1\circ\tau_2})$ conditioned on the event A that τ_1, τ_2 having disjoint support, where τ_1 and τ_2 are independent uniform random transpositions. This event has an overwhelming probability for large n, thus it suffices to construct a coupling between X^{id} and $X^{\tau_1\circ\tau_2}$ such that

$$\limsup_{n\to\infty} \mathbb{E}d(X_t^{\mathrm{id}}, X_t^{\tau_1\circ\tau_2}) \le 2(1-\theta(c)^4). \tag{3.44}$$

Indeed, it then immediately follows that the same is true with the expectation replaced by the conditional expectation given A.

Next, let X be a random walk on $\mathcal{S}_n$ which is the composition of i.i.d. uniform elements of the conjugacy class Γ. We decompose the random walk X into a walk $\tilde{X}$ which evolves by applying transpositions at each step as follows. For $t = 0, 1, \ldots,$ write out

$$X_t = \gamma_1 \circ \cdots \circ \gamma_t$$

where $\gamma_1, \gamma_2, \ldots$ are i.i.d. uniformly distributed in Γ. As before we decompose each step γ_s of the walk into a product of cyclic permutations, say

$$\gamma_s = \gamma_{s,1} \circ \ldots \circ \gamma_{s,r} \tag{3.45}$$

where $r = \sum_{j \geq 2} k_j$. The order of this decomposition is irrelevant and can be chosen arbitrarily. For concreteness, we decide that we start from the cycles of smaller sizes and progressively increase to cycles of larger sizes. We will further decompose each of these cyclic permutation into a product of transpositions, as follows: for a cycle $c = (x_1, \ldots, x_j)$, write

$$c = (x_1, x_2) \circ \ldots \circ (x_{j-1}, x_j).$$

This allows to break any step γ_s of the random walk X into a number

$$\rho := \sum_j (j-1)k_j$$

of elementary transpositions, and hence we can write

$$\gamma_s = \tau_s^{(1)} \circ \cdots \circ \tau_s^{(\rho)} \tag{3.46}$$

where $\tau_s^{(j)}$ are transpositions. Note that the vectors $(\tau_s^{(i)}; 1 \leq i \leq \rho)$ in (3.46) are independent and identically distributed for $s = 1, 2, \ldots$ and for a fixed s and $1 \leq i \leq \rho$, $\tau_s^{(i)}$ is a uniform transposition, by symmetry. However it is important to observe that they are *not* independent. Nevertheless, they obey a crucial conditional uniformity which we explain now. First we have differentiate between the set of times when a new cycle starts and the set of times when we are continuing an old cycle.

Definition 3.4.1 (Refreshment Times)**.** *We call a time s a refreshment time if s is of the form $s = \rho\ell + \sum_{j=2}^{m}(j-1)k_j$ for some $\ell \in \mathbb{N} \cup \{0\}$ and $m \in \mathbb{N} \backslash \{1\}$.*

We see that s is a refreshment time if the transposition being applied to $\tilde{X}$ at time s is the start of a new cycle. Using this we can describe the law of the transpositions being applied to $\tilde{X}$.

Proposition 3.4.2 (Conditional Uniformity)**.** *For $s \in \mathbb{N}$ and $i \leq \rho$, the conditional distribution of $\tau_s^{(i)}$ given $\tau_s^{(1)}, \ldots, \tau_s^{(i-1)}$ can be described as follows. We write $\tau_s^{(i)} = (x, y)$ and we will distinguish between the first marker x and the second marker y. There are two cases to consider:*

(i) $s\rho + i$ is a refreshment time and thus $\tau_s^{(i)}$ corresponds to the start of a new cycle

(ii) $s\rho + i$ is not a refreshment time and so $\tau_s^{(i)}$ is the continuation of a cycle.

In case (i) x is uniformly distributed on $S_i := \{1, \ldots, n\} \setminus Supp(\tau_s^{(1)} \circ \ldots \circ \tau_s^{(i-1)})$ and y is uniformly distributed on $S_i \setminus \{x\}$. In case (ii) x is equal to the second marker of $\tau_s^{(i-1)}$ and y is uniformly distributed in S_i.

Note that in either case, the second marker y is conditionally uniformly distributed among the vertices which have not been used so far. This conditional independence property is completely crucial, and allows us to make use of methods (such as that of Schramm [47]) developed initially for random transpositions) for general conjugacy classes, so long as $|\Gamma| = o(n)$. Indeed in that case the second marker y itself is not very different from a uniform random variable on $\{1, \ldots, n\}$.

We will study this random walk using this new *transposition time scale.* We thus define a process $\tilde{X} = (\tilde{X}_u : u = 0, 1, \ldots)$ as follows. Let $u \in \{0, 1, \ldots\}$ and write $u = s\rho + i$ where s, i are nonnegative integers and $i < \rho$. Then define

$$\tilde{X}_u := X_s \circ \tau_{s+1}^{(1)} \circ \cdots \circ \tau_{s+1}^{(i)}. \tag{3.47}$$

Thus it follows that for any $s \geq 0$, $\tilde{X}_{s\rho} = X_s$. Notice that $\tilde{X}$ evolves by applying successively transpositions with the above mentioned conditional uniformity rules.

Now consider our two random walks, X^{id} and $X^{\tau_1 \circ \tau_2}$ respectively, started respectively from id and $\tau_1 \circ \tau_2$, and let $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ be the associated processes constructed using (3.47), on the transposition time scale. Thus to prove (3.44) it suffices to construct an appropriate coupling between $\tilde{X}^{\mathrm{id}}_{t\rho}$ and $\tilde{X}^{\tau_1 \circ \tau_2}_{t\rho}$. Next, recall that for a permutation $\sigma \in \mathcal{S}_n$, $\mathfrak{X}(\sigma)$ denotes the renormalised cycle lengths of σ, taking values in Ω_∞ defined in (3.40). The walks $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ are invariant by conjugacy and hence both are distributed uniformly on their conjugacy class. Thus ultimately it will suffice to couple $\mathfrak{X}(\tilde{X}^{\mathrm{id}}_{t\rho})$ and $\mathfrak{X}(\tilde{X}^{\tau_1 \circ \tau_2}_{t\rho})$.

Fix $\delta > 0$ and let $\Delta = \lceil \delta^{-9} \rceil$. Define

$$\begin{aligned} s_1 &= \lfloor (cn - \delta^{-9})/k \rfloor \rho \\ s_2 &= s_1 + \Delta \\ s_3 &= t\rho. \end{aligned}$$

Our coupling consists of three intervals $[0, s_1]$, $(s_1, s_2]$ and $(s_2, s_3]$.

Let us informally describe the coupling before we give the details. In what follows we will couple the random walks $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ such that they keep their distance constant during the time intervals $[0, s_1]$ and $(s_2, s_3]$. In particular we will see that at time s_1, the walks $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ will differ by two independently uniformly chosen transpositions. Thus at time s_1 most of the cycles of $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ are identical but some cycles may be different. We will show that given that the cycles that differ at time s_1 are all reasonably large, then we can reduce the distance between the two walks to zero during the time interval $(s_1, s_2]$. Otherwise if one of the differing cycles is not reasonably large, then we couple the two walks to keep their distance constant during the time interval $[0, s_1]$, $(s_1, s_2]$ and $(s_2, s_3]$.

More generally, our coupling has the property that $d(X_t^{\mathrm{id}}, X_t^{\tau_1 \circ \tau_2})$ is uniformly bounded, so that it will suffice to concentrate on events of high probability in order to get a bound on the L^1-Kantorovitch distance $W(X_t^{\mathrm{id}}, X_t^{\tau_1 \circ \tau_2})$.

Coupling for $[0, s_1]$

First we describe the coupling during the time interval $[0, s_1]$. Let $\tilde{X} = (\tilde{X}_s : s \geq 0)$ be a walk with the same distribution as $\tilde{X}^{\mathrm{id}}$, independent of the two uniform transpositions τ_1 and τ_2. Then we have that by Lemma 3.2.4 for any $s \geq 0$, $\tilde{X}_s^{\tau_1 \circ \tau_2}$ has the same distribution as $\tilde{X}_s \circ \tau_1 \circ \tau_2$. Thus we can couple $\mathfrak{X}(\tilde{X}_{s_1}^{\mathrm{id}})$ and $\mathfrak{X}(\tilde{X}_{s_1}^{\tau_1 \circ \tau_2})$ such that

$$\begin{aligned} \mathfrak{X}(\tilde{X}_{s_1}^{\mathrm{id}}) &= \mathfrak{X}(\tilde{X}_{s_1}) \\ \mathfrak{X}(\tilde{X}_{s_1}^{\tau_1 \circ \tau_2}) &= \mathfrak{X}(\tilde{X}_{s_1} \circ \tau_1 \circ \tau_2). \end{aligned} \tag{3.48}$$

Coupling for $(s_1, s_2]$

For $s \geq 0$ define $\bar{X}_s = \mathfrak{X}(\tilde{X}_{s+s_1}^{\mathrm{id}})$ and $\bar{Y}_s = \mathfrak{X}(\tilde{X}_{s+s_1}^{\tau_1 \circ \tau_2})$. Here we will couple $\bar{X}_s$ and $\bar{Y}_s$ for $s = 0, \ldots, \Delta$. We create a matching between $\bar{X}_s$ and $\bar{Y}_s$ by matching an element of $\bar{X}_s$ to at most one element of $\bar{Y}_s$ of the same size. At any time s there may be several entries that cannot be matched. By parity the combined number of unmatched entries is an even number, and observe that this number cannot be equal to two. Now $\tilde{X}_{s_1}^{\mathrm{id}}$ and $\tilde{X}_{s_1}^{\tau_1 \circ \tau_2}$ differ by two transpositions as can be seen from (3.48). This implies that in particular initially (i.e., at the beginning of $(s_1, s_2]$), there are four, six or zero unmatched entries between $\bar{X}_0$ and $\bar{Y}_0$.

Fix $\delta > 0$ and let $A(\delta)$ denote the event that the smallest unmatched between $\bar{X}_0$ and $\bar{Y}_0$ has size greater than $\delta > 0$. We will show that on the event $A(\delta)$ we can couple the walks such that $\bar{X}_\Delta = \bar{Y}_\Delta$ with high probability. On the complementary event $A(\delta)^c$, couple the walks so that their distance remains 1 during the time interval $(s_1, s_2]$, similar to the coupling during $[0, s_1]$.

It remains to define the coupling during the time interval $(s_1, s_2]$ on the event $A(\delta)$.

We begin by estimating the probability of $A(\delta)$.

Lemma 3.4.3. *For any $c > 1$ and $\delta > 0$,*

$$\liminf_{n\to\infty} \mathbb{P}(A(\delta)) \geq [\theta(c)(1-\delta)]^4.$$

Proof. Recall that by construction $\bar{X}_0$ and $\bar{Y}_0$ only differ because of the two transpositions τ_1 and τ_2 appearing in (3.48).

Recall the hypergraph $H_{s_1/\rho}$ on $\{1,\ldots,n\}$ defined in the beginning of Section 3.3.1. Since $c > c_\Gamma$, $H_{s_1/\rho}$ has a (unique) giant component with high probability. Let A_1 be the event that the four points composing the transpositions τ_1, τ_2 fall within the largest component of the associated hypergraph $H_{s_1/\rho}$. It follows from Theorem 3.3.6 that conditionally on the event A_1, $A(\delta)$ has probability greater than $(1-\delta)^4$. Also, since the relative size of the giant component converges in probability $\theta(c)$ by Lemma 3.3.2, it is obvious that $\mathbb{P}(A_1) \to \theta(c)^4$ and thus the lemma follows. □

Recall that the transpositions which make up the walks $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1\circ\tau_2}$ obey what we called conditional uniformity in Proposition 3.4.2. For the duration of $(s_1, s_2]$ we will assume the *relaxed conditional uniformity* assumption, which we describe now.

Definition 3.4.4 (Relaxed Conditional Uniformity)**.** *For $s = s_1 + 1, \ldots, s_2$ suppose we apply the transposition (x, y) at time s. Then*

(i) if s is a refreshment time then x is chosen uniformly in $\{1,\ldots,n\}$,

(ii) if s is not a refreshment time then x is taken to be the second marker of the transposition applied at time $s-1$.

In both cases we take y to be uniformly distributed on $\{1,\ldots,n\}\backslash\{x\}$.

In making the relaxed conditional uniformity assumption we are disregarding the constraints on (x, y) given in Proposition 3.4.2. However the probability we violate this constraint at any point during the interval $(s_1, s_2]$ is at most $2(s_2 - s_1)\rho/n = 2\Delta k/n$ and on the event that this constraint is violated the distance between the random walks can increase by at most $(s_2 - s_1) = \Delta$. Hence we can without a loss of generality assume that during the interval $(s_1, s_2]$ both $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1\circ\tau_2}$ satisfy the relaxed conditional uniformity assumption.

Now we show that on the event $A(\delta)$ we can couple the walks such that $\bar{X}_\Delta = \bar{Y}_\Delta$ with high probability. The argument uses a coupling of Berestycki, Schramm, and Zeitouni [12], itself a variant of a beautiful coupling introduced by Schramm [47]. We first introduce some notation. Let

$$\Omega_n := \{(x_1 \geq \cdots \geq x_n) : x_i \in \{0/n, 1/n, \ldots, n/n\} \text{ for each } i \leq n \text{ and } \sum_{i \leq n} x_i = 1\}.$$

Notice that the walks $\bar{X}$ and $\bar{Y}$ both take values in Ω_n.

Let us describe the evolution of the random walk $\bar{X} = (\bar{X}_s : s = 0, 1, \ldots)$. Suppose that $s \geq 0$ and $\bar{X}_s = \bar{x} = (x_1, \ldots, x_n)$. Now imagine the interval $(0, 1]$ tiled using the intervals $(0, x_1], \ldots, (0, x_n]$ (the specific tiling rule does not matter). Initially for $s = 0$ we select $u \in \{1/n, \ldots, n/n\}$ uniformly at random and then call the tile that u falls into *marked*. Next if $s \geq 1$ is not a refreshment time then we keep marked the tile which was marked in the previous step. Otherwise if $s \geq 1$ is a refreshment time we select a new marked tile by selecting $u \in \{1/n, \ldots, n/n\}$ uniformly at random and marking the tile which u falls into.

Let I be the marked tile. Select $v \in \{2/n, \ldots, n/n\}$ uniformly at random and let I' be the tile that v falls in. Then if $I' \neq I$ then we merge the tiles I and I'. The new tile we created is now marked. If $I = I'$ then we split I into two tiles, one of size $v - 1/n$ and the other of size $|I| - (v - 1/n)$. The tile of size v is now marked. Now $\bar{X}_{s+1}$ is the sizes of the tiles in the new tiling we have created, ordered in decreasing order.

The evolution of $\bar{X}$ described above corresponds to the evolution of X as follows. Suppose we apply the transposition (x, y) to X_s in order to obtain X_{s+1}. The marked tile at time s corresponds to the cycle of X_s containing x: if s is a refreshment time then $x \in \{1, \ldots, n\}$ is chosen uniformly, otherwise x is the second marker from the previous step. Then we write the cycle containing x as $(x, x_1, \ldots, x_m)$ and so the point x corresponds to $1/n$ in the tiling. Then we select the second marker $y \in \{1, \ldots, n\}\backslash\{x\}$ uniformly which corresponds to the selection of the marker $v \in \{2/n, \ldots, n/n\}$.

Before we describe the coupling in detail let us make a remark. In the course of the coupling there may be several things that may go wrong; for example the size of the smallest unmatched component may become too small. We will estimate the probability of such unfortunate events and see that these tend to zero when we take $n \to \infty$ and then $\delta \to 0$. The coupling which we describe keeps the distance between walks X^{id} and $X^{\tau_1 \circ \tau_2}$ bounded by 4, hence we can safely ignore these unfortunate events.

We now recall the coupling of Berestycki, Schramm, and Zeitouni [12]. Let $s \geq 0$. Suppose that $\bar{X}_s = \bar{x} = (x_1, \ldots, x_n)$ and $\bar{Y}_s = \bar{y} = (y_1, \ldots, y_n)$. Then we can differentiate between the entries that are matched and those that are unmatched: recall that two entries from $\bar{x}$ and $\bar{y}$ are matched if they are of identical size. Our goal will be to create as many matched parts as possible and as quickly as possible. Let Q be the total mass of the unmatched parts in either $\bar{x}$ or $\bar{y}$. When putting down the tilings $\tilde{x}$ and $\tilde{y}$, associated

 86

with $\bar{x}$ and $\bar{y}$ respectively, we will do so in such a way that all matched parts are at the right of the interval $(0, 1]$ and the unmatched parts occupy the left part of the interval. Initially for $s = 0$ suppose that $u \in \{1/n, \ldots, n/n\}$ is chosen uniformly and call the tile that u falls into in each of $\tilde{x}$ and $\tilde{y}$, *marked*. As before if $s \geq 1$ is not a refreshment time then we keep marked the tiles which were marked in the previous step. Otherwise if $s \geq 1$ is a refreshment time we select new marked tiles in both $\tilde{x}$ and $\tilde{y}$ by selecting $u \in \{1/n, \ldots, n/n\}$ uniformly at random and marking the tiles which u falls into in each of $\tilde{x}$ and $\tilde{y}$.

Let $I_{\bar{x}}$ and $I_{\bar{y}}$ be the respective marked tiles of the tilings $\tilde{x}$ and $\tilde{y}$, and let $\hat{x}, \hat{y}$ be the tiling which is the reordering of $\tilde{x}, \tilde{y}$ in which $I_{\bar{x}}$ and $I_{\bar{y}}$ have been put to the left of the interval $(0, 1]$. Let $a = |I_{\bar{x}}|$ and let $b = |I_{\bar{y}}|$ be the respective lengths of the marked tiles, and assume without loss of generality that $a < b$. Let $v \in \{2/n, \ldots, n/n\}$ be chosen uniformly. We will apply v to $\hat{x}$ as we did in the transition above and obtain $\bar{X}_{s+1}$. We now describe how construct an other uniform random variable $v' \in \{2/n, \ldots, n/n\}$ which will be applied to $\hat{y}$. If $I_{\bar{x}}$ is matched (which implies that $I_{\bar{y}}$ is also matched) then we take $v' = v$ as in the coupling of Schramm [47]. In the case when $I_{\bar{x}}$ is unmatched (which implies $I_{\bar{y}}$ is also unmatched) in the coupling of Schramm [47] one again takes $v = v'$, here we do not take them equal and apply to v a measure-preserving map Φ, defined as follows.

For $w \in \{2/n, \ldots, n/n\}$ consider the map

$$\Phi(w) = \begin{cases} w & \text{if } w > b \text{ or if } 1/n \leq w \leq \gamma_n + 1/n, \\ w - \gamma_n & \text{if } a < w \leq b, \\ w + b - a & \text{if } \gamma_n + 1/n < w \leq a, \end{cases} \tag{3.49}$$

where $\gamma_n := \lceil an/2 - 1 \rceil / n$. It is not hard to check that Φ is measure preserving, thus letting $v' = \Phi(v)$ we have that v' has the correct marginal distribution.

If $v \notin I_{\bar{x}}$ then we merge the tile containing v and $I_{\bar{x}}$. The new tile is now marked. If $v \in I_{\bar{x}}$ we split the tile $I_{\bar{x}}$ into two tiles, one of length $v - 1/n$ and one of length $a - (v - 1/n)$. We mark the tile of size $v - 1/n$. Now $\bar{X}_{s+1}$ is the sizes of the tiles in the new tiling we have created, ordered in decreasing order. We obtain $\bar{Y}_{s+1}$ from the same procedure as we did to obtain $\bar{X}_{s+1}$, but we use v' instead of v. We give an example of an evolution under this coupling in Figure 3.1

The somewhat remarkable property of this coupling is that the number of unmatched entries can only decrease. Unmatched entries disappear when they are coalesced. In particular they disappear quickly when their size is reasonably large. Hence it is particularly desirable to have a coupling in which unmatched components stay large. The second

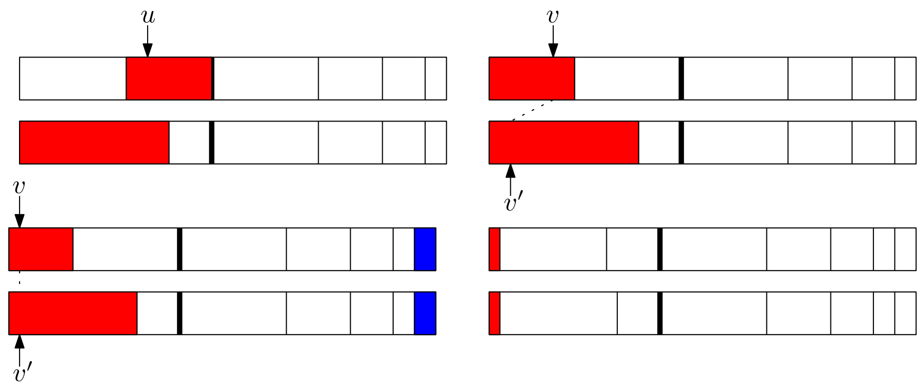


Figure 3.1: The evolution under the coupling between $\bar{X}$ and $\bar{Y}$. The red entries represent the marked entries.

crucial property of this coupling is that it does not create arbitrarily small unmatched entries: even when unmatched entry is fragmented, the size of the smallest unmatched entry cannot decrease by more than a factor of two. This is summarised by the following, which is Lemma 19 from Berestycki, Schramm, and Zeitouni [12].

Lemma 3.4.5. *Let U be the size of the smallest unmatched entry in two partitions $\bar{x}, \bar{y} \in \Omega_n$, let $\bar{x}', \bar{y}'$ be the corresponding partitions after one transposition of the coupling, and let U' be the size of the smallest unmatched entry in $\bar{x}', \bar{y}'$. Assume that $2^j \leq U < 2^{j+1}$ for some $j \geq 0$. Then it is always the case that $U' \geq U/2 - 1/n$, and moreover,*

$$\mathbb{P}(U' \leq 2^j) \leq 2^{j+2}/n.$$

Finally, the combined number of unmatched parts may only decrease.

Remark 3.4.6. *In particular, it holds that $U' \geq 2^{j-1}/n$.*

We now explain our strategy. On $A(\delta)$ we will expect that the unmatched components will remain of a size roughly of order at least δ for a while. In fact we will show that they will stay at least as big as $O(\delta^2)$ for a long time. Unmatched entries disappear when they are merged together. If all unmatched entries are of size at least δ^2, we will see that with probability at least δ^8, we have a chance to reduce the number of unmatched entries in every 4 steps. Then a simple argument shows that after time $\Delta = \lceil \delta^{-9} \rceil$, $\bar{X}_\Delta$ and $\bar{Y}_\Delta$ are perfectly matched with a probability tending to one as $\delta \to 0$.

Lemma 3.4.7. *There is δ_0 such that if $\delta < \delta_0$, during $[0, \Delta]$, both $\bar{X}_s$ and $\bar{Y}_s$ always have an entry of size greater than $\delta\theta(c)$ with probability at least $1 - 2\delta^{1/2}$ for all n sufficiently large.*

Proof. Let $\delta_0 > 0$ be such that $(1 - \delta_0)^{9!} \geq \delta_0^{1/2}$ and assume that $\delta < \delta_0$. Hence it also true that $(1 - \delta)^{9!} \geq \delta^{1/2}$. Let $Z = (Z_1, \ldots)$ be a Poisson-Dirichlet random variable on

Ω_∞ and let $(Z_1^*, \dots)$ denote the size biased ordering of Y. Recall that Z_1^* is uniformly distributed over $[0,1]$, Z_2^* is uniformly distributed on $[0, 1-Z_1^*]$, and so on. For the event $\{Z_1 \le \delta\}$ to occur it is necessary that $Z_1^* \le \delta, Z_2^* \le \delta/(1-\delta), \dots, Z_{10}^* \le \delta/(1-\delta)^9$. This has probability at most $\delta^{10}/(1-\delta)^{9!}$. Note that since $\delta < \delta_0$, we have that $(1-\delta)^{9!} \ge \delta^{1/2}$. Thus

$$\mathbb{P}(Z_1 \le \delta) \le \frac{\delta^{10}}{(1-\delta)^{9!}} \le \delta^{9+1/2}.$$

Summing over $\Delta = O(\delta^{-9})$ steps we see that the expected number of times during the interval $[0, \Delta]$ such that $\bar{X}_s$ or $\bar{Y}_s$ don't have a component of size at least $\theta(c)\delta n$ is less than $\delta^{1/2}$ as $n \to \infty$ and is thus less than $2\delta^{1/2}$ for n sufficiently large, by Theorem 3.3.6 (note that we can apply the result because this calculation involves only a finite number of components). The result follows. □

We now check that all unmatched components really do stay greater than δ^2 during $[0, \Delta]$. Let T_δ denote the first time s that either $\bar{X}_s$ or $\bar{Y}_s$ have no cycles greater than $\delta\theta(c)n$.

Lemma 3.4.8. *On $A(\delta)$, for all $s \le T_\delta \wedge \Delta$, all unmatched components stay greater than δ^2 with probability at least $1 - \delta(16/\theta(c))^{10}$.*

Proof. Say that an integer k is in scale j if $2^j/n \le k < 2^{j+1}/n$. For $s \ge 0$, let $U(s)$ denote the scale of the smallest unmatched entry of $\bar{X}_s, \bar{Y}_s$. Let j_0 be the scale of δ, and let j_1 be the integer immediately above the scale of δ^2.

Suppose for some time $s \le T_\delta$, we have $U(s) = j$ with $j_1 \le j \le j_0$, and the marked tile at time s corresponds to the smallest unmatched entry. Then after this transposition we have $U(s+1) \ge j-1$ by the properties of the coupling (Lemma 3.4.5). Moreover, $U(s+1) = j-1$ with probability at most $r_j = 2^{j+2}/n$. Furthermore, since $s \le T_\delta$, we have that this marked tile merges with a tile of size at least $\theta(c)\delta$ with probability at least $\theta(c)\delta$ after the transposition. We call the first occurrence a *failure* and the second a *mild success.*

Once a mild success has occurred, there may still be a few other unmatched entries in scale j, but no more than five since the total number of unmatched entries is decreasing. And therefore if six mild successes occur before a failure, we are guaranteed that $U(s+1) \ge j+1$. We call such an event a *good success*, and note that the probability of a good success, given that $U(s)$ changes scale, is at least $p_j = 1 - 6r_j/(r_j + \theta(c)\delta)$. We call $q_j = 1 - p_j$.

Let $\{q_i\}_{i \ge 0}$ be the times at which the smallest unmatched entry changes scale, with q_0 being the first time the smallest unmatched entry is of scale j_0. Let $\{U_i\}$ denote the scale of the smallest unmatched entry at time q_i. Introduce a birth-death chain on the integers, denoted v_n, such that $v_0 = j_0$ and

$$\mathbb{P}(v_{n+1} = j-1 | v_n = j) = \begin{cases} 1 & \text{if } j = j_0 \\ 0 & \text{if } j = j_1 \\ q_j & \text{otherwise,} \end{cases} \tag{3.50}$$

and

$$\mathbb{P}(v_{n+1} = j+1 | v_n = j) = \begin{cases} p_j, & j > j_1 \\ 0, & j = j_1. \end{cases} \tag{3.51}$$

Then it is a consequence of the above observations that $(U_i, i \geq 1)$ is stochastically dominating $(v_i, i \geq 1)$ for $s \leq T_\delta$. Set $\tau_j = \min\{n > 0 : v_n = j\}$. An analysis of the birth-death chain defined by (3.50), (3.51) gives that

$$\mathbb{P}^{j_0}(\tau_{j_1} < \tau_{j_0}) = \frac{1}{\sum_{j=j_1+1}^{j_0} \prod_{m=j}^{j_0-1} \frac{p_m}{q_m}} \leq \prod_{j=j_1+1}^{j_0-1} \frac{q_j}{p_j}$$

(see, e.g., Theorem (3.7) in Chapter 5 of Durrett [23]). Thus, by considering the 10 lowest terms in the product above (and note that for $\delta > 0$ small enough, there are at least 10 terms in this product), we deduce that $\mathbb{P}^{j_0}(\tau_{j_1} < \tau_{j_0})$ decays faster than $(16\delta/\theta(c))^{10}$. Since $T_\delta \wedge \Delta \leq \Delta = O(\delta^{-9})$ we conclude that the probability that $U(s) = j_1$ before $T_\delta \wedge \Delta$ is at most $\delta(16/\theta(c))^{10}$. □

We are now going to prove that on the event $A(\delta)$, after time Δ there are no unmatched with probability tending to one as $n \to \infty$ and $\delta \to 0$. The basic idea is that there are initially at most six unmatched parts, and this number cannot increase.

Lemma 3.4.9. *We have that for all $\delta > 0$ small enough*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(\bar{X}_\Delta \neq \bar{Y}_\Delta | A(\delta)) = 0$$

Proof. Suppose $\delta > 0$ is sufficiently small and condition throughout on the event $A(\delta)$. Let T'_δ be the first time one of the unmatched entries is smaller than δ^2 or T_δ, whichever comes first. By Lemma 3.4.7 and Lemma 3.4.8 we have that for large n,

$$\mathbb{P}(T'_\delta \geq \Delta | A(\delta)) \geq 1 - \delta(16/\theta(c))^{10} - 2\delta^{1/2}. \tag{3.52}$$

Henceforth condition on the event $\{T'_\delta \geq \Delta\}$. Initially there are at most 6 unmatched entries. Due to parity there can be either $6, 4$ or 0 unmatched entries (note in particular that 2 is excluded, as a quick examination shows that no configuration can give rise to two unmatched entries). Furthermore, by the virtue of the coupling the number of unmatched entries either remains the same or decreases sequentially. Once all the entries are matched they remain matched thereon. In order for the unmatched entries to decrease at time $s \in \{2, \ldots, \Delta\}$ it must be the case that *both* $\bar{X}_s$ and $\bar{Y}_s$ must have at least 2 unmatched entries. Call this a good configuration. Let F_s be the event that at time s the configuration is good and one of the two marked tiles at time s is the smallest unmatched tile.

We now show that $\mathbb{P}(F_s) \geq \delta^4/2$ by considering different cases:

- Suppose that at time $s-1$ the configuration is good. Then placing the second marker (v or v') inside the smallest unmatched tile will guarantee that at time s the configuration is still good. Suppose without loss of generality that v lands in the smallest unmatched tile, then it could be the case that at time $s-1$ the smallest unmatched tile was marked. In this case the smallest unmatched tile will fragment into two and the smaller of the two pieces will be matched and the resulting tile on the left will be marked. If a is the size of the smallest entry and $v \in [a/2, a]$ then both marked tiles at time s will be unmatched and furthermore of them will correspond to the smallest unmatched entry at time s. Hence the probability that F_s holds in this case is at least $\delta^2/2$.

- Suppose that the configuration at time $s-1$ is bad: that is, one copy has one unmatched entry and the other copy has either three or five unmatched entries. Suppose, without a loss of generality that $\bar{X}_{s-1}$ has one unmatched entry which means that $\bar{Y}_{s-1}$ has at least three unmatched entries. To get to a good configuration at time s it suffices to coagulate two of the unmatched entries of $\bar{Y}_{s-1}$ (as then automatically, by the properties of the coupling, the single unmatched entry in $\bar{X}_{s-1}$ fragments into two). In order for this to happen, the marked tiles at time $s-1$ must be unmatched. We force the marked entries at time $s-1$ to be unmatched as follows.

  - If $s-1$ is a refreshment time then we ask that the marker u at time $s-1$ falls inside an unmatched tile which is *not* the smallest unmatched tile. This happens with probability at least δ^2.

  - If $s-1$ is not a refreshment time then we ask for the marker v and v' at time $s-2$ to fall inside an unmatched tile which is *not* the smallest unmatched tile. This happens with probability at least δ^2. As before, once the markers v and v' fall inside unmatched tiles the probability that the marked tile at time $s-1$ is unmatched is $1/2$.

  Suppose now that at $s-1$ the marked tiles are unmatched but neither is the smallest unmatched tile. If the marker v or the marker v' at time $s-1$ falls inside the smallest unmatched tile then we are guaranteed that F_s holds and this happens with probability at least δ^2. Hence we see that the probability that F_s holds when the configuration at time $s-1$ is bad is at least $\delta^4/2$.

We have just shown that $\mathbb{P}(F_s) \geq \delta^4/2$.

Now suppose F_s holds. With probability greater than δ^2 we have that one of the marked tiles at time s is the smallest unmatched tile (in fact the probability is 1 if s is not a refreshment time). Since there are at least 2 unmatched parts in each copy, let R be the tile corresponding to a second unmatched tile in the copy that contains the larger of the two marked tiles. Then $|R| > \delta^2$, and moreover when v falls in R, we are guaranteed that a coagulation is going to occur in both copies hence decreasing the total number of unmatched entries. Let K_s denote this event and call this a success. Thus we have just shown that $\mathbb{P}(K_s|F_s) \geq \delta^4$.

Notice that for $s \in \{2, \dots, \Delta\}$ we have that the marked tiles and the markers (v, v') used in the transition from time $s+1$ to $s+2$ are independent from $\mathcal{F}_s := \sigma((\bar{X}_\ell, \bar{Y}_\ell) : \ell \leq s)$. Thus we can repeat the same argument as before to obtain that for any $s \in \{1, \dots, \lfloor (\Delta-1)/4 \rfloor\}$ we have that $\mathbb{P}(K_{4s} \cap F_{4s} | \mathcal{F}_{4s-2}) \geq \delta^8/2$. Hence it follows that the number of successes before time Δ stochastically dominates a random variable H which has the binomial distribution $\mathrm{Bin}(\lfloor (\Delta-1)/4 \rfloor, \delta^8/2)$. The event that $\{X_\Delta \neq Y_\Delta\}$ implies that there has been at most one success. Thus for $\delta > 0$ small enough

$$\mathbb{P}(\bar{X}_\Delta \neq \bar{Y}_\Delta | A(\delta) \cap \{T'_\delta \geq \Delta\}) \leq \mathbb{P}(H \leq 1) \leq \Delta(1-\delta^8/2)^{\lfloor (\Delta-1)/4 \rfloor}.$$

As $\Delta = O(\delta^{-9})$, the right hand side of the equation above converges to 0 as $\delta \downarrow 0$ and using (3.52) finishes the proof. □

Coupling for $(s_2, s_3]$

The walks $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ are uniformly distributed on their conjugacy class. Thus one can couple $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1 \circ \tau_2}$ so that

- on the event $A(\delta)^c$ we have that $d(\tilde{X}^{\mathrm{id}}_{s_2}, \tilde{X}^{\tau_1 \circ \tau_2}_{s_2}) = 2$,
- we have that using Lemma 3.4.9

$$\liminf_{\delta \downarrow 0} \liminf_{n \to \infty} \mathbb{P}(\tilde{X}^{\mathrm{id}}_{s_2} = \tilde{X}^{\tau_1 \circ \tau_2}_{s_2} | A(\delta)) = 1,$$

- on the event $\{\tilde{X}^{\mathrm{id}}_{s_2} \neq \tilde{X}^{\tau_1 \circ \tau_2}_{s_2}\}$, note that the walks $\bar{X}$ and $\bar{Y}$ have at most 6 unmatched entries. Hence there exists a coupling such that $d(\tilde{X}^{\mathrm{id}}_{s_2}, \tilde{X}^{\tau_1 \circ \tau_2}_{s_2}) \leq 4$.

Combining this with Lemma 3.4.3 we have just shown the following lemma.

Lemma 3.4.10. *There exists a coupling of $\tilde{X}^{id}$ and $\tilde{X}^{\tau_1\circ\tau_2}$ such that*

$$\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{E}[d(\tilde{X}^{id}_{s_2},\tilde{X}^{\tau_1\circ\tau_2}_{s_2})]\leq 2(1-\theta(c)^4)$$

The theorem now follows immediately.

Proof of Theorem 3.1.3. It remains to see the coupling during the time interval $(s_2, s_3]$. During this time interval we apply the same transpositions to both $\tilde{X}^{\mathrm{id}}$ and $\tilde{X}^{\tau_1\circ\tau_2}$ which keeps their distance constant throughout $(s_2, s_3]$. Thus we have that

$$d(X^{\mathrm{id}}_t, X^{\tau_1\circ\tau_2}_t) = d(\tilde{X}^{\mathrm{id}}_{s_3},\tilde{X}^{\tau_1\circ\tau_2}_{s_3}) = d(\tilde{X}^{\mathrm{id}}_{s_2},\tilde{X}^{\tau_1\circ\tau_2}_{s_2}).$$

Thus using Lemma 3.4.10 we see that (3.44) holds which finishes the proof. □

3.5 Appendices

3.5.1 Lower bound on mixing

In this section we give a proof of the lower bound on $t_{\mathrm{mix}}(\delta)$ for some arbitrary $\delta\in(0,1)$. This is for the most part a well-known argument, which shows that the number of fixed points at time $(1-\epsilon)t_{\mathrm{mix}}$ is large. In the case of random transpositions or more generally of a conjugacy class Γ such that $|\Gamma|$ is finite, this follows easily from the coupon collector problem. When $|\Gamma|$ is allowed to grow with n, we present here a self-contained argument for completeness.

Let $\Gamma\subset\mathcal{S}_n$ be a conjugacy class and set $k=k(n)=|\Gamma|$.

Lemma 3.5.1. *We have that for any $\epsilon\in(0,1)$,*

$$\lim_{n\to\infty} d_{TV}((1-\epsilon)t_{mix}) = 1$$

Proof. Let $K_m\subset\mathcal{S}_n$ be the set of permutations which have at least m fixed points. Recall that μ is the invariant measure, which is a uniform probability measure on $\mathcal{S}_n$ or $\mathcal{A}_n$ depending on the parity of Γ. Let U denote the uniform measure on $\mathcal{S}_n$. Either way,

$$\mu(K_m)\leq 2U(K_m).$$

Now, $U(K_m) \to \sum_{j=m}^{\infty} e^{-1}\frac{1}{j!}$ as $n \to \infty$, hence we deduce that

$$\limsup_{m\to\infty} \limsup_{n\to\infty} \mu(K_m) = 0. \tag{3.53}$$

Fix $\beta > 0$ and let

$$t_\beta = \frac{1}{k} n(\log n - \log \beta).$$

Assume that β is such that t_β is an integer. For each $i \geq 0$, γ_i write $N(\gamma_i) \subset \{1, \dots, n\}$ for the set of non-fixed points of γ_i. Then we have that for each $i \geq 0$, $|N(\gamma_i)| = k$ and further $\{N(\gamma_i)\}_{i=1}^{\infty}$ are i.i.d. subsets of $\{1, \dots, n\}$ chosen uniformly among the subsets of size $k = |\Gamma|$.

Consider for $1 \leq i \leq n$ the event A_i that the i-th card is not collected by time t_β, that is $i \notin \bigcup_{\ell=1}^{t_\beta} N(\gamma_\ell)$. Thus for $1 \leq i_1 < \cdots < i_\ell \leq n$ and $\ell \leq n - k$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_\ell}) = \left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}} \right)^{t_\beta}.$$

Let $N = N(n) \in \mathbb{N}$ be increasing to infinity such that $N^2 = o(n)$ and $N = o(n^2 k^{-2})$. Then by the inclusion-exclusion formula we have that

$$\mathbb{P}(A_1 \cup \cdots \cup A_N) = \sum_{\ell=1}^{N} (-1)^{\ell+1} \binom{n}{\ell} \left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}} \right)^{t_\beta}. \tag{3.54}$$

Writing out the fraction of binomials on the right hand side we have

$$\left(1 - \frac{k}{n-\ell}\right)^{\ell t_\beta} \leq \left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}} \right)^{t_\beta} \leq \left(1 - \frac{k}{n}\right)^{\ell t_\beta}.$$

Now $-x/(1-x) \leq \log(1-x) \leq -x$ for $x \in (0,1)$ thus we have that

$$\exp\left(-\frac{\ell k t_\beta}{n-k-\ell}\right) \leq \left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}} \right)^{t_\beta} \leq \exp\left(-\frac{\ell k t_\beta}{n}\right). \tag{3.55}$$

On the other hand we have that

$$\frac{(n-\ell)^\ell}{\ell!} \leq \binom{n}{\ell} \leq \frac{n^\ell}{\ell!}. \tag{3.56}$$

Note that $ne^{-t_\beta k/n} = \beta$, then combining (3.55) and (3.56) we get

$$\left(1-\frac{\ell}{n}\right)^{\ell} \exp\left(-\frac{k(k+\ell)\ell t_\beta}{n(n-k-\ell)}\right)\frac{\beta^\ell}{\ell!} \leq \binom{n}{\ell}\left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}}\right)^{t_\beta} \leq \frac{\beta^\ell}{\ell!}. \tag{3.57}$$

Let us lower bound the error term on the left hand side of (3.57). First $(1-\ell/n)^\ell \geq e^{-\ell^2/(n-\ell)}$, hence it follows that

$$\inf_{\ell\leq N}\left(1-\frac{\ell}{n}\right)^{\ell} \exp\left(-\frac{k(k+\ell)\ell t_\beta}{n(n-k-\ell)}\right) \geq \inf_{\ell\leq N} \exp\left(-\frac{\ell^2}{n-\ell} - \frac{k(k+\ell)\ell t_\beta}{n(n-k-\ell)}\right).$$

It is easy to see that the right hand side above converges to 1 as $n\to\infty$. Using this and (3.57) it follows that

$$\lim_{n\to\infty}\sum_{\ell=1}^{N}(-1)^{\ell+1}\binom{n}{\ell}\left(\frac{\binom{n-\ell}{k}}{\binom{n}{k}}\right)^{t_\beta} = \lim_{n\to\infty}\sum_{\ell=1}^{N}(-1)^{\ell+1}\frac{\beta^\ell}{\ell!} = 1-e^{-\beta}.$$

For integers $a<b$ let Let $K_{[a,b]} = A_{a+1}\cup A_{a+2}\cup\ldots A_b$. Then we have shown

$$\liminf_{n\to\infty}\mathbb{P}(X_{t_\beta}\in K_{[1,N]}) \geq 1-e^{-\beta}.$$

Likewise, for any $j < \lfloor n/N \rfloor$,

$$\liminf_{n\to\infty}\mathbb{P}(X_{t_\beta}\in K_{[jN,(j+1)N]}) \geq 1-e^{-\beta}.$$

Hence

$$\liminf_{n\to\infty}\mathbb{P}(X_{t_\beta}\in \cap_{j=1}^m K_{[jN,(j+1)N]}) \geq 1-me^{-\beta}.$$

Let $\epsilon>0$. Then for any $\beta>0$, if $t=(1-\epsilon)t_{\mathrm{mix}}$ then $t<t_\beta$ for n sufficiently large, and hence

$$\liminf_{n\to\infty}\mathbb{P}(X_t\in \cap_{j=1}^m K_{[jN,(j+1)N]}) = 1.$$

But it is obvious that $\cap_{j=1}^m K_{[jN,(j+1)N]}\subset K_m$ and hence for $t=(1-\epsilon)t_{\mathrm{mix}}$,

$$\liminf_{n\to\infty}\mathbb{P}(X_t\in K_m) = 1. \tag{3.58}$$

Comparing with (3.53) the result follows. □

3.5.2 Proof of Lemma 3.3.2 for the case of k-cycles

Suppose here that Γ is the set of k-cycles where $k=o(n)$. Note that assumption (3.13) implies that either k is constant or strictly increasing to infinity. We will be adapting the proof of Theorem 2.3.2 in Durrett [24].

Suppose that $s = s(n)$ is such that $sk/n \to c$ for some $c > c_\Gamma$ as $n \to \infty$ for some $c \geq 0$. We reveal the vertices of the component containing a vertex $v \in \{1, \ldots, n\}$ as follows. There are three states that each vertex can be: unexplored, removed or active. Initially v is active and all the other vertices are unexplored. At each step of the iteration we select an active vertex w according to some prescribed rule (say with the smallest label). The vertex w becomes removed and every unexplored vertex which joined to w by a hyperedge becomes active. We repeat this exploration procedure until there are no more active vertices. At the i-th step of the exploration procedure we let A_i, R_i and U_i denote the set of active, removed and unexplored vertices respectively. Initially $A_0 = \{v\}$, $U_0 = \{1, \ldots, n\}\backslash\{v\}$ and $R_0 = \emptyset$.

Suppose that w is the vertex being explored on the i-the step of the exploration and let $\mathcal{M}_i(s) = \{t \leq s : w \in \mathrm{Supp}(\gamma_t)\}$. Consider the set

$$\mathcal{N}_i(s) := \mathcal{M}_i(s)\backslash \bigcup_{m=1}^{i} \mathcal{M}_m(s).$$

Then we have that $|A_i\backslash A_{i-1}| \leq (k-1)|\mathcal{N}_i(s)| - 1$. Indeed for each $t \in \mathcal{N}_i(s)$ we have that w is a point composing the k-cycle γ_t. There are $k-1$ many other points composing γ_t which may be new.

Conditionally on $\mathcal{F}_i := \sigma(|A_1|, \ldots, |A_i|)$ the probability that $t \in \mathcal{N}_i(s)$ for some $t \leq s$ is at most k/n. Hence it follows that conditionally on $\mathcal{F}_i$, $|\mathcal{N}_i(s)|$ is stochastically dominated by $D = (k-1)M(s)$ where $M(s)$ is a Binomial$(s, k/n)$. Let $S_0 = 1$ and for $i \geq 1$ define $S_i - S_{i-1} = D_i - 1$ where $D_1, D_2, \ldots$ is a sequence of i.i.d. random variables with distribution D. Hence it follows that we can couple $S = (S_i : i = 0, 1, \ldots)$ and $(|A_i| : i = 0, 1, \ldots)$ so that $|A_i| \leq S_i$ for $i \leq \tau_S := \inf\{i \geq 0 : S_i = 0\}$.

Let us now find a random variable which $|A_i\backslash A_{i-1}|$ stochastically dominates. Fix $\delta \in (0, 1)$ small, and condition on the event that for some $\ell \in \mathbb{N}$ we have that $|A_\ell| + \ell \leq \delta n$. This implies that $|U_i| \geq (1-\delta)n$ for each $i \leq \ell$ and that we discovered at most $\delta n/k$ many hyperedges by time i. Consequently we have that conditionally on $\mathcal{F}_i$, $|\mathcal{N}_i(s)|$ stochastically dominates $M(s - \delta n/k)$. Suppose that $t \in \mathcal{N}_i(s)$ and write $\gamma_t = (w, x_1, \ldots, x_{k-1})$. Then it might be the case that $\{x_1, \ldots, x_{k-1}\} \cap A_i \neq \emptyset$. However we have that $|A_i| \leq \delta n$ hence conditionally on $\mathcal{F}_i$, $|\{x_1, \ldots, x_{k-1}\} \cap A_i|$ is stochastically dominated by N with distribution Binomial$(k-1, \delta)$. It follows that conditionally on $\mathcal{F}_i$, $|A_i\backslash A_{i-1}|$ stochastically dominates $\tilde{D} = (k-1-N)M(s - \delta n/k)$.

Let $W_0 = 1$ and for $i \geq 1$ define $W_i - W_{i-1} = \tilde{D}_i - 1$ where $\tilde{D}_1, \tilde{D}_2, \ldots$ is a sequence of i.i.d. random variables with distribution $\tilde{D}$. Hence it follows that we can couple $W = (W_i : i = 0, 1, \ldots)$ and $(|A_i| : i = 0, 1, \ldots)$ so that $W_i \leq |A_i|$ for $i \leq \tau' := \inf\{i \geq$

$0 : |A_i| = 0$ or $|A_i| + i > \delta n/k\}$.

We now follow the four step proof of Theorem 2.3.2 in [24].

Step 1: Suppose that $W_0 \geq C \log(n/k)$ for some constant $C > 0$ to be determined later. Define $T_0 := \inf\{i \geq 0 : W_i = 0\}$ and for $x \in [0,1]$ let $\tilde{G}_n(x) = \mathbb{E}[x^{\tilde{D}}]$, then we have that

$$\begin{aligned}\lim_{n\to\infty} \tilde{G}_n(x) &= \lim_{n\to\infty} \left(1 - \frac{k}{n}\left(1 - [1-(1-\delta)(1-x)]^{k-1}\right)\right)^{s-\delta n/k} \\ &= \lim_{n\to\infty} \exp\left\{-(c-\delta)\left(1 - [1-(1-\delta)(1-x)]^{k-1}\right)\right\} \\ &= \Psi((1-\delta)(1-x), c-\delta).\end{aligned}$$

Suppose that $\delta > 0$ is small enough so that $c - \delta > c_\Gamma$. Then similar to the proof of Lemma 3.2.1 we have that there exists a $y \in (0,1)$ such that $\Psi((1-\delta)(1-y), c-\delta) < y$. Suppose that n is sufficiently large so that

$$\phi_n := \frac{y}{\tilde{G}_n(y)} > 1 + \epsilon$$

for some $\epsilon > 0$.

Now we have that $M_i := y^{W_i}\phi_n^i$ is a non-negative martingale, hence by the optional stopping theorem we have that for any $m \in \mathbb{N}$

$$y^{C\log(n/k)} \geq \mathbb{E}[M_0] = \mathbb{E}[M_{T_0 \wedge m}] \geq \mathbb{E}[\phi_n^{T_0}\mathbb{1}_{\{T_0 \leq m\}}] \geq \mathbb{P}(T_0 \leq m).$$

Taking $m \uparrow \infty$ we see that $\mathbb{P}(T_0 < \infty) \leq y^{C\log(n/k)}$. Taking $C > 3/\log(1/y)$ we conclude that

$$\mathbb{P}(T_0 < \infty | W_0 \geq C\log(n/k)) \leq \frac{k^3}{n^3}.$$

Step 2: Let $r = \beta \log(n/k)$ for some constant $\beta > 0$ to be determined later. Suppose that $|A_r| > 0$ then the event $\{|A_r| \leq r\}$ implies that $r < \tau'$. Hence it follows that

$$\begin{aligned}\mathbb{P}(0 < |A_r| \leq r) \leq \mathbb{P}(W_r \leq r) &= \mathbb{P}(y^{W_r} \geq y^r) \\ &\leq \left(\frac{\tilde{G}_n(y)}{y}\right)^r \\ &\leq (1+\epsilon)^{-\beta\log(n/k)}.\end{aligned}$$

Let $\beta > C \vee 3\log(1+\epsilon)$, then we have that $\mathbb{P}(0 < |A_r| \leq r) \leq k^3 n^{-3}$. It follows from Step 1 that if $|A_r| > 0$ then with high probability, the lower bounding random walk will never hit 0.

Step 3: Let $m = n^{2/3}k^{-2/3}$, then

$$\begin{aligned}\mathbb{P}(|A_m| + m > \delta n) &\leq \mathbb{P}\left(\sum_{i=1}^{m} D_i \geq \delta n\right) \\ &\leq \mathbb{E}[e^{D/(k-1)}]^m e^{-(\delta n)/(k-1)} \\ &= \left(1 + \frac{k}{n}\left(e^1 - 1\right)\right)^{sm} e^{-\delta n/(k-1)} \\ &\leq \kappa e^{-\delta n/k}\end{aligned}$$

for some constant $\kappa > 0$. Combined with Step 1, we have that if $|A_r| > 0$ then the coupling between W_i and $|A_i|$ will be valid for all $i \leq m = n^{2/3}k^{-2/3}$ with high probability and thus $\mathbb{P}(m > \tau') = O(e^{-\delta n/k})$. On the other hand suppose that $0 < K < \log[(1+\epsilon)/(1+\epsilon/2)]/\log[y]$, then

$$\mathbb{P}(|A_m| \leq Km | m < \tau') \leq \mathbb{P}(W_m \leq Km) = \mathbb{P}(y^{W_m} \geq y^{Km}) \leq \left(\frac{G_n(y)}{y^{K+1}}\right)^m \leq (1+\epsilon/2)^m.$$

Thus we see that at time $n^{2/3}k^{-2/3}$ there are at least $Kn^{2/3}k^{-2/3}$ many active vertices with high probability. Suppose now that $v, v' \in \{1, \ldots, n\}$ are vertices such that when we run two exploration processes, one started at v and one started v', we find that both processes at time $n^{2/3}k^{-2/3}$ have at least $Kn^{2/3}k^{-2/3}$ active vertices. There are two possibilities: either the exploration processes intersect by time $n^{2/3}k^{-2/3}$ or at time $n^{2/3}k^{-2/3}$ the set of active vertices for the two processes are disjoint. In the former situation v and v' are in the same component. Let us show that in the latter situation v and v' are in the same component with high probability. The probability a uniformly chosen k-hyperedge connects the set of active vertices of v to the set of active vertices of v' is at least

$$\frac{2(Kn^{2/3}k^{-2/3})^2}{k(k-1)} \frac{\binom{n-2}{k-2}}{\binom{n}{k}} \leq \kappa' \frac{k^{2/3}}{n^{2/3}}$$

for some constant $\kappa' > 0$. On the other hand we know that with high probability there are at least $(1-\delta)n$ unexplored vertices and hence there are at least $s - \delta n/k = (c-\delta)n/k$ many hyperedges that are unexplored. Hence we see that the probability that there is no hyperedge connecting the active vertices of v and v' is at most

$$(1 - \kappa' k^{2/3} n^{-2/3})^{(c-\delta)n/k} \leq e^{-\kappa'(c-\delta)n^{1/3}k^{-1/3}}.$$

For $v \in \{1, \ldots, n\}$ let $\mathcal{C}_v$ denote the size of the component containing v and for

$v, v' \in \{1, \ldots, n\}$ define

$A(v, v') = \{v \text{ and } v' \text{ are in the same component}\} \cup \{\mathcal{C}_v \leq C \log(n/k)\} \cup \{\mathcal{C}_v \leq C \log(n/k)\}$.

Altogether we have just shown that

$$\mathbb{P}(A(v, v')^c) \leq 2k^3 n^{-3} + \kappa e^{-\delta n/k} + e^{-\kappa'(c-\delta)n^{1/3}k^{-1/3}}.$$

Suppose now that v, v' are such that $A(v, v')^c$ holds. Then it follows that there are two disjoint hyperedges e_v and $e_{v'}$ such that $v \in e_v$ and $v' \in e_{v'}$. Now we have that $A(w, w')^c$ holds for any $w \in e_v$ and $w' \in e_{v'}$. Hence

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{(v,v') \in \{1,\ldots,n\}^2} A(v, v') \right) &= \mathbb{P}(\#\{(v, v') \in \{1, \ldots, n\}^2 : A(v, v')^c \text{ holds}\} \geq k^2) \\
&\leq k^{-2} \mathbb{E}[\#\{(v, v') \in \{1, \ldots, n\}^2 : A(v, v')^c \text{ holds}\}] \\
&= \frac{n^2}{k^2} \mathbb{P}(A(v, v')) \\
&= o(1).
\end{aligned}$$

Hence we see that with high probability the set of vertices which are in components of size greater than $C \log(n/k)$ are connected.

Step 4: Now it suffices to show that $Y^{(n)} := |\{v : \mathcal{C}_v \leq C \log(n/k)\}|/n \to 1 - \theta(c)$ in probability as $n \to \infty$. Suppose that $W_0 = S_0 = 1$ and let $T_0 := \inf\{i \geq 0 : W_i = 0\}$ and $T_0' := \inf\{i \geq 0 : S_i = 0\}$. Similar to before one can show that $G_n(x) = \mathbb{E}[x^D]$ converges point-wise in $(0, 1)$ as $n \to \infty$ to $\Psi(1 - x, c)$. Let $\rho_n(c)$ be the minimal solution in $[0, 1]$ to the equation $\rho_n(c) = G_n(\rho_n(c))$. Then it is not hard to show that $\rho_n(c) \to 1 - \theta(c)$ as $n \to \infty$. Now suppose that n is large so that $\rho_n(c) < 1$. Then we have that $M_i := \rho_n(c)^{S_i}$ is a martingale hence by the optional stopping theorem we have that

$$\mathbb{P}(T_0' \leq C \log(n/k)) + \mathbb{E}[M_{C \log(n/k)} \mathbb{1}_{\{T_0' > C \log(n/k)\}}] = \mathbb{E}[M_{T_0' \wedge C \log(n/k)}] = \rho_n(c)$$

and hence for any $v \in \{1, \ldots, n\}$,

$$\liminf_{n \to \infty} \mathbb{P}(\mathcal{C}_v \leq C \log(n/k)) \geq \lim_{n \to \infty} \mathbb{P}(T_0' \leq C \log(n/k)) = \lim_{n \to \infty} \rho_n(c) = 1 - \theta(c).$$

Using a similar argument with T_0 yields that $\limsup_{n \to \infty} \mathbb{P}(\mathcal{C}_v \leq C \log(n/k)) \leq 1 - \theta(c -$

99

δ). As $\delta > 0$ is arbitrary it follows that

$$\lim_{n\to\infty} \mathbb{E}[Y^{(n)}] = \lim_{n\to\infty} \mathbb{P}(\mathcal{C}_v \le C\log(n/k)) = 1-\theta(c).$$

For $v \in \{1,\dots,n\}$ let $Y_v = \mathbb{1}_{\{\mathcal{C}_v \le C\log(n/k)\}}$ and suppose that $Y_1 = 1$. Then there is a set of vertices $V \subset \{1,\dots,n\}$ which make up the component containing 1 and $|V| \le C\log(n/k)$. The probability that a uniformly chosen k-hyperedge is disjoint from the set V is at least

$$\left(1-\frac{C\log(n/k)}{n}\right)\cdots\left(1-\frac{C\log(n/k)}{n-k+1}\right) \ge \left(1-\frac{C\log(n/k)}{n-k}\right)^k.$$

Thus the probability that the exploration process started from 2 does not reveal an element of V is at least

$$\left(1-\frac{C\log(n/k)}{n-k}\right)^{kC\log(n/k)} \ge \kappa'' e^{-\frac{kC^2\log(n/k)^2}{n}}$$

for some constant $\kappa'' > 0$. Hence it follows that

$$\limsup_{n\to\infty} \mathrm{Var}(Y^{(n)}) \le \limsup_{n\to\infty} n^{-2}\left(n+\binom{n}{2}(1-\kappa'' e^{-\frac{kC^2\log(n/k)^2}{n}})\right) = 0.$$

Thus it follows from Chebychev's inequality that for any $\eta > 0$

$$\limsup_{n\to\infty} \mathbb{P}(|Y^{(n)}-(1-\theta(c))| > \eta) = \limsup_{n\to\infty} \mathbb{P}(|Y^{(n)}-\mathbb{E}[Y^{(n)}]| > \eta) \le \limsup_{n\to\infty} \frac{\mathrm{Var}(Y^{(n)})}{\eta^2} = 0.$$

3.5.3 Proof of Theorem 3.3.6

Let $\Gamma \subset \mathcal{S}_n$ be a conjugacy class with cycle structure $(k_2, k_3, \dots)$. Let $X = (X_t : t = 0, 1, \dots)$ be a random walk on $\mathcal{S}_n$ which at each step applies an independent uniformly random element of Γ. Let $\rho = \sum_j (j-1)k_j$ and let $\tilde{X}$ be the transposition walk associated to the walk X using (3.47). In particular for $t \ge 0$, $\tilde{X}_{t\rho} = X_t$. Finally let $Z = (Z_1, Z_2, \dots)$ denote a Poisson–Dirichlet random variable.

For convenience we restate Theorem 3.3.6 here.

Theorem 3.5.2. *Let $s \ge 0$ be such that $sk/(n\rho) \to c$ for some $c > c_\Gamma$. Then for each $m \in \mathbb{N}$ we have that as $n \to \infty$,*

$$\left(\frac{\bar{\mathfrak{X}}_1(\tilde{X}_s)}{\theta(c)}, \dots, \frac{\bar{\mathfrak{X}}_m(\tilde{X}_s)}{\theta(c)}\right) \to (Z_1, \dots, Z_m)$$

in distribution where $\theta(c)$ is given by (3.20).

The proof of this result is very similar to the proof of Theorem 1.1 in Schramm [47]. We give the details here.

Recall the hypergraph process $H = (H_t : t = 0, 1, \dots)$ associated with the walk X defined in Section 3.3.1. Analogously let $\tilde{G} = (\tilde{G}_t : t = 0, 1, \dots)$ be a process of graphs on $\{1, \dots, n\}$ such that the edge $\{x, y\}$ is present in $\tilde{G}_t$ if and only if the transposition (x, y) has been applied to $\tilde{X}$ prior to and including time t. Hence we have that for each $t = 0, 1, \dots$, $\tilde{G}_{t\rho} = H_t$.

Recall that $\tilde{X}$ satisfies conditional uniformity as described in Proposition 3.4.2. Using the graph process $\tilde{G}$ above and the conditional uniformity of $\tilde{X}$ the following lemma, which is the analogue of Lemma 2.4 in Schramm [47], follows almost verbatim from Schramm's arguments.

Lemma 3.5.3. *Let $s \geq 0$ be such that $sk/(n\rho) \to c$ for some $c > c_\Gamma$ and let $\epsilon \in (0, 1/8)$. Let $M = M(\epsilon, n, s)$ be the minimum number of cycles of $\tilde{X}_s$ which are needed to cover at least $(1-\epsilon)$ proportion of the vertices in the giant component of $\tilde{G}_s$. Then for $\alpha \in (0, 1/8)$ we have that*

$$\limsup_{n\to\infty} \mathbb{P}(M > \alpha^{-1}|\log(\alpha\epsilon)|^2) \leq C\alpha$$

for some constant C which does not depend on α nor ϵ.

Henceforth fix some time $s \geq 0$ such that $sk/(n\rho) \to c$ for some $c > c_\Gamma$. Fix $\epsilon \in (0, 1/8)$ and define

$$\Delta := \lfloor \epsilon^{-1} \rfloor$$
$$s_0 := s - \Delta.$$

For $t = 0, \dots, \Delta$ define $\bar{X}_t = \mathfrak{X}(\tilde{X}_{s_0+t})$. We can assume that for $t \leq \Delta$, $\tilde{X}_{s_0+t}$ satisfies the relaxed conditional uniformity assumption described in Definition 3.4.4. Indeed by making this assumption we are disregarding the constraint on the transpositions described in Proposition 3.4.2 applied to $\tilde{X}_t$ for $t = s_0, \dots, s$. However the probability that we violate this constraint is at most $2\Delta k/n$.

Colour an element of $\bar{X}_0 = \mathfrak{X}(\tilde{X}_{s_0})$ green if the cycle whose renormalised cycle length is this element lies in the giant component of $\tilde{G}_{s_0}$. We colour all the other elements of $\bar{X}_0$ red. Thus asymptotically in n, the sum of the green elements is $\theta(c)$ and the sum of the red elements is $1 - \theta(c)$. In the evolution of $(\bar{X}_t : t = 0, 1, \dots)$ we keep the colour scheme as follows. If an element fragments, both fragments retain the same colour. If we coagulate two elements of the same colour then the new element retains the colour of the previous two elements. If we coagulate a green element and a red element, then the colour of the resulting element is green.

Define $\bar{X}' = (\bar{X}'_t : t = 0, \ldots, \Delta)$ and $\bar{X}'' = (\bar{X}''_t : t = 0, \ldots, \Delta)$ as follows. Initially $\bar{X}'_0 = \bar{X}''_0 = \bar{X}_0$. Apply the same colouring scheme to $\bar{X}'$ and $\bar{X}''$ as we did to $\bar{X}$. Each step evolution is described as follows. Then the walks evolve as follows.

- $\bar{X}'_t$: Evolves the same as $\bar{X}$ except we ignore any transition which involves a red entry.

- $\bar{X}''_t$: Evolves the same as $\bar{X}'$ except that the markers u, v used in the transitions of $\bar{X}''$ are distributed uniformly on $[0, 1]$.

Lemma 3.3.2 states that the second largest component of $\tilde{G}_{s_0}$ has size $o(n)$. Hence, initially each red element has size $o(1)$ as $n \to \infty$. Now Δ does not increase with n, hence for any $s = 0, 1, \ldots, \Delta$, we are unlikely to make a coagulation (or fragmentation) in $\bar{X}'_s$ without coagulating (or fragmenting) entries of $\bar{X}_s$ of similar size. Similar considerations for the processes $\bar{X}'$ and $\bar{X}''$ leads to the following lemma.

Lemma 3.5.4. *There exists a coupling between the walks $\bar{X}$ and $\bar{X}'$, and between $\bar{X}'$ and $\bar{X}''$ such that for each $\eta > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{i\in\mathbb{N}} |\bar{X}_i(\Delta) - \bar{X}'_i(\Delta)| > \eta\right) = \lim_{n\to\infty} \mathbb{P}\left(\sup_{i\in N} |\bar{X}'_i(\Delta) - \bar{X}''_i(\Delta)| > \eta\right) = 0.$$

Using the preceding lemma, it suffices now to find an appropriate coupling between $\bar{X}''$ and Z. To do this we modify Schramm's coupling in Schramm [47]. First we let $\{J_1, \ldots, J_L\}$ be the set of times $s \in \{0, \ldots, \Delta\}$ such that that $\bar{X}''_{s-1} \neq \bar{X}''_s$. It is easy to see that $\lim_{n\to\infty} \mathbb{P}(L > \sqrt{\Delta}) = 1$ and henceforth we will condition on the event that $\{L > \sqrt{\Delta}\}$ and set $\Delta' = \lfloor\sqrt{\Delta}\rfloor$. Define a process $\bar{Y} = (\bar{Y}_t : t = 0, \ldots, \Delta')$ as follows. Initially $\bar{Y}_0 = \bar{X}''_0$. For $t = 1, \ldots, \Delta'$ we let $\bar{Y}_t$ be X''_{J_t} renormalised so that $\sum_i \bar{Y}_i(t) = 1$ where $\bar{Y}_i(t)$ is the i-th element of $\bar{Y}_t$.

We define a process $\bar{Z} = (Z_t : t = 0, 1, \ldots, \sqrt{\Delta})$ as follows. Initially Z_0 has the distribution of a Poisson–Dirichlet random variable, independent of $\bar{Y}$. Then for $t = 1, \ldots, \Delta'$ define $\bar{Z}_t$ by applying the coupling in Section 3.4.2 to $\bar{Y}$ and $\bar{Z}$ but with the following modifications:

- the markers $u, v \in [0, 1]$ are taken uniformly at random,

- we always take $v' = v$,

- we modify the definition of a refreshment time: s is a refreshment time if either $J_{s-1} + 2 \leq J_s$ or $J_s + s_0$ is a refreshment time in the sense of Definition 3.4.1,

- when a marked tile of size a fragments, it creates a tile of length v and a tile of length $a - v$. We mark the tile of length $a - v$.

It is not hard to check that the Poisson–Dirichlet distribution is invariant under this evolution and hence we have that for each $t = 0, 1, \ldots, \Delta'$, Z_t has the law of a Poisson–Dirichlet.

Our coupling agrees with the coupling in Schramm [47, Section 3] when $\Gamma = T$ is the set of all transpositions. In this case each time s is a refreshment time and hence the marked tile at time s is always chosen by the marker u. One can adapt the arguments in Chapter 3 of Schramm's paper to our case by using the following idea. Note first that all the estimates of Schramm apply at s when s is a refreshment time. When s is not a refreshment time and Schramm considers the event that the marker u at time s falls inside an unmatched tile, instead we consider the event that the marker v at time $s-1$ falls inside an unmatched tile. By the properties of the coupling, this guarantees that at time s the marked tile is unmatched.

Adapting Schramm's arguments leads to the following lemma, which is the analogue of Schramm [47, Corollary 3.4].

Lemma 3.5.5. *Define*

$$N^0 := \#\{i \in \mathbb{N} : \bar{Y}_i(0) > \epsilon\} + \#\{i \in \mathbb{N} : \bar{Z}_i(0) > \epsilon\}$$

and let

$$\bar{\epsilon} := \epsilon + \sum_{i=1}^{\infty} \bar{Y}_i(0) \mathbb{1}_{\{\bar{Y}_i(0) < \epsilon\}} + \sum_{i=1}^{\infty} \bar{Z}_i(0) \mathbb{1}_{\{\bar{Z}_i(0) < \epsilon\}}.$$

Define the event

$$\mathcal{B} = \left\{ \bar{\epsilon}^{4/5} \leq \frac{1}{\Delta'} \leq \frac{\bar{\epsilon}^{1/5}}{N^0 \vee 1} \right\}.$$

Let $q \in \{1, \ldots, \Delta'\}$ be distributed uniformly, independent of the processes $\bar{Y}$ and $\bar{Z}$. Then we have that for each $\rho > 0$,

$$\mathbb{P}(\sup_{i \in \mathbb{N}} |\bar{Y}_i(q) - \bar{Z}_i(q)| > \rho) \leq C \frac{\mathbb{P}(\mathcal{B})}{\rho \log \Delta'}$$

for some constant $C > 0$.

Using Lemma 3.5.5 it suffices to show that $\mathbb{P}(\mathcal{B}) / \log \Delta' \to 0$ as $\epsilon \downarrow 0$. The following lemma shows a stronger result.

Lemma 3.5.6. *Suppose that $\mathcal{B}$ is defined as in Lemma 3.5.5, then*

$$\lim_{\epsilon \downarrow 0} \mathbb{P}(\mathcal{B}) = 1.$$

Proof. Let

$$\mathcal{B}_1 := \left\{\bar{\epsilon}^{4/5} \le \frac{1}{2}\epsilon^{1/2}\right\}$$
$$\mathcal{B}_2 := \left\{2\epsilon^{1/2} \le \frac{\bar{\epsilon}^{1/5}}{N^0 \vee 1}\right\}$$

Now as $(1/2)\epsilon^{-1/2} \le \Delta' \le 2\epsilon^{-1/2}$ we have that $\mathcal{B} \supset \mathcal{B}_1 \cap \mathcal{B}_2$. First let us bound $\mathbb{P}(\mathcal{B}_1^c)$. Note that on the event $\mathcal{B}_1^c$ we have that $\bar{\epsilon} > 2^{-5/4}\epsilon^{5/8}$. Note that a size biased sample from a Poisson–Dirichlet random variable has a uniform law on $[0,1]$. Hence it follows that

$$\mathbb{E}\left[\sum_{i=1}^{\infty} \bar{Z}_i(0)\mathbb{1}_{\{\bar{Z}_i(0)<\epsilon\}}\right] = \epsilon$$

and thus

$$\mathbb{P}\left(\sum_{i=1}^{\infty} \bar{Z}_i(0)\mathbb{1}_{\{\bar{Z}_i(0)<\epsilon\}} > \epsilon^{5/6}\right) \le \frac{\mathbb{E}\left[\sum_{i=1}^{\infty} \bar{Z}_i(0)\mathbb{1}_{\{\bar{Z}_i(0)<\epsilon\}}\right]}{\epsilon^{5/6}} \le \epsilon^{1/6}.$$

Next consider the random variable M in Lemma 3.5.3 at time $s_0 = s - \Delta$ where we recall that $\bar{Y}(0) = \mathfrak{X}(\tilde{X}_{s_0})$. We have that

$$\sum_{i=1}^{\infty} Y_i(0)\mathbb{1}_{\{\bar{Y}_i(0):\bar{Y}_i(0)<\epsilon\}} \le \epsilon(M+1)$$

Then applying Lemma 3.5.3 at time s_0 we have that

$$\mathbb{P}\left(\sum_{i=1}^{\infty} \bar{Y}_i(0)\mathbb{1}_{\{\bar{Y}_i(0)<\epsilon\}} > \epsilon^{5/6}\right) \le \mathbb{P}(M > \epsilon^{-1/5}) \le C\epsilon^{1/6}$$

for some constant $C > 0$. Hence it follows that for $\epsilon > 0$ small

$$\mathbb{P}(\mathcal{B}_1^c) = \mathbb{P}(\bar{\epsilon} > 2^{-5/4}\epsilon^{5/8}) \le \mathbb{P}(\bar{\epsilon} > \epsilon^{5/6}) \le \epsilon^{1/6} + C\epsilon^{1/6}.$$

which shows that $\mathbb{P}(\mathcal{B}_1) \to 1$ as $\epsilon \downarrow 0$.

Now we bound $\mathbb{P}(\mathcal{B}_2^c)$. Firstly we use the bound $\bar{\epsilon} \ge \epsilon$ and so we are left to bound N^0 from above. Using the stick breaking construction of Poisson–Dirichlet random variables (see for example [10, Definition 1.4]) one can show that

$$\mathbb{P}\left(\#\{i \in \mathbb{N} : Z_i(0) > \epsilon\} > \epsilon^{-1/4}\right) \le C'\epsilon$$

 104

for some constant $C' > 0$. On the other hand we have that

$$\#\{i \in \mathbb{N} : Y_i(0) > \epsilon\} \leq M$$

and hence using Lemma 3.5.3 we obtain

$$\mathbb{P}\left(\#\{i \in \mathbb{N} : Y_i(0) > \epsilon\} > \epsilon^{-1/4}\right) \leq C''\epsilon^{1/5}$$

for some constant $C'' > 0$. Hence it follows that $\mathbb{P}(\mathcal{B}_2^c) \leq C'\epsilon + C''\epsilon^{1/5}$ and the result now follows.

Theorem 3.5.2 now follows from Lemma 3.5.4, Lemma 3.5.5 and Lemma 3.5.6.

Bibliography

[1] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit. "A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces". In: *Electron. J. Probab.* 18 (2013), no. 14, 21. ISSN: 1083-6489. DOI: 10.1214/EJP.v18-2116. URL: http://dx.doi.org/10.1214/EJP.v18-2116.

[2] David Aldous. "Random walks on finite groups and rapidly mixing Markov chains". In: *Seminar on probability, XVII.* Vol. 986. Lecture Notes in Math. Springer, Berlin, 1983, pp. 243–297. DOI: 10.1007/BFb0068322. URL: http://dx.doi.org/10.1007/BFb0068322.

[3] David Aldous. "Brownian excursions, critical random graphs and the multiplicative coalescent". In: *Ann. Probab.* 25.2 (1997), pp. 812–854. ISSN: 0091-1798. DOI: 10.1214/aop/1024404421. URL: http://dx.doi.org/10.1214/aop/1024404421.

[4] David Aldous. "Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists". In: *Bernoulli* 5.1 (1999), pp. 3–48. ISSN: 1350-7265. DOI: 10.2307/3318611. URL: http://dx.doi.org/10.2307/3318611.

[5] Krishna B. Athreya and Peter E. Ney. *Branching processes.* Die Grundlehren der mathematischen Wissenschaften, Band 196. Springer-Verlag, New York-Heidelberg, 1972, pp. xi+287.

[6] Itai Benjamini and Oded Schramm. "Recurrence of distributional limits of finite planar graphs". In: *Electron. J. Probab.* 6 (2001), no. 23, 13 pp. (electronic). ISSN: 1083-6489. DOI: 10.1214/EJP.v6-96. URL: http://dx.doi.org/10.1214/EJP.v6-96.

[7] Julien Berestycki and Nathanaël Berestycki. "Kingman's coalescent and Brownian motion". In: *ALEA Lat. Am. J. Probab. Math. Stat.* 6 (2009), pp. 239–259. ISSN: 1980-0436.

[8] Julien Berestycki, Nathanaël Berestycki, and Vlada Limic. "The Λ-coalescent speed of coming down from infinity". In: *Ann. Probab.* 38.1 (2010), pp. 207–233. ISSN: 0091-1798. DOI: 10.1214/09-AOP475. URL: http://dx.doi.org/10.1214/09-AOP475.

[9] Julien Berestycki, Nathanaël Berestycki, and Jason Schweinsberg. “Small-time behavior of beta coalescents”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 44.2 (2008), pp. 214–238. ISSN: 0246-0203. DOI: 10.1214/07-AIHP103. URL: http://dx.doi.org/10.1214/07-AIHP103.

[10] Nathanaël Berestycki. *Recent progress in coalescent theory*. Vol. 16. Ensaios Matemáticos [Mathematical Surveys]. Rio de Janeiro: Sociedade Brasileira de Matemática, 2009, p. 193. ISBN: 978-85-85818-40-1.

[11] Nathanaël Berestycki. “Emergence of giant cycles and slowdown transition in random transpositions and k-cycles”. In: *Electron. J. Probab.* 16 (2011), no. 5, 152–173. ISSN: 1083-6489. DOI: 10.1214/EJP.v16-850. URL: http://dx.doi.org/10.1214/EJP.v16-850.

[12] Nathanaël Berestycki, Oded Schramm, and Ofer Zeitouni. “Mixing times for random k-cycles and coalescence-fragmentation chains”. In: *Ann. Probab.* 39.5 (2011), pp. 1815–1843. ISSN: 0091-1798. DOI: 10.1214/10-AOP634. URL: http://dx.doi.org/10.1214/10-AOP634.

[13] Nathanaël Berestycki and Batı Şengül. “Cutoff for conjugacy-invariant random walks on the permutation group”. In: *arXiv:1410.4800* (2014).

[14] Jean Bertoin. *Random fragmentation and coagulation processes*. Vol. 102. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2006, pp. viii+280. ISBN: 978-0-521-86728-3; 0-521-86728-2. DOI: 10.1017/CBO9780511617768. URL: http://dx.doi.org/10.1017/CBO9780511617768.

[15] Jean Bertoin and Jean-Francois Le Gall. “Stochastic flows associated to coalescent processes. III. Limit theorems”. In: *Illinois J. Math.* 50.1-4 (2006), 147–181 (electronic). ISSN: 0019-2082. URL: http://projecteuclid.org/getRecord?id=euclid.ijm/1258059473.

[16] Russ Bubley and Martin Dyer. “Path coupling: A technique for proving rapid mixing in Markov chains”. In: *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*. IEEE. 1997, pp. 223–231.

[17] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*. Vol. 33. Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2001, pp. xiv+415. ISBN: 0-8218-2129-6.

[18] Brigitte Chauvin and Alain Rouault. “KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees”. In: *Probab. Theory Related Fields* 80.2 (1988), pp. 299–314. ISSN: 0178-8051. DOI: 10.1007/BF00356108. URL: http://dx.doi.org/10.1007/BF00356108.

[19] Nicolas Curien and Jean-François Le Gall. "The Brownian Plane". English. In: *Journal of Theoretical Probability* (2013), pp. 1–43. ISSN: 0894-9840. DOI: 10.1007/s10959-013-0485-0. URL: http://dx.doi.org/10.1007/s10959-013-0485-0.

[20] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988, pp. vi+198. ISBN: 0-940600-14-5.

[21] Persi Diaconis and Mehrdad Shahshahani. "Generating a random permutation with random transpositions". In: *Z. Wahrsch. Verw. Gebiete* 57.2 (1981), pp. 159–179. ISSN: 0044-3719. DOI: 10.1007/BF00535487. URL: http://dx.doi.org/10.1007/BF00535487.

[22] Peter Donnelly and Thomas G. Kurtz. "Particle representations for measure-valued population models". In: *Ann. Probab.* 27.1 (1999), pp. 166–205. ISSN: 0091-1798. DOI: 10.1214/aop/1022677258. URL: http://dx.doi.org/10.1214/aop/1022677258.

[23] Rick Durrett. *Probability: theory and examples*. Fourth. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010, pp. x+428. ISBN: 978-0-521-76539-8. DOI: 10.1017/CBO9780511779398. URL: http://dx.doi.org/10.1017/CBO9780511779398.

[24] Rick Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010, pp. x+210. ISBN: 978-0-521-15016-3.

[25] Paul Erdős and Alfred Rényi. "On the evolution of random graphs". In: *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5 (1960), pp. 17–61.

[26] Steven N. Evans. "Kingman's coalescent as a random metric space". In: *Stochastic models (Ottawa, ON, 1998)*. Vol. 26. CMS Conf. Proc. Providence, RI: Amer. Math. Soc., 2000, pp. 105–114.

[27] Steven N. Evans. *Probability and real trees*. Vol. 1920. Lecture Notes in Mathematics. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Berlin: Springer, 2008, pp. xii+193. ISBN: 978-3-540-74797-0. DOI: 10.1007/978-3-540-74798-7. URL: http://dx.doi.org/10.1007/978-3-540-74798-7.

[28] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. New York: John Wiley & Sons Inc., 1971, pp. xxiv+669.

 108

[29] Leopold Flatto, Andrew M. Odlyzko, and David B. Wales. "Random shuffles and group representations". In: *Ann. Probab.* 13.1 (1985), pp. 154–178. ISSN: 0091-1798. URL: http://links.jstor.org/sici?sici=0091-1798(198502)13:1<154:RSAGR>2.0.CO;2-0&origin=MSN.

[30] Bruce Hughes. "Trees and ultrametric spaces: a categorical equivalence". In: *Adv. Math.* 189.1 (2004), pp. 148–191. ISSN: 0001-8708. DOI: 10.1016/j.aim.2003.11.008. URL: http://dx.doi.org/10.1016/j.aim.2003.11.008.

[31] Mark Jerrum. "A very simple algorithm for estimating the number of k-colorings of a low-degree graph". In: *Random Structures Algorithms* 7.2 (1995), pp. 157–165. ISSN: 1042-9832. DOI: 10.1002/rsa.3240070205. URL: http://dx.doi.org/10.1002/rsa.3240070205.

[32] Michał Karoński and Tomasz Łuczak. "The phase transition in a random hypergraph". In: *J. Comput. Appl. Math.* 142.1 (2002). Probabilistic methods in combinatorics and combinatorial optimization, pp. 125–135. ISSN: 0377-0427. DOI: 10.1016/S0377-0427(01)00464-2. URL: http://dx.doi.org/10.1016/S0377-0427(01)00464-2.

[33] J. F. C. Kingman. "The coalescent". In: *Stochastic Process. Appl.* 13.3 (1982), pp. 235–248. ISSN: 0304-4149. DOI: 10.1016/0304-4149(82)90011-4. URL: http://dx.doi.org/10.1016/0304-4149(82)90011-4.

[34] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, 2009, pp. xviii+371. ISBN: 978-0-8218-4739-8.

[35] John Lott and Cédric Villani. "Ricci curvature for metric-measure spaces via optimal transport". In: *Ann. of Math. (2)* 169.3 (2009), pp. 903–991. ISSN: 0003-486X. DOI: 10.4007/annals.2009.169.903. URL: http://dx.doi.org/10.4007/annals.2009.169.903.

[36] Nathan Lulov and Igor Pak. "Rapidly mixing random walks and bounds on characters of the symmetric group". In: *J. Algebraic Combin.* 16.2 (2002), pp. 151–163. ISSN: 0925-9899. DOI: 10.1023/A:1021172928478. URL: http://dx.doi.org/10.1023/A:1021172928478.

[37] Yann Ollivier. "Ricci curvature of Markov chains on metric spaces". In: *J. Funct. Anal.* 256.3 (2009), pp. 810–864. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2008.11.001. URL: http://dx.doi.org/10.1016/j.jfa.2008.11.001.

[38] Yann Ollivier and Cédric Villani. "A curved Brunn-Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube?" In: *SIAM J. Discrete Math.* 26.3 (2012), pp. 983–996. ISSN: 0895-4801. DOI: 10.1137/11085966X. URL: http://dx.doi.org/10.1137/11085966X.

[39] Jim Pitman. "Coalescents with multiple collisions". In: *Ann. Probab.* 27.4 (1999), pp. 1870–1902. ISSN: 0091-1798. DOI: 10.1214/aop/1022677552. URL: http://dx.doi.org/10.1214/aop/1022677552.

[40] Max-K. von Renesse and Karl-Theodor Sturm. "Transport inequalities, gradient estimates, entropy, and Ricci curvature". In: *Comm. Pure Appl. Math.* 58.7 (2005), pp. 923–940. ISSN: 0010-3640. DOI: 10.1002/cpa.20060. URL: http://dx.doi.org/10.1002/cpa.20060.

[41] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion.* Third. Vol. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Berlin: Springer-Verlag, 1999, pp. xiv+602. ISBN: 3-540-64325-7.

[42] Yuval Roichman. "Upper bound on the characters of the symmetric groups". In: *Invent. Math.* 125.3 (1996), pp. 451–485. ISSN: 0020-9910. DOI: 10.1007/s002220050083. URL: http://dx.doi.org/10.1007/s002220050083.

[43] Yuval Roichman. "Characters of the symmetric groups: formulas, estimates and applications". In: *Emerging applications of number theory (Minneapolis, MN, 1996).* Vol. 109. IMA Vol. Math. Appl. Springer, New York, 1999, pp. 525–545. DOI: 10.1007/978-1-4612-1544-8_21. URL: http://dx.doi.org/10.1007/978-1-4612-1544-8_21.

[44] Sandrine Roussel. "Marches aléatoires sur e groupe symétrique." PhD thesis. Toulouse, 1999.

[45] Sandrine Roussel. "Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique". In: *Colloq. Math.* 86.1 (2000), pp. 111–135. ISSN: 0010-1354.

[46] Serik Sagitov. "The general coalescent with asynchronous mergers of ancestral lines". In: *J. Appl. Probab.* 36.4 (1999), pp. 1116–1125. ISSN: 0021-9002.

[47] Oded Schramm. "Compositions of random transpositions". In: *Israel J. Math.* 147 (2005), pp. 221–243. ISSN: 0021-2172. DOI: 10.1007/BF02785366. URL: http://dx.doi.org/10.1007/BF02785366.

[48] Jason Schweinsberg. “A necessary and sufficient condition for the Λ-coalescent to come down from infinity”. In: *Electron. Comm. Probab.* 5 (2000), 1–11 (electronic). ISSN: 1083-589X. DOI: 10.1214/ECP.v5-1013. URL: http://dx.doi.org/10.1214/ECP.v5-1013.

[49] Jason Schweinsberg. “Coalescents with simultaneous multiple collisions”. In: *Electron. J. Probab.* 5 (2000), Paper no. 12, 50 pp. (electronic). ISSN: 1083-6489. DOI: 10.1214/EJP.v5-68. URL: http://dx.doi.org/10.1214/EJP.v5-68.

[50] Batı Şengül. “Scaling Limits of Coalescent Processes Near Time Zero”. In: *arXiv:1309.0494* (2013).

[51] Leon Simon. *Lectures on geometric measure theory.* Vol. 3. Proceedings of the Centre for Mathematical Analysis, Australian National University. Canberra: Australian National University Centre for Mathematical Analysis, 1983, pp. vii+272. ISBN: 0-86784-429-9.

[52] Karl-Theodor Sturm. “On the geometry of metric measure spaces. I”. In: *Acta Math.* 196.1 (2006), pp. 65–131. ISSN: 0001-5962. DOI: 10.1007/s11511-006-0002-8. URL: http://dx.doi.org/10.1007/s11511-006-0002-8.

[53] Karl-Theodor Sturm. “On the geometry of metric measure spaces. II”. In: *Acta Math.* 196.1 (2006), pp. 133–177. ISSN: 0001-5962. DOI: 10.1007/s11511-006-0003-7. URL: http://dx.doi.org/10.1007/s11511-006-0003-7.

[54] A. M. Vershik and S. V. Kerov. “Asymptotic theory of the characters of a symmetric group”. In: *Funktsional. Anal. i Prilozhen.* 15.4 (1981), pp. 15–27, 96. ISSN: 0374-1990.

[55] Guofang Wei. “Manifolds with a lower Ricci curvature bound”. In: *Surveys in differential geometry. Vol. XI.* Vol. 11. Surv. Differ. Geom. Int. Press, Somerville, MA, 2007, pp. 203–227.