

Phase transitions for the distance of random  
walks with applications to genome  
rearrangements

Nathanaël Berestycki

May 2005



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.1.1	From genome evolution to random walks . . . . .	1
1.1.2	Phase transitions of random walks . . . . .	5
1.1.3	Informal overview of the results in my thesis . . . . .	7
1.1.4	Relation to classical theory of card shuffling and the cut-off phenomenon . . . . .	7
1.2	Erdős-Renyi Random Graphs . . . . .	9
1.3	Random transpositions: the phase transition. . . . .	14
1.3.1	Coupling with an Erdős-Renyi random graph. . . . .	14
1.3.2	Law of large numbers behavior for random transpositions	17
1.3.3	Fluctuations . . . . .	19
1.4	The (hyperbolic) geometric picture . . . . .	22
1.4.1	Gromov hyperbolicity . . . . .	23
1.4.2	Results . . . . .	24
1.5	Limiting behavior for the distance of random walks . . . . .	26
1.5.1	Random $p$ -cycles . . . . .	26
1.5.2	Random adjacent transpositions . . . . .	28
1.5.3	Random walk on a random 3-regular graph . . . . .	30
1.5.4	Other random walks and some open problems . . . . .	32
1.6	On a result of Oded Schramm . . . . .	33
1.6.1	The Poisson-Dirichlet distribution . . . . .	33
1.6.2	A process of coalescence and fragmentation . . . . .	34
1.6.3	Schramm's result . . . . .	35
1.6.4	More open problems . . . . .	37

<b>2</b>	<b>A phase transition in the random transposition random walk</b>	<b>43</b>
2.1	General motivation . . . . .	43
2.2	The coagulation-fragmentation process and the random graph process . . . . .	47
2.3	Limit Theorems . . . . .	49
2.3.1	The subcritical regime . . . . .	49
2.3.2	The critical regime . . . . .	51
2.3.3	The supercritical regime . . . . .	52
2.3.4	Results for Reversals. . . . .	54
2.3.5	Emergence of a giant cycle? . . . . .	54
2.4	The subcritical regime . . . . .	56
2.4.1	Preliminary results : comparison with a branching process . . . . .	57
2.4.2	Proof of Theorem 1 . . . . .	61
2.5	The critical regime . . . . .	63
2.6	The supercritical regime . . . . .	68
2.6.1	The cluster queuing system . . . . .	69
2.6.2	Completion of the proof of Theorem 24 . . . . .	71
2.6.3	The initial mass upstairs . . . . .	73
2.6.4	A sharper estimate for the mass upstairs . . . . .	74
<b>3</b>	<b>The hyperbolic geometry of random transpositions</b>	<b>82</b>
3.1	Introduction . . . . .	82
3.1.1	Asymptotic hyperbolicity . . . . .	83
3.1.2	The geometry of $G_n$ . . . . .	85
3.1.3	The hyperbolic constant under the uniform measure . . . . .	88
3.2	Asymptotic hyperbolicity under $\nu$ . . . . .	89
3.3	Large deviations and volume growth . . . . .	94
3.4	The uniform measure on $\partial B(an)$ . . . . .	98
3.5	Asymptotic singularity between $\mu$ and $\nu$ . . . . .	102
3.6	Number of geodesics and Radon-Nikodym derivative . . . . .	105
3.7	The size of the support of the hitting distribution . . . . .	106
3.8	Asymptotic hyperbolicity under the uniform measure . . . . .	112
3.8.1	8.0 . . . . .	115
3.8.2	Branching process asymptotics . . . . .	116
3.8.3	A modified branching process . . . . .	118
3.8.4	Fragmentations in the subcritical case . . . . .	120
3.8.5	Mean in the supercritical regime and duality . . . . .	123

3.8.6	Fragmentations in the supercritical range . . . . .	125
<b>4</b>	<b>Limiting behavior of the distance of a random walk</b>	<b>127</b>
4.1	Introduction . . . . .	127
4.2	Summary of main results . . . . .	129
4.2.1	Random $p$ -cycles . . . . .	129
4.2.2	Random adjacent transpositions . . . . .	130
4.2.3	Random walk on a random 3-regular graph . . . . .	133
4.2.4	Informal speculation . . . . .	134
4.2.5	Other random walks and some open problems . . . . .	135
4.3	Random $p$ -cycles . . . . .	136
4.3.1	Random hypergraphs, definitions and notations . . . . .	136
4.3.2	Critical point for random hypergraphs . . . . .	137
4.3.3	Bounds for the Cayley distance on the symmetric group	142
4.3.4	Phase transition for the 3-cycle random walk . . . . .	144
4.4	Random adjacent transpositions . . . . .	145
4.4.1	Trivial scaling . . . . .	146
4.4.2	Random walk scaling . . . . .	147
4.4.3	Brownian scaling . . . . .	151
4.5	Random walk on a random 3-regular graph . . . . .	153
4.5.1	Proof of subcritical regime . . . . .	154
4.5.2	Proof of the supercritical regime . . . . .	157

# List of Figures

1.1	<i>Of mice and men.</i> . . . . .	1
1.2	<i>Bourque-Pevzner's simulations, and Berestycki-Durrett's theoretical predictions.</i> . . . . .	4
1.3	Graph of the function $t \mapsto f(t)$ and of $t \mapsto f(t) - t$ . . . . .	6
1.4	In a hyperbolic space, $[y, z]$ is $\delta$ -close to $[x, y] \cup [y, z]$ . . . . .	23
1.5	Subcritical random graph with $c = 0.95$ for $n = 100$ vertices. On the left, the cluster $\mathcal{C}_1$ containing 1 has 13 vertices. On the right, the complement of $\mathcal{C}_1$ contains many vertices, but all clusters are small. . . . .	38
1.6	Critical random graph ( $c = 1$ ) for $n = 100$ vertices. On the left, the cluster $\mathcal{C}_1$ containing 1 has now 19 vertices. On the right, the complement of $\mathcal{C}_1$ is made up of comparable clusters. . . . .	39
1.7	Supercritical random graph with $c = 1.05$ for $n = 100$ vertices. On the left, the cluster $\mathcal{C}_1$ containing 1 is the giant component with 44 vertices. On the right, the complement of $\mathcal{C}_1$ is made up of small clusters. . . . .	40
1.8	Cayley graph of $\mathcal{S}_n$ generated by $S = \{\text{all transpositions}\}$ for $n = 5$ . At the bottom is the identity element, and permutations with equal distance from the identity are represented on the same level. The graph has $5! = 120$ vertices and $n! \binom{n}{2} / 2 = 600$ edges. . . . .	41
1.9	Cayley graph of $\mathcal{S}_n$ generated by $S = \{\text{adjacent transpositions}\}$ for $n = 5$ . At the bottom is the identity element, and permutations with equal distance from the identity are represented on the same level. The graph has $5! = 120$ vertices and $n!(n - 1)/2 = 240$ edges. . . . .	42
2.1	Breakpoint graph for human-mouse X chromosome comparison	76

2.2	Simulations with $n = 100$ markers. Average values of $k - D_k$ for 10,000 simulations of the random transposition chain, 10,000 simulations of $k - d_0(\pi_k)$ and Pevzner's 100 simulations of $k - d(\pi_k)$ for the reversal chain. The smooth curve of small triangles gives the limiting behavior as $n \rightarrow \infty$ of $(cn/2 - D_{cn/2})$ from Theorem 3. . . . .	78
2.3	Comparison of the distribution of the number of fragmentations in 10,000 simulations of the random transposition chain with the Poisson distribution with the same mean. . . . .	79
2.4	Growth of the average fraction of vertices in the cycle containing 1 at time $cn/2$ in 10,000 simulations of the random transposition chain with $n = 100$ and $n = 1000$ compared to $\theta(c)^2/2$ ( $\theta(c)$ being the percolation probability of the corresponding random graph with $p = c/n$ ). . . . .	80
2.5	Histogram of the size of the cycle containing 1 in 100,000 simulations of the random transposition chain with $n = 100$ at time 100 ( $c = 2$ ). The open squares give the distribution of the size of finite clusters in the corresponding random graph. . . . .	81
3.1	Two independent random walks run until they hit the sphere of radius $an$ . . . . .	90
3.2	Numerical evaluation by Maple of the limiting behavior of the quantity $-n^{-1} \log \frac{ S_n }{ \partial B }$ . The result is plotted as a function of $\xi = f^{-1}(a)$ , where $f(\xi) = 1 + \log(1 - \xi)(1 - \xi)/\xi$ . $f^{-1}$ is an increasing function of $a$ . The plot is only relevant for $a < 1/2$ i.e. $\xi < f^{-1}(1/2) \approx 0.715331863\dots$ . . . . .	126

# List of Tables

2.1	Order of the genes in <i>D. repleta</i> compared to their order in <i>D. melanogaster</i> . . . . .	77
-----	---	----

## **Abstract**

We study phase transition phenomena for the distance of random walks on graphs. In particular, we show how this question relates to the theory of random graphs and to stochastic processes of coalescence and fragmentation. This question is also intimately connected to some problems in genome rearrangement.

## Biographical sketch

Nathanaël Berestycki was born in Paris on December 6<sup>th</sup>, 1980. He studied mathematics successively at Lycée Henri IV, Ecole Normale Supérieure de Cachan, and Université Paris VI, before he moved to Cornell University first as a visitor and then as a full-time Ph.D. student. His thesis was co-supervised by Rick Durrett in Cornell and Jean-François Le Gall in E.N.S. Now aged 24, his first position will be at the University of British Columbia in Vancouver, Canada, as a postdoctoral fellow.

*To my parents.*

# Acknowledgements

Among all the people to whom I would like to express my gratitude, one of them deserves very special thanks. I would like to say to my Cornell advisor Rick Durrett what an exemplary advisor he has been to me, and what an inspiring experience it has been to be working closely with him for almost three years. Perhaps even more than his support and his availability, his open mind and his seemingly never-ending curiosity have had a major impact on me. Also, he is the one who made my stay in Cornell University financially possible, and his generosity in this matter is here acknowledged.

It is a great honor for me that Jean-François Le Gall has accepted to co-supervise this thesis. His many suggestions and advices, on mathematical questions, but also on more general ones, have meant a lot to me.

Throughout my thesis I have had the privilege to have extensive mathematical discussions with the following people: Gérard Ben Arous, Jason Fulman, Jim Pitman, Laurent Saloff-Coste, Oded Schramm, Marc Yor. I have always learnt a lot from them. In particular, I thank Gérard Ben Arous, Jason Fulman, and Oded Schramm for receiving me in their respective institutions (N.Y.U., U. Pitt. and Microsoft Research), and Marc Yor for his contagious enthusiasm. In fact, were it not for his brilliant teaching, my research would probably be on something else than probability theory. Thanks to Jim Pitman, Jason Schweinsberg and my brother (I'll come back to him!) for an opportunity to work together. Thanks also to Ruth Williams at UCSD and to Timo Seppalainen at the University of Wisconsin for inviting me to speak at their seminars. Finally, many thanks to Greg Lawler and Laurent Saloff-Coste for seating in my committee.

Around me, everyday, there were other graduate students and postdocs with whom it was always fun to talk and work. I'd like to thank, in Cornell, Christian Beneš, Christophe Garban (who proofread so carefully the

introduction to this thesis - thanks, man!), Paul Jung, Mike Kozdron, José Trujillo Ferreras, and Brigitta Vermesi. In Paris, I should thank the V6 as a whole - Florent Benaych-Georges, Mathieu Merle, Mathilde Weill. Thanks in particular to Florent and Mathilde for an unforgettable hot summer in 2003! I also want to thank all my friends. Even though I was so often more than 5000km away, they have managed to remain great friends. It is of course impossible to list them all, but it would really be unfair not to mention the following people for constant support and fun: Guillaume, Olivier Z., Flora, and, of course, Dan. Thanks for being such an awesome roommate!

This list is getting way too long. But how can I not say a few words about my family? It is impossible to thank appropriately my parents, but I want to give it a try by dedicating this thesis to them. It's not just that I wouldn't be writing this without them, but it's their lovingness, their patience, and, simply put, it's how much I've learnt from them, that I have in mind here. I want to say my love for my brother and sisters Julien, Clara and Noémie, to my grandmother, to Marie-Anne, and to my cousins. In particular at the time of writing this we are preparing the celebration of the wedding of my brother Julien with Valentine. I want to wish them all the best and tell them how much it means to me to be so close to them.

Oh, I almost forgot. Thanks to Bach and Mozart for writing so many masterworks. I am sure they have also played a role in this thesis.

# Chapter 1

## Introduction

### 1.1 Introduction

#### 1.1.1 From genome evolution to random walks

We summarize here briefly how a problem in genome rearrangement leads to questions about the distance of a random walk from its starting point.

A major problem in modern computational biology is to understand in a quantitative way the effect of large-scale mutations on genomes and how this may affect the rate of divergence in the evolution of two species. A typical situation is the following.

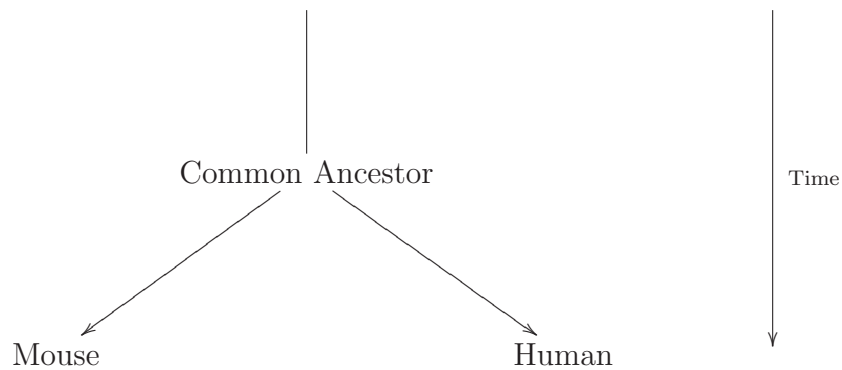


Figure 1.1: *Of mice and men.*

Suppose we ask the following question about the mouse and the human:

*How much time has elapsed since their most recent common ancestor?*, or in other words, what is the *divergence time* between the mouse and the human?

As can be expected, this relatively old question may be answered in quite many different ways. One possible approach to this question, often used by computer scientists and biologists working in this field (see for instance Bourque and Pevzner (2002) and references therein), may be called the *parsimony method* for large-scale mutations. To explain this method, we make the following remark. There are many genes that humans share with mice, and moreover within one species there is little polymorphism as to how those genes are ordered on a chromosome, so that it makes sense to compare the relative order of a given set of common genes on chromosomes in the two species. This order may evolve with time due to certain *large-scale mutations*, i.e., mutations that affect entire regions of a chromosome (containing from dozens to hundreds of genes), rather than single nucleotides. These mutations are assumed to affect different species independently.

The parsimony method proceeds as follows. We focus on one particular type of those large-scale mutations. One possible example is that of *reversals*, which select two genes on a chromosome and flips over the whole segment between those two extremities. Then, on some suitably chosen portion of a suitably chosen chromosome, we compare the order of corresponding genes between the two species, and ask: what is the minimal number of reversals that must be performed on this portion of the human genome to match the corresponding portion of the mouse genome? This approach makes sense if the chromosome chosen for comparison is the X chromosome (or if we compare mitochondrial DNA), since in these two cases reversals are the only large-scale mutations that change the relative order of genes within the chromosome (translocations, for instance, involve necessarily another chromosome, and so never involve the X chromosome).

Indeed, given this number  $d$  (which, as we will see, is in fact a distance between two permutations for some suitable graph on the symmetric group), we may give the following estimate on the divergence time. It is possible to approximate the rate  $\rho$  of such mutations. On each branch of the evolutionary paths (common ancestor to mouse and common ancestor to human) there have been at least roughly  $d/2$  reversals performed, corresponding to a duration of approximately

$$\tau = \frac{d}{2\rho}$$

Actually, on the examples where we have carried out numerical investigations, it is more meaningful to proceed backwards: there already exist good estimates on the divergence time of mouse and human, and we can use  $d$  and  $\tau$  to estimate the rate  $\rho$  at which those mutations happen. In the case of mouse and human we find that  $\rho$  is, roughly, 1 to 2 mutations per genome per million years. We can also use this method with other species and ask whether this rate is the same for every species. Surprisingly, some species have much higher rates of mutations. The explanation has to be found in the way large-scale mutations affect species.

Coming back to the problem of estimating the distance  $d$  between two genomes (i.e., the minimal number of reversals or other mutations needed to transform one genome into the other), the point is that it is sometime possible to compute this quantity very efficiently. In particular, in the case of reversals, Hannehalli and Pevzner (1995) came up with a polynomial-time algorithm that computes the required quantity, making this approach to estimating the divergence time  $\tau$  (or the rate  $\rho$ ) a very convenient one.

However, this approach has one major drawback, which lies in the fact that the parsimony approach only gives a *lower bound* on the true number of reversals that have occurred on the evolutionary path leading from the mouse to the human. Therefore it becomes of critical importance for this method to know when the parsimony approach gives a reliable estimate of the true number of reversals performed, i.e., when this lower bound is accurate and, in fact, gives the actual number of reversals performed.

Bourque and Pevzner (2002) have approached this problem through simulations. They start with  $n$  genes in order

$$1 \ 2 \ 3 \ \dots \ n$$

and successively perform a certain number of *random* reversals, i.e., at each step they choose uniformly at random two points on this string, and then flip over the segment in between those two points. After each random reversal, they compute  $d(k)$ , i.e., the minimum number of reversals that one needs to perform to transform the original string  $1 \ 2 \ 3 \ \dots \ n$  into the one obtained after  $k$  random reversals. Of course, by definition one has  $d(k) \leq k$ .

The question of whether the parsimony approach is reliable translates directly into knowing whether  $d(k)$  is approximately  $k$  or not. Indeed when  $d(k) \approx k$  this means that the distance gives an answer close to the actual

number of reversals that have been performed. On the other hand if  $d(k) - k$  is comparable to  $k$  this means that the parsimony approach underestimates by a large amount the true number of reversals performed, and thus the time elapsed since the most recent common ancestor. Figure 1.2 below summarizes the results of their simulations, where they have plotted the function  $d(k) - k$  against the number  $k$  of random reversals performed.

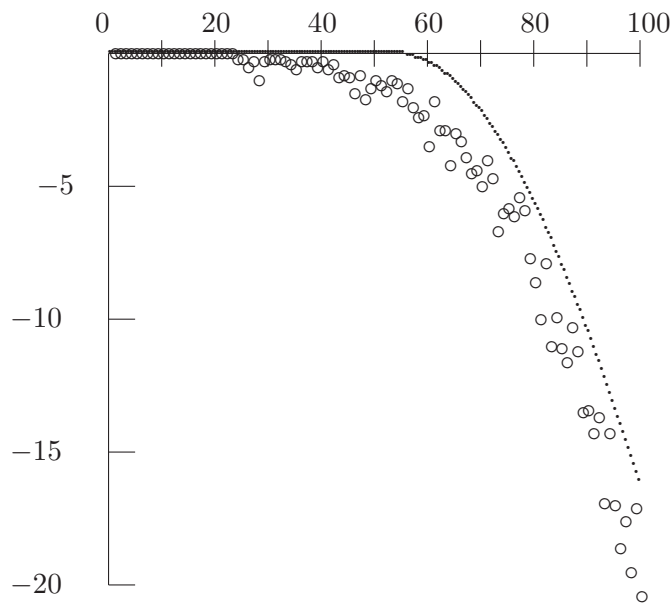


Figure 1.2: *Bourque-Pevzner's simulations, and Berestycki-Durrett's theoretical predictions.*

These simulations have been performed for  $n = 100$  markers, and the conclusion drawn from those simulations by Bourque and Pevzner was that the parsimony approach gives a reliable estimate only as long as  $k \leq 0.4n$ , or, equivalently, only as long as the computed distance  $d(k) \leq 0.4n$ . The solid line next to the circles represents the behavior of  $d(k) - k$  predicted by myself and Rick Durrett when we let  $n \rightarrow \infty$ . This has the advantage of correcting the factor 0.4 from Bourque and Pevzner into a 0.5. Moreover, at least from a theoretical point of view, it is now possible to actually infer  $k$  from  $d(k)$  for all  $k$  (even beyond  $k \geq 0.5n$ ) since, as will be seen in Theorem

1, the function  $d(k)$  is asymptotically increasing with an explicit expression, so this step requires only an inversion of the function, which is trivial to do numerically. However since the function becomes pretty flat this leads to poor estimates.

### 1.1.2 Phase transitions of random walks

Implicitly in the simulations of Bourque and Pevzner, the problem has been translated into a mathematically more tractable one about random walks. Indeed, consider the graph whose set of vertices consists of all rearrangements  $\sigma$  of the string  $\{1, 2, \dots, n\}$ , and put an edge between two rearrangements  $\sigma$  and  $\sigma'$  if applying some reversal to  $\sigma$  leads to  $\sigma'$ . Then performing random reversals to the string  $\{1, 2, \dots, n\}$  is exactly equivalent to follow the successive positions occupied by a *random walk* on this graph, i.e., a random process that starts at the *identical* rearrangement  $\sigma(i) = i$  for  $1 \leq i \leq n$ , and then chooses one of its neighbors uniformly at random to move there at the next step. Moreover, in this context, the parsimony distance between the initial ordering  $\{1, 2, \dots, n\}$  and the current state is simply the *graph distance* between the current position of the random walk and its starting point, i.e., the minimal number of edges that one needs to cross on this graph to reach the current state starting from the identical ordering. In this framework, the observations of Bourque and Pevzner (2002) are made rigorous by the following Theorem, which is a consequence of Theorem 6. Let  $d(k)$  be the parsimony distance of the current position after  $k$  random reversals.

**Theorem 1.** *Let  $t > 0$ . The following convergence holds in probability as  $n \rightarrow \infty$*

$$\frac{1}{n}d(\lfloor tn \rfloor) \rightarrow_p f(t)$$

where

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

Thus the random walk framework makes it possible to give a rigorous treatment of the biological question explained in the previous question, but it also suggests interesting mathematical behavior. Indeed, the result above, or, more importantly from a purely mathematical point of view, Theorem

6 below, describes the behavior of the distance of a certain random walk, (random reversals, or the random transposition random walk, which is a simplified version of the random reversals). It tells us that this behavior has a very unexpected *phase transition*: despite all the regularity and the symmetries of the underlying graph, the distance evolves *linearly* for a certain range (with a coefficient equal to 1), until at some critical time it abruptly becomes *sublinear*. Most surprising is the fact that the function  $f$ , which describes the asymptotic behavior of the distance of the walk, has a non-analytic point at  $t = 1/2$ , corresponding to this critical time. The change of behavior from linear to sublinear is also exactly the point until which the parsimony method gives an accurate approximation.  $t = 1/2$  then corresponds to the fact that the correct factor in Bourque and Pevzner's simulations should be a 0.5 asymptotically, rather than 0.4.

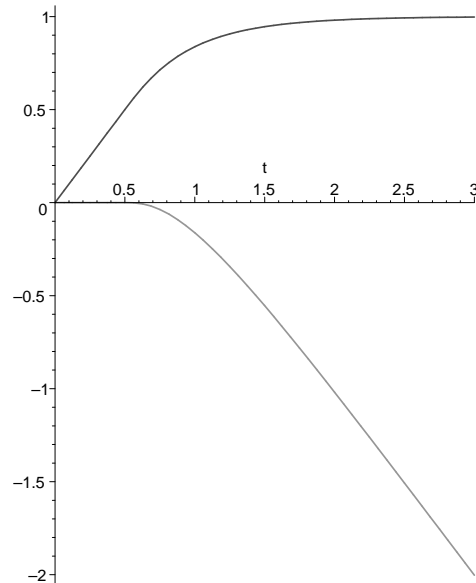


Figure 1.3: Graph of the function  $t \mapsto f(t)$  and of  $t \mapsto f(t) - t$ .

The second curve representing  $f(t) - t$  is of course the same as the solid line shown in the picture of Bourque and Pevzner's simulations.

### 1.1.3 Informal overview of the results in my thesis

In my thesis I have studied from a mathematical point of view the existence of phase transitions for the distance of a random walk, as described above, for different examples of random walks in various contexts.

In many examples, I have been able to show how the existence of such phase transitions can be related to the existence of large clusters in certain random graphs. In particular Erdős-Renyi random graphs (i.e., mean-field percolation) play a crucial role. Often these systems exhibit a phase transition from a phase where all clusters are small to a phase where there is a single "giant" component. By studying the properties of these objects we can often gain very useful insight on the behavior of the random walks.

I have also approached the problem of these phase transitions from the point of view of the geometry of the underlying graph, and have proved a connection with hyperbolic geometry.

Finally, discussing different examples has led me to construct an approximate picture of what may be the different possibilities for the behavior of the distance versus time for a random walk. Based on the examples that we study, it seems that there are only two types of possible behaviors. The first one is the diffusive behavior (this is for instance the case of simple random walk on  $\mathbb{Z}^d$ ), and the other one is the case of a phase transition from a linear to a sublinear behavior. The case of random transpositions is the prototypical example of this type of behavior, but we will also discuss other random walks with similar phase transitions.

### 1.1.4 Relation to classical theory of card shuffling and the cut-off phenomenon

Classically, the literature about random walks on finite graphs and groups has focused on one major question, which concerns the rate of convergence of a given Markov chain, reversible or non-reversible, towards its equilibrium distribution. A useful example to bear in mind for this theory is that of card shuffling. We can represent the ordering of a deck of 52 cards by a permutation in the set  $\mathcal{S}_{52}$  of the  $52!$  permutations of those 52 cards. Repeatedly shuffling the deck of cards according to some method is then represented by the successive states of a certain Markov chain with transition probabilities  $p(\sigma, \pi)$  corresponding to the particular way the shuffling is done. For

instance, if the method chosen consists of selecting two cards at random among the  $n = 52$  cards in the deck, and then interchanging them (which, albeit a highly unrealistic approach to actually shuffling cards, is easy to implement in a computation), then  $p$  has the form

$$p(\sigma, \pi) = \begin{cases} 1 / \binom{n}{2} & \text{if } \sigma = \pi\tau \text{ for some transposition } \tau \\ 0 & \text{otherwise} \end{cases} \quad (1.1)$$

This Markov chain is an example of a *random walk* on the symmetric group  $\mathcal{S}_n$  where the set of generators is all transpositions. In general, if  $S$  is a set generating the group  $\mathcal{S}_n$  we say that the Markov chain  $(X_k; k \geq 0)$  is the random walk with set of generators  $S$  if the transition probabilities of  $X$  are given by  $p(x, y) = |S|^{-1}$  if  $yx^{-1} \in S$  and 0 otherwise. Actually, it will often be more convenient for us to work with *continuous-time* (rate 1) random walks, meaning that at each state we wait an independent exponential (mean 1) amount of time before jumping according to the transition kernel  $p$ . The embedded discrete-time random walk is sometimes called the *discrete skeleton* of the walk.

In the context of card shuffling, the interesting cases are those when the equilibrium distribution of the Markov chain  $X$  is the uniform distribution; the main question of interest is then to ask *How many moves does it take to bring the deck of cards close to the uniform distribution?*

The first breakthrough about this question came in a well-known paper by Diaconis and Shahshahani in (1981), where they proved the following. If we take the random transposition random walk (1.1) in continuous time (to avoid parity/periodicity problems) then the convergence to the equilibrium (uniform) distribution occurs abruptly, after  $\frac{1}{2}n \log n$  shuffles. This phenomenon is known as the *cut-off* phenomenon. It has the following precise mathematical meaning:

**Theorem 2.** Diaconis and Sahshahani, 1981. *For any  $\varepsilon > 0$ , the total variation distance between the law of the chain at time  $(1 - \varepsilon)\frac{1}{2}n \log n$  and the uniform distribution converges to 1 as  $n \rightarrow \infty$ , while the same quantity at time  $(1 + \varepsilon)\frac{1}{2}n \log n$  converges to 0.*

Since then, the cut-off phenomenon has been observed in a wide variety of contexts. An exposition of this fascinating subject can be found in the

well-known book of Diaconis (1988). See also the survey by Saloff-Coste (2003) for a more recent presentation of this area with a slightly different perspective. We refer the reader to Durrett (2003) for applications of these ideas to questions in evolutionary biology.

As we see, this theory is concerned with the behavior of random walks at a rather later stage of their evolution than we presently are in Theorem 1. In some sense the question we are interested in is a refinement of the mixing time question, since for instance clearly mixing cannot occur before the walk has reached a distance corresponding to the average diameter of the group. Also, from this question, one gains a better understanding of the way the random walk converges to its equilibrium distribution. For this, see in particular the result of Oded Schramm on random transpositions (that we will discuss at the end of the introduction), and the case of the random regular graph, where the analysis of the distance question (Theorem 17) leads to a conjecture about the mixing time on this graph.

## 1.2 Erdős-Renyi Random Graphs

Erdős-Renyi random graphs play a major role in the proof of the phase transition for random transpositions, so we start by recalling some basic properties of these objects.

The theory of random graphs was born towards the end of the 50's in a series of papers by Erdős and Renyi, where they asked the following question. Suppose we take  $n$  vertices and a number  $p = p(n)$  with  $0 < p < 1$ . Let  $G(n, p(n))$  be the graph such that there is an edge between any two given vertices  $i$  and  $j$  with probability  $p(n)$ , and the existence of an edge between two vertices are independent events for different pairs of vertices. We denote by  $\mathbf{Q}_p = \mathbf{Q}_{p(n)}^n$  the law of this random variable on the space of all graphs with  $n$  vertices. (The graph  $G(n, p(n))$  is now known as an Erdős-Renyi random graph). The question asked by Erdős and Renyi was the following: what can be said about the properties of the graph for various scaling of the function  $p(n)$ ? To avoid trivialities, (there is always a positive probability that there is no edge at all in the graph, for instance), we make the following definition.

**Definition 1.** *A property  $\Pi$  is said to hold with high probability (or asymptotically almost surely, a.a.s.) if*

$$\mathbf{Q}_{p(n)}(\Pi) \rightarrow 1$$

as  $n \rightarrow \infty$ .

With this definition, Erdős and Renyi were able to answer questions like: how should we choose  $p(n)$  so that *with high probability* the graph  $G(n, p(n))$  is connected? Clearly, the larger  $p$ , the likelier it is to be connected, so it is interesting to see what is the "smallest" value of  $p(n)$  that will make  $G(n, p)$  connected with high probability.

Erdős and Renyi discovered the fascinating fact that many properties of random graphs appear all of a sudden, with a threshold:

**Definition 2.**  $\hat{p} = \hat{p}(n)$  is a threshold function for the property  $\Pi$  if

$$\begin{cases} \text{if } p \ll \hat{p} & \text{then } \mathbf{Q}_p(\Pi) \rightarrow 0 \\ \text{if } p \gg \hat{p} & \text{then } \mathbf{Q}_p(\Pi) \rightarrow 1 \end{cases}$$

Here  $a_n \ll b_n$  means  $a_n/b_n \rightarrow 0$ , which we will also denote by  $a_n = o(b_n)$ . Chapter 1 of Janson et al. (2000) contains many general results about the existence of thresholds and how to quantify their sharpness.

A typical example of this phenomenon is connectedness of the random graph.

**Theorem 3.** Erdős and Renyi (1960). *Connectivity of  $G(n, p)$  has a sharp threshold function  $\hat{p}(n) = \log n/n$ .*

In this result the word *sharp* refers to the fact that for any  $\delta > 0$  if  $p = (1 - \delta)\hat{p}(n)$  then  $\mathbf{Q}_p(G \text{ is connected}) \rightarrow 0$  while if  $p = (1 + \delta)\hat{p}(n)$  then  $\mathbf{Q}_p(G \text{ is connected}) \rightarrow 1$ .

As we will see when we describe the basic correspondence between the random transposition random walk and the random graphs (Theorem 5), the above result is related to the convergence to equilibrium of the random walk in time  $(1/2)n \log n$  (Theorem 2), although it seems hard to mathematically deduce Theorem 2 from Theorem 3. However Theorem 20, which is due to Schramm, will shed some unexpected light on this question.

As was mentioned, we will be interested in what happens to the random walk *before* it converges to equilibrium, so it is necessary to describe what does the random graph look like for smaller values of  $p$ .

It turns out that such a description reveals a fascinating feature of random graphs, which is sometime called the birth of the giant component. This phenomenon is perhaps better described by first introducing a time-evolving graph process.

To each edge of the complete graph  $K_n$  (i.e., the graph consisting of all  $\binom{n}{2}$  possible edges between pairs of vertices), we associate an independent exponential random variable with mean  $\binom{n}{2}$ . We then look at the process  $\{G(n, t); t \geq 0\}$  where for all  $t \geq 0$   $G(n, t)$  is defined as the graph whose edges have ringed by time  $t$ , by which we mean that we put an edge between  $i$  and  $j$  (with  $i < j$ ) if and only if the corresponding exponential random variable  $X_{ij} < t$ .

**Properties.** The following properties are obvious:

- (i) As time evolves the only modifications we make to  $G(n, t)$  are to add edges to existing edges.
- (ii) Conditioned on having  $k$  edges at time  $t$ ,  $G(n, t)$  is uniform on all graphs with  $k$  edges on  $n$  vertices.
- (ii) At a fixed time  $t$ ,  $G(n, t)$  has the distribution of an Erdős-Renyi random graph  $\mathbf{Q}_p$  with  $p = 1 - \exp(-t/\binom{n}{2})$ .

This procedure is analogous to the *coupling* in percolation, where a realization of  $\mathbf{P}_p$  has been constructed on the same probability space for all values of  $0 < p < 1$ . However here the underlying graph is the complete graph rather than the usual  $\mathbf{Z}^d$  lattice. In this "coupled" version, Theorem 3 can be stated as: it takes time  $(n \log n)/2$  for  $G(n, t)$  to be connected with high probability.

The result announced concerns the emergence of a giant cluster that encompasses a positive fraction of all vertices much before the graph becomes completely connected. More precisely, the size of the largest connected component has a phase transition at  $p = c/n$  for  $c = 1$ , or equivalently, at times  $cn/2$  with  $c = 1$  in the graph process.

**Theorem 4.** (*Erdős-Renyi, 1959*). Suppose  $p(n) = c/n$  for  $c > 0$ . As  $n \rightarrow \infty$ ,

<i>Regime</i>	<i>Size of largest component</i>
$c < 1$ ( <i>subcritical regime</i> )	$\frac{3}{(1-c)^2} \log n$
$c = 1$ ( <i>critical regime</i> )	$O(n^{2/3})$
$c > 1$ ( <i>supercritical regime</i> )	<i>Largest</i> = $\theta(c)n$  <i>Second</i> = $O(\log n)$ .

Here,  $\theta(c)$  is the probability of survival for a branching process whose offspring distribution is a Poisson random variable with mean  $c$ .

This has the precise meaning that the ratio of the size of the largest component to  $3(1 - c)^{-2} \log n$  (when  $c < 1$ ), or to  $\theta(c)n$  (when  $c > 1$ ), converges to 1 in probability. When  $c = 1$  there is a non-degenerate limit in distribution, although a satisfactory description of the limiting distribution has only recently been given by Aldous (1997).

Figures 1.5, 1.6 and 1.7, at the end of this introduction, illustrate this phase transition. The simulations have been realized using Matlab, and taking  $n = 100$  vertices. The figures represent the state of an Erdős-Renyi graph at  $c = 0.95$ ,  $c = 1$  and  $c = 1.05$  and show a drastic change in the cluster containing a fixed vertex, say 1. For comparison, we have also represented on the right-hand side of the figures the complement of this cluster - notice how this gets smaller and smaller.

We will sketch the proof of this result a little below, but we first wish to comment on the content of Theorem 4. Looking at things from the point of view of the graph process, this can be rephrased as the following. At times  $cn/2$ , if  $c < 1$ , all clusters are smaller than  $O(\log n)$ . However when  $c = 1$  the logarithmic factor blows up: we have reached a phase where clusters are large enough and they get interconnected more easily, which gives rise to several large components of order  $n^{2/3}$ . (Aldous (1997) has given a representation of

the asymptotic state of a random graph near the point of its phase transition involving the ranked lengths of the excursions of a certain Brownian motion with parabolic drift). However when  $c$  becomes greater than 1, all of the large clusters merge into a single *giant component*, whose size is approximately  $\theta(c)n$ , whereas all other clusters are only of order  $\log n$  at most. From then on until the end of the process, as soon as a cluster becomes reasonably large, it get absorbed into the giant component, until finally all vertices are in it, which happens by Theorem 3 at time  $n \log n/2$ .

*Sketch of the proof.* The proof of Theorem 4 relies on a limit theorem for the size of the connected component containing a distinguished vertex, which can be identified as the total progeny of a branching process with offspring distribution a Poisson random variable with mean  $c$  (abbreviated into  $PGW(c)$ , for Poisson Galton-Watson process). This limit theorem will also be very useful in the understanding of the random transposition random walk. Recall that  $(Z_i, i \geq 0)$  is a  $PGW(c)$  process if it is a Markov chain on the nonnegative integers  $\mathbf{N}$ , with transition probabilities  $p(k, \cdot) = \mu^{*k}(\cdot)$  where  $\mu$  is the Poisson( $c$ ) distribution and  $\mu^{*k}$  denotes  $k$ -fold convolution of  $\mu$  with itself. Let  $Z = \sum_{i=0}^{\infty} Z_i$  be the total progeny of this process when started with  $Z_0 = 1$  individual.

**Proposition 1.** *Let  $\mathcal{C}_1$  be the cluster that contains vertex 1. Let  $c > 0$  and  $p = c/n$ . Then as  $n \rightarrow \infty$*

$$\mathbf{Q}_p(|\mathcal{C}_1| = k) \rightarrow P(Z = k)$$

*Proof.* The number of children of vertex 1,  $Z_n^1 = |Y_1|$  has distribution Binomial  $(n - 1, c/n)$  since it has  $n - 1$  potential neighbors, each being an actual neighbor with probability  $c/n$ . This converges to a Poisson( $c$ ) limit by elementary computations with Stirling's formula. We can proceed one generation further: given that 1 is connected to exactly  $n_1$  other vertices, each of the  $n_1$  children is connected to a Binomial( $n - 1 - n_1, c/n$ ) number of individuals. This also converges to Poisson( $c$ ) distribution. However this time we need to be a little careful, since some of these children in the second generation could be the same for different individuals in the first. An easy calculation shows that these can be neglected asymptotically. Proceeding like this with further generations gives the proof.  $\square$

Armed with this Proposition, we now finish the sketch of the proof of Theorem 4. When  $c < 1$ , the expected number of offsprings is  $< 1$  so the

branching process dies out. Roughly, there are  $n$  clusters of finite size  $Z$  with exponential tails, so assuming independence of the size of different clusters gives that the largest cluster is of order  $O(\log n)$ .

When  $c > 1$ , there is a probability  $1 - \theta(c)$  that  $PGW(c)$  dies out. This means that for roughly  $(1 - \theta(c))n$  vertices, their cluster dies very quickly. For the remaining  $\theta(c)n$  vertices, their cluster grows quite large, but this observation is not enough to understand the existence of the giant component: it remains to see why there is only one such large component. In other words, we need to see why all vertices whose cluster is quite large are actually part of the same cluster. In the Proposition above, we proved finite-dimensional distribution convergence toward those of a  $PGW(c)$ . However a more sophisticated argument (a consequence of the birthday problem) shows that this approximation is reasonable at least until  $n^{1/2}\omega(n)$  have been exposed in this growing procedure, where  $\omega(n) \rightarrow \infty$  slowly enough. Let  $x$  and  $y$  be two vertices whose clusters are larger than  $n^{1/2}\omega(n)$ , and suppose that by the time each has size  $n^{1/2}\omega(n)$ , they haven't intersected yet. We show that it is likely that they will intersect at the next generation. Indeed, there are roughly  $n^{1/2}\omega(n)$  vertices in the last generation of each cluster, which means that there are potentially  $N = \binom{n^{1/2}\omega(n)}{2}$  edges between these two sets. Therefore, the probability that there is no edge between the two sets of vertices is smaller than

$$(1 - c/n)^N \leq \exp(-\omega(n)^2) \rightarrow 0$$

which explains the uniqueness of the giant component. For more on this approach (and real proofs), see the recent notes of Durrett (2005).

## 1.3 Random transpositions: the phase transition.

### 1.3.1 Coupling with an Erdős-Renyi random graph.

We first define properly the main object of this study, the random transposition random walk. This is the Markov chain in the symmetric group  $\mathcal{S}_n$  (i.e., the space of all  $n!$  permutations of  $n$  symbols), starting at the identity element, and whose transition probabilities are described by the kernel  $p$  given in (1.1). In words, at each step, we pick two symbols uniformly at random

and exchange them. This might seem a very unrealistic simplification of the random reversals process that we considered at the beginning, where, instead of simply exchanging the two extremities, we also flipped around the whole segment. In fact, this is not the case, since careful analysis of Hannehalli and Pevzner's (1995) algorithm shows that it shares an essential feature with this random transposition random walk, namely, the existence of a discrete coagulation-fragmentation chain, which allows us to translate our results for random transpositions to random reversals.

In fact, it will be more convenient for us to use a continuous-time version of this walk. Let  $(\sigma_t; t \geq 0)$  be the random Markov process in continuous time taking its values in  $\mathcal{S}_n$  with  $\sigma_0 = e$  the identity, and whose discrete skeleton has transition probabilities (1.1). Let  $D(t)$  be the minimum number of transpositions needed to construct  $\sigma_t$  starting at the identity. Equivalently, if  $\mathcal{S}_n$  is viewed as a graph  $G_n$  with edges between permutations  $\sigma$  and  $\sigma'$  when  $p(\sigma, \sigma') > 0$  (i.e., there is some transposition  $\tau$  such that  $\sigma = \sigma'\tau$ ), then  $D(t)$  is the graph distance between  $\sigma_t$  and the identity, i.e. the minimum number of edges one needs to cross to go from the starting point of the walk to its current position.

The aim of this section is to present convincing evidence that there is a phase transition for  $D(t)$  when  $t = n/2$ , and to relate this phase transition to that of the Erdős-Renyi random graph.

The first step in proving this result (Theorem 6) is to see that if  $\sigma$  is a permutation then its distance to the identity is related to the number of cycles in the cycle decomposition of  $\sigma$ . Recall that the cycle decomposition of a permutation is the decomposition of  $\{1, \dots, n\}$  into disjoint orbits under the action of  $\sigma$ : hence the cycle that contains 1 consists of  $1, \sigma(1), \sigma^2(1), \dots$  and eventually  $\sigma^k(1) = 1$  again, so the cycle has been completed and we can proceed with the next cycle, which is the one containing the smallest integer not used by the previous cycles. For instance the permutation which maps

$$\begin{array}{c|cccccccc} k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \sigma(k) & 3 & 9 & 6 & 8 & 2 & 1 & 5 & 4 & 7 \end{array}$$

has cycle decomposition  $(136)(2975)(48)$ .

Then  $D(\sigma)$  is obtained by the simple formula, first discovered by Cayley (1889)

$$D(\sigma) = n - \#\text{cycles of } \sigma \tag{1.2}$$

Therefore to evaluate  $D(t)$  it suffices to know how many cycles  $\sigma_t$  is made of. We now make a fundamental remark that under the random transposition random walk, the cycles of  $\sigma_t$  evolve according to a *discrete coagulation and fragmentation chain*. Suppose that at a given time  $t$  we apply a transposition  $(x, y)$  to the permutation  $\sigma_{t-}$ , with  $x$  and  $y$  in different cycles  $C$  and  $C'$  of this permutation. Then the cycle decomposition of  $\sigma_t$  is the same as that of  $\sigma_{t-}$  except that the cycles  $C$  and  $C'$  have been replaced by a unique cycle  $C \cup C'$ : we say that  $C$  and  $C'$  have merged. On the other hand, suppose that  $x$  and  $y$  belong to the same cycle  $C = (x_1, \dots, x_k)$  of the permutation  $\sigma_{t-}$ . If  $x = x_i$  and  $y = x_j$  with  $x_i \neq x_j$  then  $C$  breaks into  $(x_1, \dots, x_{i-1}, x_j, x_{j+1}, \dots, x_k)$  on the one hand and  $(x_i, \dots, x_{j-1})$  on the other hand: we say that  $C$  is being *fragmented*. In particular, conditionally of making a fragmentation of the cycle  $C$ , then the two pieces have respective size  $U$  and  $|C| - U$ , where  $U$  is a uniform random variable on  $\{1, \dots, |C| - 1\}$ .

We are now ready to describe the coupling with the Erdős-Renyi random graph. As we perform the random walk, we draw a graph  $G(t)$  by the simple following procedure: whenever a transposition is applied to  $\sigma_{t-}$  involving  $i$  and  $j$ , we also draw an edge between  $i$  and  $j$  in  $G(t)$ . For convenience we change the time-parametrization: we work at times  $cn/2$  rather than  $tn$  (as in Theorem 1). As proved by the next result, this has the great advantage of making  $G(cn/2)$  an Erdős-Renyi random graph with standard scaling parameters. Although very simple to prove, this is an essential result of my thesis:

**Theorem 5.** *At time  $cn/2$ ,  $G(cn/2)$  is a realization of the Erdős-Renyi random graph  $\mathbf{Q}_{p_n}$  with  $p_n \sim c/n$ : in fact  $G(t)$  has the law of the evolving graph process of Proposition 1.2. Moreover, in this coupling, any cycle  $C$  of  $\sigma_t$  is a subset of some connected component of  $G(t)$ .*

This result is quite simple to understand. In  $G(t)$  the edges are independent because we can think of the edges (or, equivalently, the transpositions) arriving according to some Poisson point process on the space of all  $\binom{n}{2}$  edges. Since the arrivals of points are independent in a Poisson point process, this proves the independence of edges in  $G(t)$ . Moreover, if  $e$  is an edge,

$$P(e \in G(t)) = 1 - \exp(-t/\binom{n}{2})$$

by the same Poisson point process approach. When  $t = cn/2$ , this gives  $p_n \sim c/n$ . The fact that the cycles of  $\sigma_t$  are subset of connected components

of  $G(t)$  is equally simple. As we draw the graph  $G(t)$  the clusters can only experience coagulation, whereas from time to time there is a fragmentation in the cycles of  $\sigma_t$ : hence a cluster in  $G(t)$  may correspond to several cycles in  $\sigma_t$ . In fact we can be a little more precise in this description. We see that, except for double or multiple edges, which cannot be seen on  $G(t)$ , fragmentations are related to components with cycles in  $G(t)$ . We say that  $C$  is a *cyclic* component of  $G$ , or that it has *positive complexity* if there is a closed path (i.e., starting and ending at the same vertex) that uses only edges in  $C$  and no edge more than once. In the opposite case,  $C$  is a *tree*. So, if  $t$  is a time such that there is a fragmentation at this time, caused by some transposition  $(i, j)$ , then  $i$  and  $j$  must be part of the same cycle in  $\sigma_{t-}$ . Hence we see that except if  $(i, j)$  was already an edge in  $G(t^-)$ , this transposition puts an edge between two vertices which were previously in the same cycle of  $\sigma_{t-}$ , and hence in the same connected component. Therefore this move increases the complexity of the component: in particular, if it was a tree before, it now has a cycle. This fact will be very convenient when dealing with the subcritical case, where few fragmentations happen (Theorem 7). In the other regimes, and in particular in the supercritical regime, clusters of the random graph may have a large complexity, so the relationship between fragmentation and complex components is less clear.

### 1.3.2 Law of large numbers behavior for random transpositions

We recall that  $D(t)$  denotes the distance to the identity from  $\sigma_t$ . The following result describes the mean behavior for  $D(cn/2)$ .

**Theorem 6.** *Let  $c > 0$ . The following convergence holds in probability as  $n \rightarrow \infty$*

$$\frac{1}{n}D(cn/2) \rightarrow_p u(c)$$

where

$$u(c) = \begin{cases} c/2 & \text{for } c \leq 1 \\ 1 - \sum_{k=1}^{\infty} \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k < c/2 & \text{for } c > 1 \end{cases} \quad (1.3)$$

*Sketch of the proof.*

A simple consequence of formula (1.2) is the following expression for  $D(t)$

$$D(t) = N_t - 2 \sum_{s \leq t} \mathbf{1}_{\{F_s\}} \quad (1.4)$$

where  $N_t$  is the total number of transpositions applied before time  $t$  and  $\mathbf{1}_{\{F_s\}}$  denotes the indicator of the event that there is a fragmentation at time  $s$ , so the sum counts the number of such fragmentations. Indeed, the distance can only increase or decrease by 1 at each move, and (1.2) tells us that it increase by 1 with a coagulation and decreases by 1 with each fragmentation. Hence, if we count ”+1” for each transposition (adding up to  $N_t$ ), we also have to subtract 2 for each fragmentation.

If  $c < 1$  it is well-known that there are only a bounded (in expectation) number of complex components in  $G(n, c/n)$ . Therefore the number of fragmentation prior to  $cn/2$  is  $o(n)$  and by (1.4), it follows that  $D(cn/2) \approx N_{cn/2}$ , which is a Poisson random variable with mean  $cn/2$ .

The case  $c > 1$  is much more delicate, since it seems at first that by (1.4) we have to show that there is a lot of fragmentation, meaning that using  $G(n, c/n)$  facts may not be very relevant. However, a quick argument shows that at time  $cn/2$  the difference between the number of cycles in  $\sigma_{cn/2}$  and the number of clusters in  $G(cn/2)$  is  $o(n)$ . This may seem paradoxical at first, but our work on the supercritical regime (Theorem 9) sheds light on this phenomenon. In fact what is happening is that even though there are many fragmentations happening, many of those fragments get reabsorbed by larger cycles fairly quickly. Therefore, when we count the cycles at a given time, compared to the random graph, we only have to add those extra cycles that have been generated by fragmentation and have not yet been reabsorbed by a larger cycle. This is of course much smaller than the total number of fragmentations.

On the other hand, counting the number of clusters in  $G(n, c/n)$  is an easy task. Indeed each given tree on  $k$  given vertices has probability

$$\begin{aligned} & \left(\frac{c}{n}\right)^{k-1} \left(1 - \frac{c}{n}\right)^{\binom{k}{2} - (k-1) + k(n-k)} \\ & \sim n^{-(k-1)} \frac{1}{c} (ce^{-c})^k \end{aligned}$$

to be one of the connected components of the random graphs. By Cayley’s (1889) formula, there are  $k^{k-2}$  ways to draw a tree on  $k$  given vertices, so,

since  $\binom{n}{k} \sim n^k/k!$  for fixed  $k$ , the expected number of trees is, approximately,

$$n \sum_{k=1}^{\infty} \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k$$

On the other hand trees are all that we need to count since even in the supercritical regime there is only one component (the giant one) which has a large complexity. Using (1.2) then finishes the proof. Note the amazing fact that we haven't used  $c < 1$  for this proof: in fact the above series simplifies into  $1 - c/2$  when  $c < 1$ , but not for  $c > 1$ !

### 1.3.3 Fluctuations

The study of fluctuations around the mean obtained in the last section is much more delicate and we outline our results with very short sketches of the main underlying ideas.

Let  $Z_c = \sum_{s \leq cn/2} \mathbf{1}_{\{F_s\}}$  be the number of fragmentations up to time  $cn/2$ .

**Theorem 7.** *Let  $c < 1$ . Then*

$$Z_c \rightarrow_d \text{Poisson}(\kappa(c)) \tag{1.5}$$

where  $\kappa(c) = (-\log(1 - c) - c)/2$ . In fact, the process  $(Z_c, c < 1)$  converges with respect to the Skorokhod topology on compact sets to a Poisson process whose compensator is  $(\kappa(c), c < 1)$ .

This result should really be viewed as a result on the fluctuations of  $D(cn/2)$  about its mean. We have chosen to formulate it in terms of fragmentations rather than the distance, since by (1.4),  $D_{cn/2} - cn/2 \approx N_{cn/2} - cn/2 = O(n^{1/2})$ . That is, in continuous time and in the subcritical regime the fluctuations of the distance are due to those of the Poisson process. However for the embedded discrete time chain, if  $k = \lfloor cn/2 \rfloor$ , then

$$(k - D_k)/2 \Rightarrow \text{Poisson}(\kappa(c)) \text{ as } n \rightarrow \infty \tag{1.6}$$

(We divide by 2 since a fragmentation reduces the distance by 1 instead of increasing it by 1). To deduce (1.6) from (1.5) we note that time  $k$  in the discrete walk corresponds to time  $N^{-1}(k) \approx cn/2$  in the continuous time walk.

Very roughly speaking, this  $O(1)$  bound on the number of fragmentations can be understood by remarking that Theorem 4 for random graphs implies that all cycles are very small in the subcritical regime. Therefore, the probability of breaking a cycle (and thus of increasing the value of  $Z_c$ ) is small, since we need to pick two integers in the same cycle. When cycles are of order 1, this happens with probability of order  $O(1/n)$ , so on the order of a time-scale  $\text{const} \cdot n$  this happens only a bounded number of time. To get the exact limit, the key idea is then to use exchangeability and a limit theorem for the size  $|\mathcal{C}_1|$  of a cycle containing a distinguished vertex, say 1. Since fragmentations are somehow negligible, it is not surprising by Proposition 1 that

$$|\mathcal{C}_1| \rightarrow_d Z$$

the total progeny of a  $PGW(c)$ .

When we enter the critical regime, the fluctuations start getting bigger.

**Theorem 8.** *If  $c = 1$ , then*

$$\left(\frac{6}{\log n}\right)^{1/2} (Z_c - \frac{1}{6} \log n) \rightarrow_d \mathcal{N}(0, 1) \quad (1.7)$$

*a standard gaussian random variable.*

Again, we emphasize that this result really describes the fluctuations of  $D(n/2)$  about its mean,  $n/2$ .

To prove this result, we proceed in two steps. It is well-known (see, for instance, Aldous' description of the critical random graph), that the correct scaling to observe the critical random graph is at times  $cn/2$  with  $c = 1 + \lambda n^{-1/3}$ ,  $\lambda \in \mathbb{R}$ . So, our first step will be to use subcritical estimates up to time  $cn/2$  with  $c = 1 - n^{-1/3}$ , and prove that for this value of  $c$ , the limit theorem (1.7) holds. To see this, define for  $0 \leq r \leq 1$

$$W_n(r) = \left(\frac{6}{\log n}\right)^{1/2} (Z_{1-n^{-r/3}} - \frac{r}{6} \log n)$$

Then  $(W_n(r); r \leq 1)$  is a martingale whose jumps may be neglected asymptotically. Moreover, it can be proved using subcritical estimates that the quadratic variation of  $W_n(r)$  converges to  $r$  because of the special form of the time-change  $c = 1 - n^{-r/3}$ . Therefore

$$(W_n(r), r \leq 1) \rightarrow_d (B_t, t \leq 1)$$

a standard Brownian motion (with respect to the Skorokhod topology). In particular, its terminal variable converges as announced to a standard gaussian random variable.

Our second step is to use random graph estimates to show that only a bounded number of fragmentations may happen between  $c = (1 - n^{-1/3})$  and  $c = 1$ . Indeed, by Aldous' theorem, the largest components have size  $O(n^{2/3})$  in the random graph, so the probability of making a fragmentation is  $O(\frac{n^{2/3}}{n} \frac{n^{2/3}}{n}) = O(n^{-2/3})$ . But  $c = 1 - n^{-1/3}$  to  $c = 1$  corresponds to a physical time-span of  $O(n^{2/3})$ , so the expected number of fragmentations in this regime has to be bounded.

Finally, we may state the result for the fluctuations in the supercritical regime,  $c > 1$ . This is the most complicated but also the most interesting case.

Recall that  $u(c) = 1 - \sum_{k=1}^{\infty} \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k$ .

**Theorem 9.** *If  $c > 1$ , then the following convergence holds*

$$\frac{D(cn/2) - nu(c)}{n^{1/2}} \rightarrow_d \mathcal{N}(0, \sigma^2)$$

for some explicit  $\sigma = \rho[1 + \rho(c/2 - 1)]$ , and  $\rho = 1 - \theta(c)$  is the extinction probability of a supercritical PGW( $c$ ).

The core of the argument consists in proving that approximating the number of cycles in  $\sigma_{cn/2}$  by the number of clusters in  $G(cn/2)$  is accurate within  $o(n^{1/2})$ . Indeed, if this is true, we can use a result by Pittel (1990) which proves the central limit theorem with the indicated variance for the number of clusters in the random graph.

We have already indicated that it is easy to show that the random graph approximation is accurate within  $o(n)$ . In fact this simple argument also yields a result up to  $O(n^{1/2}a_n)$  for any sequence  $a_n \rightarrow \infty$ . But proving that this holds up to  $o(n^{1/2})$  is a much more complex task, and we need to study the structure of coalescence and fragmentation in much more detail, and exploit fully the fact that fragments get reabsorbed by larger clusters fast enough.

## 1.4 The (hyperbolic) geometric picture

The phase transition described in the previous section is very surprising if we think geometrically about it. Indeed, we have a graph  $G_n$ , *transitive*, (i.e., the group of automorphisms of  $G_n$  acts transitively on  $G_n$ , or, in words, the graph looks the same from every vertex), with the following property. If we do simple random walk on this graph, the *speed* of the random walk (distance divided by time) is 1 until a certain critical time  $n/2$ , at which point, it starts decelerating in a non-analytic way. Moreover there is only one point at which we observe non-analyticity. It is the goal of this section to investigate what are the geometric properties of  $G_n$  that may be lying behind such a behavior.

Relating geometric properties of graphs to the speed of random walk on these graphs is not a new subject. For instance, the celebrated Carne-Varopoulos bound (Theorem 12.1 in Lyons and Peres (2005)) is often used to prove subdiffusive behavior of a random walk on a graph:

$$p_n(u, v) \leq 2 \sqrt{\frac{\pi_v}{\pi_u}} \|P\|_\pi^n e^{-\frac{d(u,v)^2}{2n}} \quad (1.8)$$

where  $p_n$  are the  $n$ -step transition probabilities,  $\pi_v$  denotes a stationary measure of the walk, and  $\|P\|_\pi$  denotes the norm of the operator  $P$  in  $\ell^2(\pi)$ . This works especially well when the growth of the graph is subexponential.

For random walks where we expect the speed to be linear, existing results in the literature relate the speed of the random walk to the *growth* of the graph and its *branching number*. Briefly, the lower growth rate  $\underline{\mathbf{gr}}$  of a graph  $T$  is defined as  $\liminf_{n \rightarrow \infty} |T_n|^{1/n}$  where  $T_n$  is the set of vertices at distance  $n$  from a fixed vertex (the "root"),  $o$ . (This definition does not depend on the choice of  $o$ ). An infinite graph  $T$  such that  $\underline{\mathbf{gr}} > 1$  is said to have exponential growth. The branching number is defined as the supremum of those  $\lambda$  such that there is a positive flow from  $o$  to  $\infty$  when the amount that can flow through an edge  $e$  is at most  $\lambda^{-|e|}$ , where  $|e|$  denotes the distance between  $o$  and  $e$  (see Lyons and Peres (2005) for intuition, background and much more about these quantities). In general,  $\mathbf{br} \leq \underline{\mathbf{gr}}$ . Peres (see Lyons and Peres (2005), Proposition 12.5) has proved that if  $T$  is a connected locally finite graph and  $\liminf_{n \rightarrow \infty} \frac{d(X_n)}{n} \geq s$  with positive probability, then  $\mathbf{br} \geq e^{s/2} > 1$ , which also implies  $\underline{\mathbf{gr}} > 1$ . Conversely, if in addition the graph is *nonamenable*, then having an exponential growth rate ( $\underline{\mathbf{gr}} > 1$ ) is sufficient

to have positive speed (see Proposition 5.6 and Theorem 5.7 in Lyons and Peres (2005)). See also the work of Virág (2000) for related questions.

Here the geometric phenomenon that seems to be causing the phase transition of the random walk is quite different and slightly more subtle. It is linked with the change of a constant related to Gromov hyperbolicity, that approximately describes *how hyperbolic does the space look like* from the point of view of the random walk.

In order to describe our results it is useful to recall some basic definitions from the theory of Gromov hyperbolic graphs.

### 1.4.1 Gromov hyperbolicity

The notion of hyperbolicity for a discrete structure such as a group is a notion that goes back essentially to Gromov (1987). As there is no derivative, and thus no curvature available to work with in a discrete space, the idea is to define what hyperbolic means using only elementary properties of the space.

One way to do is as follows. Let  $(X, |\cdot|)$  be a metric space, where  $|x - y|$  denotes the distance between  $x$  and  $y$ .

**Definition 3.** A triangle  $(x, y, z)$  with geodesic sides  $s_1, s_2, s_3$  is said to be  $\delta$ -thin if any side, say  $s_1$ , lies entirely within distance at most  $\delta$  of the two remaining sides (see Figure 1.4.1).

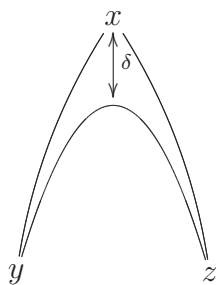


Figure 1.4: In a hyperbolic space,  $[y, z]$  is  $\delta$ -close to  $[x, y] \cup [y, z]$

**Definition 4.** The space is called  $\delta$ -hyperbolic if all geodesic triangles are  $\delta$ -thin, and it is simply called hyperbolic if it is  $\delta$ -hyperbolic for some  $\delta \geq 0$ .

When  $\delta = 0$ , it can be seen that the space isometrically embeds into a tree.

Of course a bounded space (in particular, a finite space such as  $\mathcal{S}_n$ ) is trivially hyperbolic, but we will be interested in situations where the constant  $\delta$  may or may not stay bounded as the size of the space tends to  $\infty$ .

## 1.4.2 Results

We need some notations before we state our results. For  $0 < a < 1$ , let  $\partial B(an)$  be the sphere of radius  $an$ , i.e. the set of points at distance  $\lfloor an \rfloor$  from the origin. We let  $\nu$  be the hitting distribution of  $\partial B(an)$  by  $\sigma_t$ , i.e.  $\nu$  is the law on  $\partial B(an)$  of  $\sigma_T$  where  $T = \inf\{t > 0, d(\sigma_t) = \lfloor an \rfloor\}$ . Asking that  $0 < a < 1$  is of course the natural choice since by (1.2), the points furthest away from the identity are permutations with only one cycle containing all  $n$  elements, in which case the distance is  $n - 1$ .

**Theorem 10.** *Let  $x, y$  be sampled from  $\nu$  independently, and set  $p = I$ , the identity element.*

1. *If  $a < 1/4$ , then with probability asymptotically 1, there is a geodesic between  $x$  and  $y$  that comes within expected distance  $\delta$  of  $p$  for some  $\delta < \infty$ .*
2. *If  $a > 1/4$ , then no geodesic between  $x$  and  $y$  can approach  $p$  closer than  $\delta n$  with probability asymptotically 1, for some  $0 < \delta < \infty$ .*

To rephrase less rigorously, this result tells us that when  $a < 1/4$ , Definition 4 is satisfied "asymptotically  $\nu$ -almost surely", but not when  $a > 1/4$ .

At this point we should emphasize that the result in Theorem 10 involves hyperbolic constants that are different from the standard definitions discussed above in several important ways. The most obvious difference comes from the randomness of  $x$  and  $y$ , and from the fact that the role played by  $x, y$  and  $p$  are somewhat different. Here  $p$  is a fixed reference point, whereas Gromov's definition requires that every triangle should be thin. Another issue is that, corresponding to the second definition of hyperbolicity with thin triangles, we show that there exists a certain geodesic between  $x$  and  $y$  having the desired properties. But in the corresponding paper I have also

given examples of points connected by a great many geodesics, which, more importantly, may be far apart.

Despite these differences, this result says something about the geometry of the graph *as perceived by the walk*, since the tips of the triangle are chosen according to the random walk hitting measure  $\nu$ .

It is natural to ask if this choice is representative of the whole space, but unfortunately the next result tells us that it is not the case.

**Theorem 11.** *Let  $\mu$  be the uniform measure on  $\partial B(an)$ . Then  $\mu$  and  $\nu$  are asymptotically singular, i.e.  $d_{TV}(\mu, \nu) \rightarrow 0$ . Moreover there exists a sequence of sets  $S = S_{a,n} \subset \partial B$  such that  $\nu(S) \rightarrow 1$  ( $S$  is almost the support of  $\nu$ ) but*

$$\frac{|S_{a,n}|}{|\partial B(an)|} \leq C \exp(-\gamma n)$$

for some  $\gamma > 0$ .

Similar results exist for random walks on supercritical Galton-Watson trees, for which it is known (see Lyons and Peres (2005), chapter 15) that the harmonic measure at infinity is singular with respect to the branching measure, and they have different Hausdorff dimensions.

An essential part of the above result consists in giving a good asymptotic description of the uniform measure  $\mu$  on  $\partial B(an)$ . One of these results is the following.

**Theorem 12.** *Let  $|\mathcal{C}_1|$  be the length of the cycle that contains 1. Under  $\mu$ , the uniform distribution on  $\partial B(an)$ ,*

$$|\mathcal{C}_1| \Rightarrow \mathbf{G}$$

where  $\mathbf{G}$  is a geometric r.v. with  $P(\mathbf{G} > k) = (b/(1+b))^k$  and  $b$  satisfies  $\log(1+b)/b = 1-a$ .

The proof of this result uses a description of the uniform measure known as the Feller coupling (a variant of the Pitman-Dubins *Chinese restaurant process*, see Pitman (2002)). Conditioning this permutation to have distance  $an$ , i.e. to have  $(1-a)n$  cycles, we get a description of  $|\mathcal{C}_1|$  by a mild generalization of the *Gibbs conditioning principle* well-known to statistical physicists and to probabilists familiar with large deviation theory.

Since the "support" of the random walk is so small compared to the size of the space, it is natural to ask whether Theorem 10 can be extended to the case where the tips of the triangle are chosen according to the uniform distribution  $\mu$  rather than the hitting distribution. Interestingly, the next result shows that the qualitative behavior remains the same under the uniform distribution, but with a different critical radius.

**Theorem 13.** *Let  $x, y$  be sampled from  $\mu$  independently, and set  $p = I$ , the identity element.*

1. *If  $a < 1 - \log 2$ , then with probability asymptotically 1, there is a geodesic between  $x$  and  $y$  that comes within expected distance  $\delta$  of  $p$  for some  $\delta < \infty$ .*
2. *If  $a > 1 - \log 2$ , then no geodesic between  $x$  and  $y$  can approach  $p$  closer than  $\delta n$  with probability asymptotically 1, where  $0 < \delta < \infty$*

This result is another example of the use of the connection developed throughout this thesis between phase transition of walks and existence of giant clusters in related random graphs, although this time the graph process is more complicated than the simple  $G(n, p)$  model of Erdős and Renyi.

## 1.5 Limiting behavior for the distance of random walks

In my third paper (joint with Rick Durrett), we ask about the existence of phase transitions similar to that of Theorem 1 and 6 for other random walks. More generally, let  $G_n$  be a family of graphs whose size tend to infinity, and consider  $X^n$  the simple random walk on this graph. What can be said of the limiting behavior of of the distance from a distinguished point,  $\text{dist}(X_{t\theta_n}^n)$ , as a function of  $t$ , and for various scaling parameters  $\theta_n$ ? As mentioned, we were particularly interested in the existence of a phase transition for a suitable scaling.

### 1.5.1 Random $p$ -cycles

A first walk that we consider is a simple generalization of the result for random transpositions. Suppose that the step distribution is now uniform

on cycles of lengths  $p$ , where  $p \geq 3$  (so  $p = 2$  is exactly the case of random transpositions). To avoid complications we explain our results for  $p = 3$ . As Theorem 14 shows, this random walk displays a phase transition very similar to that of random transpositions.

For  $c > 0$ , let

$$u(c) = 1 - \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

For  $t > 0$  denote by  $D_t$  the distance to the identity of  $\sigma_t$ , where  $\sigma_t$  denotes the random walk which is a product of a Poisson number of 3-cycles.

**Theorem 14.** *Let  $\omega(n)$  be any sequence with  $\omega(n) \rightarrow \infty$  as slowly as we want. Then*

$$\frac{D_{cn} - u(c)n}{\omega(n)\sqrt{n}} \rightarrow 0$$

*in probability.*

Of course the function has similar characteristics as in the case of random transpositions for  $c < 1/6$ ,  $u(c) = c$ , it has no-second derivative at  $c = 1/6$  and  $u(c) < c$  for  $c > 1/6$ . The value  $c = 1/6$  for the phase transition is perhaps surprising. Indeed the naive reasoning that we are about to present suggests the value  $c = 1/4$  as a first guess. Indeed, any 3-cycle  $(a b c)$  can be decomposed as the product of 2 transpositions  $(a b c) = (a b)(b c)$ , so it seems at first that this random walk goes "twice as fast" as the random transposition, since there are "two steps at once". Of course, in this decomposition the two transpositions share a common point  $(b)$ , but intuitively this should actually slow down the walk - whereas in fact we observe the phase transition *before*  $c = 1/4$ , at  $c = 1/6$ .

Our proof sheds light on this rather surprising fact. We use random hypergraphs as a substitute for Erdős-Renyi random graphs. Random hypergraphs are graphs where edges ("hyperedges") may connect several vertices at the same time. The natural thing to do is then to couple the random walk with a hypergraph process, adding a hyperedge connecting  $a, b$  and  $c$  whenever the 3-cycle  $(a b c)$  is applied. The result of doing this is, morally, adding the edges  $(a, b)$ ,  $(b, c)$  *but also the edge  $(c, a)$* : therefore this walk goes 3 times as fast as random transpositions, and this is where the value  $c = 1/6$  comes in. More generally, for a random  $p$ -cycle, the phase transition is located at time  $cn$  with  $c = 1/p(p-1)$ . Again, loosely speaking, this walk has been sped up by a factor of  $\binom{p}{2}$  compared to random transpositions.

## 1.5.2 Random adjacent transpositions

As we will see, the random adjacent transposition provides a first example of a walk which has a very different behavior, with different formulae describing the asymptotic behavior of the distance in different regimes, but no sharp phase transition. The walk  $(X_t, t \geq 0)$  is once again a  $\mathcal{S}_n$ -valued Markov process where the step distribution of the discrete-time walk is uniform on transpositions  $(i, i+1)$  for  $1 \leq i \leq n-1$ . To evaluate the (adjacent) distance of a permutation  $\sigma$ , i.e., the minimum number of adjacent transpositions needed to build  $\sigma$ , or the graph distance between  $I$  and  $\sigma$  on the Cayley graph of  $\mathcal{S}_n$  generated by the set of adjacent transpositions, there is a very convenient formula

$$d_{\text{adj}}(\sigma) = \text{Inv}(\sigma) := \#\{1 \leq i < j \leq n : \sigma(i) > \sigma(j)\} \quad (1.9)$$

$\text{Inv}(\sigma)$  is called the number of inversions of  $\sigma$ . It is hard to trace the origins of this formula. Let us mention Diaconis and Graham (1977), which includes earlier references to Kendall (1970) and Knuth (1973), section 5.1.1.

Having in mind roughly the same applications to computational biology as those exposed earlier in the introduction, Eriksson et al. (2000) considered the problem of evaluating the distance after  $k$  random adjacent transpositions were applied. Relying heavily on formula (1.9) they were able to carry some explicit combinatorial analysis, to obtain various exact formulae for this expected distance, such as this one:

$$Ed_{\text{adj}}(X_{t(k)}) = \sum_{r=0}^k \frac{(-1)^r}{n^r} \left[ \binom{k}{r+1} 2^r C_r + 4d_r \binom{k}{r} \right] \quad (1.10)$$

where  $t(k) = \inf\{s > 0 : N_s = k\}$  is the time at which the  $k^{\text{th}}$  random adjacent transposition is applied,  $C_r$  are the Catalan numbers (i.e.,  $C_r = \frac{1}{r+1} \binom{2r}{r}$ ) and  $d_r$  is some non-negative integer sequence. An explicit form for  $d_r$  is obtained by Eriksen (2005) (who also proves that the formula holds for a wider range of  $k$ ). Proofs are based on the observation that the matrix with entries  $p_{i,j}(t) = P(X_t(i) > X_t(j))$  performs a discrete heat-flow process, a fact that will become obvious from our approach.

However getting useful asymptotics from exact formulae such as (1.10) seems an almost untractable task. On the other hand, we give a much more probabilistic approach, which has the advantage of yielding intuitive results

about the asymptotic behavior of this walk. Also, in certain regimes our results go beyond the simple expected value analysis.

Let us explain very briefly the starting point of our analysis. By (1.9), we get

$$d_{\text{adj}}(X_t) = \sum_{i < j} \mathbf{1}_{\{X_t(i) > X_t(j)\}}$$

On the other hand, if  $1 \leq i \leq n$  is fixed, the trajectory  $(X_t(i), t \geq 0)$  of the  $i^{\text{th}}$  particle under the action  $X_t$  is nothing but a continuous time simple random walk, on  $\{1, \dots, n\}$  with reflecting boundaries, starting at  $i$ , and with rate  $2/(n-1)$  except at the boundaries where the rate is  $1/(n-1)$ .

Taking two such trajectories, say those of particles  $i$  and  $j$  with  $i < j$ , they now both perform independent simple random walk, except when they are adjacent, because it is never the case that  $X_t(i) = X_t(j)$ . When they are adjacent, the only thing that can happen is an exchange of the two particles, or one of them moves away from the other. In other words, this pair of particles performs simple exclusion on  $\{1, \dots, n\}$  (with reflecting boundaries and modified rates near the boundaries).

Taking expected values for instance, this simply becomes

$$E d_{\text{adj}}(X_t) = \sum_{i < j} P^{i,j}(Y_t > Y'_t)$$

where the law of  $(Y_t, Y'_t)$  under  $P^{i,j}$  is just simple exclusion of two particles started at  $i$  and  $j$ .

From this representation we can deduce the following result.

**Theorem 15.** *Let  $t > 0$ . Then*

$$\frac{1}{n} E d_{\text{adj}}(X_{nt}) \rightarrow f(t)$$

*as  $n \rightarrow \infty$  for an explicit function  $f(t)$ .  $f(t)$  is infinitely differentiable, and moreover it has the asymptotic behavior*

$$\lim_{t \rightarrow \infty} \frac{f(t)}{\sqrt{t}} = \frac{1}{2} E[\max_{0 \leq s \leq 1} B_{2s}] = \sqrt{2}/2$$

*where  $B_t$  is a standard Brownian motion.*

This should be regarded as a "diffusive behavior" type of result. The next result looks at the distance of the random walk on time-scales of the order  $n^3t$ . On this time-scale, any two particles may have had a chance to be exchanged. Indeed, it takes about  $n^2$  unit of time for particles  $n$  units of space apart to meet, and recall that a given pair of particles evolves with rate  $O(1/n)$ , which accounts for the extra factor  $n$  in the scaling.

**Theorem 16.** *Let  $t > 0$ .*

$$\frac{1}{n^2}d_{\text{adj}}(X_{n^3t}) \rightarrow_p \mathbf{P}[B_1(t) > B_2(t)]$$

where  $B_1$  and  $B_2$  are two reflecting Brownian motions started uniformly on  $0 \leq B_1(0) < B_2(0) \leq 1$  evolving independently.

The last two results contrast sharply with the behavior of random transpositions. Although there are different regimes, we never observe a non-analyticity. Moreover the distance is a strictly sublinear function of time in both regimes. Figures 1.8 and 1.9, at the end of this introduction, represent the two Cayley graphs (respectively, for all transpositions and for the adjacent transpositions)<sup>1</sup>. Notice how the two graphs are different, and how in particular the one for the adjacent looks similar to a lattice.

### 1.5.3 Random walk on a random 3-regular graph

Here is an example of a random walk which has a phase transition in a context much different from the case considered so far of random walks on  $\mathcal{S}_n$ .

A *3-regular* graph is a graph where all vertices have degree equal to 3. If  $n$  is even, it is always possible to draw such a graph on  $n$  vertices, and a *random 3-regular graph* is a uniform member of such graphs. Bollobàs and de la Vega (1982) have given a representation of a random 3-regular graph in terms of random matchings, which we explain briefly. Expand each vertex  $i$  into 3 vertices  $3i$ ,  $3i + 1$  and  $3i + 2$ . If  $n$  is even, we may consider a random matching of those  $3n$  vertices. A random 3-regular graph  $G_n$  is then obtained by collapsing back the groups of 3 vertices together and keeping the edges of the random matching. (In fact this only holds on the event that there are no self-loops or multiple edges in  $G_n$  but this can be shown to have positive

---

<sup>1</sup>These figures were made by Drew Armstrong (Cornell University).

probability asymptotically). Alternatively, a random 3-regular graph is also a special case of the Molloy and Reed (1995) graph model where vertices have a prescribed degree distribution.

We make use of these two (equivalent) representations of a random 3-regular graph. The result that we prove for the random walk  $(X_k, k \geq 0)$  on this graph is the following, assuming that  $X_0 = 1$ .

**Theorem 17.** *For fixed  $t > 0$*

$$\frac{d(X_{\lfloor t \log_2 n \rfloor})}{\log_2 n} \xrightarrow{p} f(t) := \min\left(\frac{1}{3}t, 1\right)$$

An intuitive description of a random 3-regular graph as seen from vertex 1 can be given as follows. We grow the cluster containing 1 by exposing successive vertices with increasing distance away from 1. As long as we have not exposed more than  $o(n)$  vertices this should look very much like a regular tree (i.e., each vertex has 2 edges going away from the root and 1 leading back towards the root). Therefore this phase roughly lasts until distance  $\log_2 n$ . After distance  $\log_2 n$  however, this tree approximation is no longer valid and there are many edges inter connecting the vertices.

Therefore, as long as we are below critical distance  $\log_2 n$ , we expect that  $d(X_k)$  evolves like a biased random walk on  $\mathbf{Z}_+$ , with transition probabilities  $p(x, x+1) = 2/3$  and  $p(x, x-1) = 1/3$ . After  $k$  moves we expect this walk to be at distance  $k(2/3 - 1/3) = k/3$ .

On the other hand, once the walk reaches a distance corresponding to the average diameter of the graph ( $\log_2 n$  by Bollobàs - de la Vega (1982), or Theorem 2.13 in Wormwald (2005)) it should remain at this level. Indeed it cannot go any further since this is the diameter, and on the other hand the tree structure below makes it hard for it to come down back toward the root.

As we already mentioned, this gives a lower bound on the time to reach equilibrium for a random walk on a random 3-regular graph. Indeed, the walk cannot have reached equilibrium before it has reached a distance corresponding to the average diameter of the graph. In fact, we conjecture that this lower bound gives the right answer.

**Conjecture 1.** *This lower bound is sharp: random walk on a random 3-regular graph mixes in time  $3 \log_2 n$ .*

Right now it is known that  $O(\log n)$  is the correct order of magnitude for the mixing time. More precisely, if we let  $\Delta(t) = \max_{i,j} \frac{|p_{i,j}^t - \pi_j|}{\pi_j}$  be the relative pairwise distance between the chain and the invariant (here uniform) measure  $\pi$ , then  $\Delta(b \log n) \rightarrow 0$  for  $b$  large enough. See Durrett (2005) for a proof. The conjecture says that we can take  $b = (3 + \varepsilon)$  if we want.

### 1.5.4 Other random walks and some open problems

We hope to initiate a programme of identifying the random walks for which the distance has a phase transition and those for which it remains smooth. More generally, we view the analysis carried on the examples above as the start of a more complete investigation of what are the possible behaviors for the distance of a random walk.

Some more work is already going in this direction. For instance, Fulman (2004) studies the evolution of the distance when the random walk is the so-called Gilbert-Shannon-Reeds riffle shuffle. For this shuffling method (which, we emphasize, is *nonreversible*), Bayer and Diaconis (1992) proved an explicit formula for the distribution of  $\sigma_t$  after  $r$  shuffles. In particular from this formula it follows that the riffle-shuffle distance of a permutation is given by

$$d(\sigma) = \lceil \log_2(\text{Des}(\sigma) + 1) \rceil \text{ where } \text{Des}(\sigma) = \#\{1 \leq i \leq n-1 : \sigma(i) > \sigma(i+1)\}$$

$\text{Des}(\sigma)$  is called the number of descents of  $\sigma$ . The main result of Fulman (2004) is

**Theorem 18.** Fulman (2004). *After  $r = \log_2(\alpha n)$  shuffles*

$$\frac{1}{n} E(\text{Des}(\sigma_r)) \rightarrow \alpha - \frac{1}{e^{1/\alpha} - 1}$$

*at least if  $\alpha > 1/(2\pi)$ .*

This says that for this range of  $r$  the walk is already in a sublinear regime. In particular, as  $\alpha \rightarrow \infty$  we get that  $d(\sigma_r) \sim \log_2 n$  since the expression  $\alpha - \frac{1}{e^{1/\alpha} - 1}$  is asymptotic to  $1/2$ . This is not surprising since this is the diameter of the graph. It is not clear at this point whether this formula also holds for smaller values of  $\alpha$ , although it is tempting to let  $\alpha \rightarrow 0$  and get that for small values of  $\alpha$  the walk is "almost" linear (the fraction term with the exponential is much smaller than the other term).

Also, the techniques developed for the random walk on a 3-regular graph should be very useful when dealing with graph which have a strong "tree"-like geometry. For instance, the giant cluster of a supercritical Erdős-Renyi random graph should be another example where the random walk exhibits a phase transition. Essentially, the arguments will be the same, except that there is no available good upper bound in the supercritical regime of the walk, where the random walk should stay at a distance corresponding to the average distance in the graph ( $\log n / \log c$ ). In the case of a random 3-regular graph this was easy since the average distance is the same as the topological diameter, which is well-known (this is a 20-year old result of Bollobàs and de la Vega). But in the supercritical random graph, the topological diameter is larger than the average distance. These questions are rather delicate, see Chung and Lu (2001) for partial results about the asymptotics about the topological diameter. Indeed it is not even known if the diameter of the graph is realized on the giant component or on some "small" cluster!

## 1.6 On a result of Oded Schramm

The introduction to this thesis would be incomplete without a discussion of Oded Schramm's (2004) result about the Poisson-Dirichlet structure of the random transposition random walk. To make this discussion accessible to the reader, we will recall basic definitions of the Poisson-Dirichlet distribution, and then describe Schramm's result and its meaning.

### 1.6.1 The Poisson-Dirichlet distribution

There are many ways to think about this distribution which occurs naturally in an amazing variety of different contexts. Perhaps the simplest possible definition is the so-called *stick-breaking* construction of a *PD* random variable. Let  $\mathcal{S}^\downarrow$  be the space of nonincreasing sequences  $(x_i)$  with  $x_i \geq 0$  and  $\sum_i x_i \leq 1$  (i.e.,  $\mathcal{S}^\downarrow$  is the infinite-dimensional simplex).  $\mathcal{S}^\downarrow$  can be endowed with a distance that makes it a complete metric space.

**Definition 5.** Let  $(U_i, i \geq 1)$  be a collection of i.i.d. uniform  $(0, 1)$  random variables. Define  $W_1 = U_1$  and for  $i \geq 1$   $W_{i+1} = \bar{U}_1 \bar{U}_2 \dots U_i$  where  $\bar{U}_i = 1 - U_i$ . The Poisson-Dirichlet distribution  $\mu$  is defined as the probability distribution on  $\mathcal{S}^\downarrow$  given by the law of the decreasing rearrangement of the variables  $(W_i)_{i \geq 1}$

In other words, we start with a stick of length 1. We break it at a uniformly chosen point. One of those pieces is then chosen and broken again uniformly. At each stage, we keep repeating this procedure by selecting one of the two pieces resulting from the breaking of the last stage, and break it again uniformly. The Poisson-Dirichlet distribution is the distribution of the masses of the pieces of the initial stick, after decreasing rearrangement.

It is also possible to define it as the rescaled points of a Poisson point process. For this, let  $X_i$  be the points (in decreasing order) of a Poisson point process on  $(0, \infty)$  with intensity  $x^{-1}e^{-x}dx$ . Let  $S = \sum_{i \geq 1} X_i$ , then  $S < \infty$  a.s., and the points  $Y_i = (X_i/S)$  have the Poisson-Dirichlet distribution. (Moreover,  $S$  has the standard exponential distribution, and is independent of  $(Y_i)_{i \geq 1}$ .)

For us, it will be important to see its connection with random uniform permutations. If  $\sigma$  is uniformly distributed, then it can be seen that the ranked lengths of the cycles of  $\sigma$ , normalized by  $n$ , converge weakly (for the topology induced by the metric on  $\mathcal{S}^\downarrow$ ) to  $\mu$ , the Poisson-Dirichlet distribution. This is easily seen because the cycles of  $\sigma$  possess the uniform stick-breaking property even at the discrete level (a fact which is obvious with the Pitman-Dubins Chinese Restaurant process).

Other natural occurrences of this distribution (and related distributions) include for instance the study of the ranked lengths of the excursions of a one-dimensional Brownian motion, probabilistic number theory (see Donnelly and Grimmett (1993)), processes of coalescence and fragmentation (see below), or, more recently, in the study of global properties of typical compact Riemann surfaces (Gamburd (2005))<sup>2</sup>.

## 1.6.2 A process of coalescence and fragmentation

Consider the following Markov chain  $X$  with values in the space  $\mathcal{S}^\downarrow$ , whose transition probabilities can be described as follows. If the state of the chain is  $s \in \mathcal{S}^\downarrow$ , we let  $I$  and  $J$  be two independent random variables in  $\mathbf{N}$  obtained by size-biased picking from  $s$ , i.e.  $P(I = i) = s_i$ , so  $I$  has a natural tendency to pick large components of  $s$  rather than small components. The transformation we apply to the chain depends on whether  $I = J$  or not. If  $I \neq J$  we merge the components  $s_I$  and  $s_J$ : we replace the two entries  $s_I$  and  $s_J$  by a

---

<sup>2</sup>Thanks to Ofer Zeitouni for bringing this paper to my attention

single entry of size  $s_I + s_J$ , and then rerank. When  $I = J$  on the other hand, we split the component uniformly at random and rerank.

It is not hard to see that this process mimics the dynamics of the cycle lengths of the random transposition random walk, but in a continuous setting. In particular, since the random transposition random walk converges to the uniform distribution on  $\mathcal{S}_n$ , the cycle distribution after a "long time" should be that of a uniform permutation. On the other hand we have already argued that this is asymptotically a Poisson-Dirichlet distribution, after normalization by  $n$ . Writing carefully this argument, this shows by weak convergence that

Poisson-Dirichlet is invariant with respect to the split-merge process  $X$

This observation led Vershik to conjecture that Poisson-Dirichlet was the unique invariant measure for the process  $X$ . Around 2000 and 2001, several approaches were proposed to try to prove this conjecture (see Pitman (2002), Mayer-Wolf et al. (2002)), including the weak convergence sketched above, and the conjecture was finally proved in 2002 by Diaconis et al. (2004).

**Theorem 19.** Diaconis, Mayer-Wolf, Zeitouni, Zerner (2004). *The Poisson-Dirichlet distribution is the unique invariant measure for the process  $X$ .*

This difficult result was proved using the representation theory of the symmetric group, by looking at precise estimates for the decoupling time between the discrete and the continuous process when the starting distribution puts enough mass on reasonably large clusters, and then proving *a priori* that any invariant measure necessarily satisfies this requirement.

### 1.6.3 Schramm's result

Consider the random transposition random walk on the symmetric group and recall the coupling of Theorem 5 with an Erdős-Renyi random graph.

We have insisted many times during this introduction on the existence of a giant component in the supercritical regime of the random graph, i.e. at times  $cn/2$  when  $c > 1$ . But what about the cycles of the permutation? Do there exist cycles of order  $n$  at times  $cn/2$  in the permutation, when  $c > 1$ ? Clearly, since the random graph gives an upper-bound on the sizes of the cycles of the random walk, when  $c < 1$  it is impossible for the random walk

to cycles larger than  $O(\log n)$ . But when  $c > 1$ , it is plausible that such cycles exist despite fragmentation.

The first evidence in this direction was given by Theorem 5 from Berestycki and Durrett (2004). Indeed it is proved there that the total mass of cycles which are reasonably large (larger than  $n^{2/3-\varepsilon}$  for some  $\varepsilon > 0$ ) equals the size of the giant component of the random graph, up to  $o(n)$ . Of course this is not enough to conclude to the existence of cycles of order  $O(n)$ .

The existence of such cycles, and their precise structure, is the content of the following striking result of Oded Schramm. Let  $\sigma_t$  be the random transposition random walk on  $\mathcal{S}_n$  and let  $G(t)$  be the random graph process of Theorem 5. For  $c > 1$  we let  $\Gamma(t)$  be the largest component of  $G(t)$ , which is the giant component when  $t = cn/2$  and  $c > 1$ .

**Theorem 20.** Schramm (2004). *Let  $c > 1$ . Consider  $X^n$  the random vector of  $\mathcal{S}^\downarrow$  which consists of the non-increasing rearrangement of the size  $|C|$  of those cycles of  $\sigma_{cn/2}$  which belong to  $\Gamma(cn/2)$ , normalized by  $|\Gamma(cn/2)| \sim \theta(c)n$ , and appended with 0's to make it an infinite vector:*

$$X^n = \text{rank} \left( \frac{|C|}{\theta(c)n}, C \text{ cycle of } \sigma_{cn/2} \text{ and } C \subset \Gamma(cn/2) \right)$$

Then

$$X^n \rightarrow_d \text{Poisson-Dirichlet}$$

*In particular at time  $cn/2$  the largest cycle of  $\sigma_{cn/2}$  is asymptotic to  $Wn$ , where  $W$  is the first coordinate of a Poisson-Dirichlet random variable, so  $W > 0$  a.s.*

Along the way, another proof of Theorem 19 is given. Remarkably, this proof uses only probabilistic arguments, and, in particular, does not appeal to the representation of the symmetric group.

This result was first conjectured by David Aldous, although it first appeared in Berestycki and Durrett (2004). Aldous was indeed motivated by our coupling result (Theorem 5) as well as his own result (1997) about the multiplicative coalescent and the critical random graph.

**Interpretation.** An intuitive interpretation of this result is the following. Consider the pieces of the giant component in the random walk, i.e. those cycles which belong to the giant component in the coupling. As time evolves,

two things happen to those cycles. They first experience coagulation and fragmentation according to the discrete analogue of the split-merge process of Theorem 19, and, second, the system receives mass continuously due to coalescence with small clusters outside of the giant component. Having in mind the content of Theorem 19 which says that Poisson-Dirichlet is the equilibrium distribution of the split-merge transformation, Theorem 20 says that as the mass of the system increases *it is instantly in equilibrium with itself*, until eventually the giant component encompasses all vertices, which has to be when the random walk has reached uniformity. On the other hand, by the connectivity result for random graphs of Theorem 3, this happens at time  $n \log n/2$ . Although we have not tried to make the above argument rigorous, we believe that it definitely provides some wonderful probabilistic insight about Diaconis and Shahshahani's result (Theorem 2).

#### 1.6.4 More open problems

It is worth noting that Theorem 20 is the mean-field analogue of the following question. Given some infinite graph  $G = (V, E)$ , attach to each edge a Poisson clock. We think of each vertex as the location of some particle. When an edge  $e$  rings, the two particles at both ends are exchanged. This defines a random permutation  $\sigma_t$  of the vertices of  $V$  for all  $t \geq 0$ . So, when  $G$  is the complete graph  $K_n$  on  $n$  vertices, this is exactly the random transposition random walk.

It is natural to ask whether there exist some infinite cycle for some  $t > 0$ . (In fact, it has been shown by Sütő (2002) that the occurrence of this event with positive probability is equivalent to Bose-Einstein condensation in a bosonic gas, a fact of significant importance to physicists since Bose-Einstein condensation is generally believed to be one of the mechanisms leading to superconductivity and superfluidity).

However on the mathematical level there are very few results available right now, if we set aside Theorem 20 of Schramm. Angel (2003) considers the case of a  $d$ -regular tree for  $d \geq 5$ . By using a representation of this process due to Tóth (1993) in terms of a certain self-interacting random walk he was able to prove that for  $t$  in a finite interval  $(a, b)$  (dimension-dependent) then *a.s.* there are some infinite cycles. It is conjectured that for  $G = \mathbf{Z}^d$  infinite cycles occur only when  $d \geq 3$ , i.e. those dimensions for which simple random walk is transient.

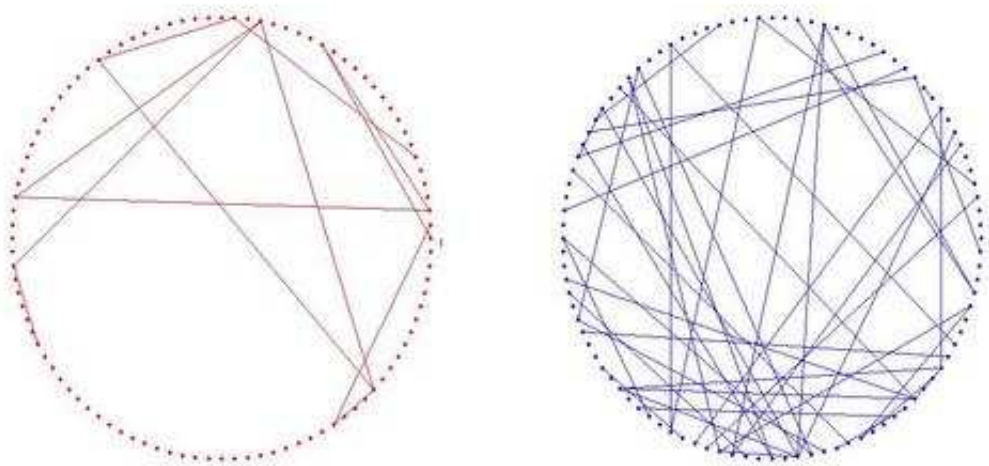


Figure 1.5: Subcritical random graph with  $c = 0.95$  for  $n = 100$  vertices. On the left, the cluster  $\mathcal{C}_1$  containing 1 has 13 vertices. On the right, the complement of  $\mathcal{C}_1$  contains many vertices, but all clusters are small.

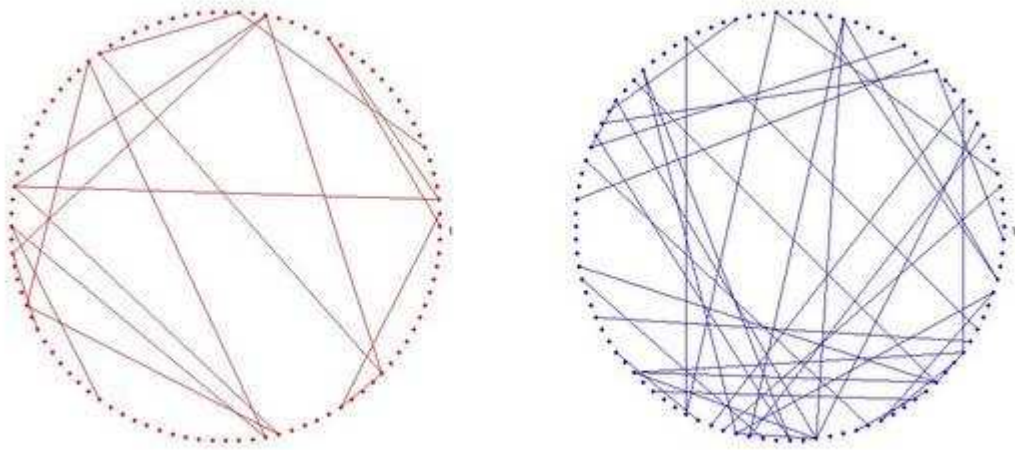


Figure 1.6: Critical random graph ( $c = 1$ ) for  $n = 100$  vertices. On the left, the cluster  $\mathcal{C}_1$  containing 1 has now 19 vertices. On the right, the complement of  $\mathcal{C}_1$  is made up of comparable clusters.

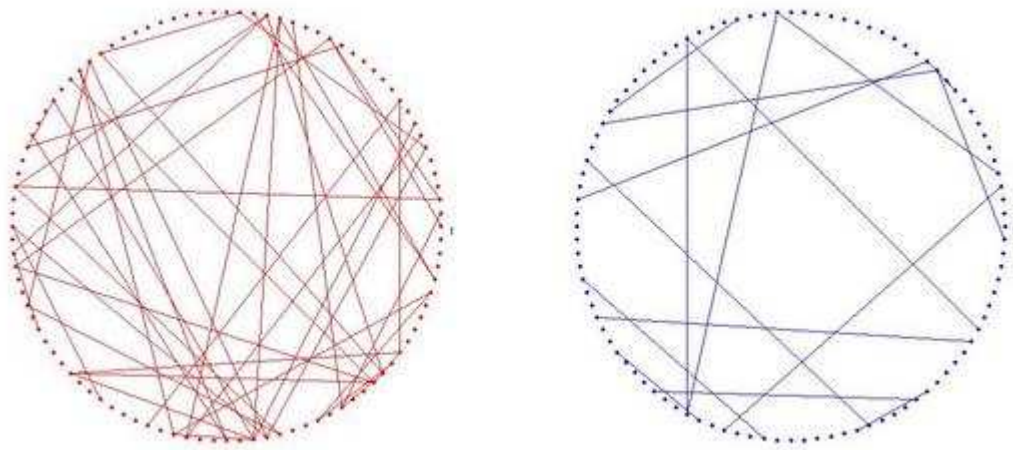


Figure 1.7: Supercritical random graph with  $c = 1.05$  for  $n = 100$  vertices. On the left, the cluster  $\mathcal{C}_1$  containing 1 is the giant component with 44 vertices. On the right, the complement of  $\mathcal{C}_1$  is made up of small clusters.

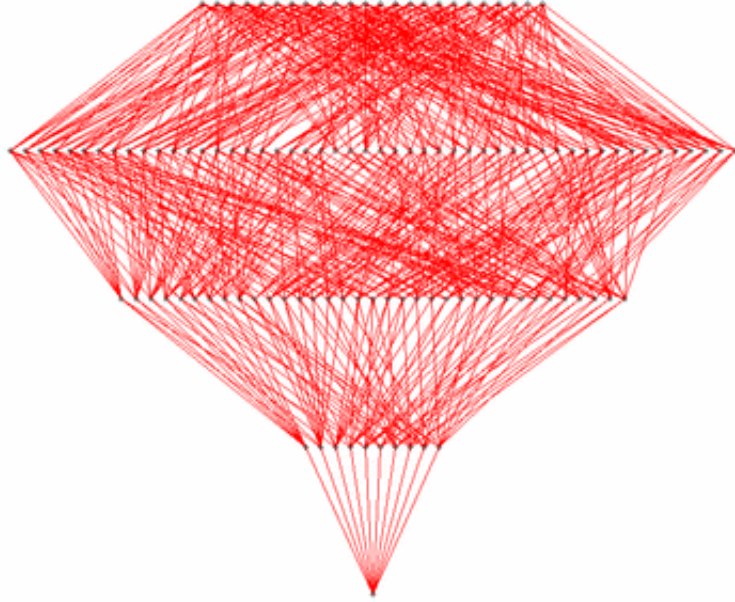


Figure 1.8: Cayley graph of  $\mathcal{S}_n$  generated by  $S = \{\text{all transpositions}\}$  for  $n = 5$ . At the bottom is the identity element, and permutations with equal distance from the identity are represented on the same level. The graph has  $5! = 120$  vertices and  $n! \binom{n}{2} / 2 = 600$  edges.

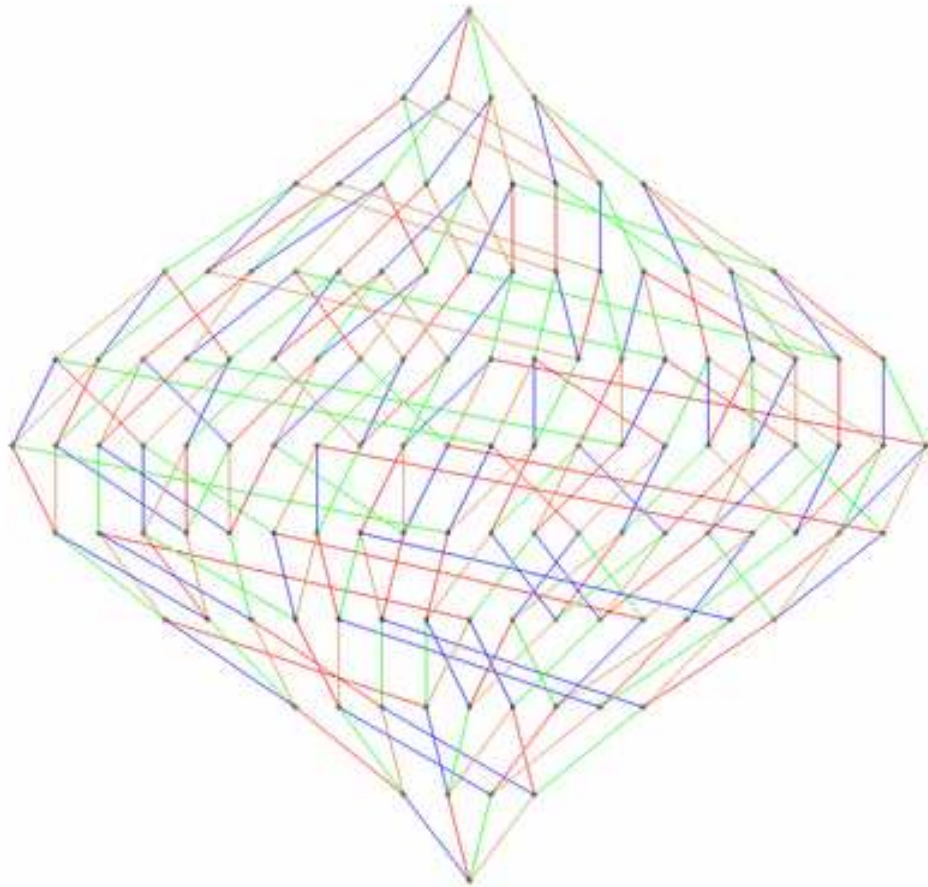


Figure 1.9: Cayley graph of  $\mathcal{S}_n$  generated by  $S = \{\text{adjacent transpositions}\}$  for  $n = 5$ . At the bottom is the identity element, and permutations with equal distance from the identity are represented on the same level. The graph has  $5! = 120$  vertices and  $n!(n - 1)/2 = 240$  edges.

# Chapter 2

## A phase transition in the random transposition random walk

by Nathanaël Berestycki<sup>1,2</sup> and Rick Durrett<sup>2</sup>

1. Ecole Normale Supérieure, D.M.A.
2. Cornell University, Department of Mathematics.

### 2.1 General motivation

The relationship between the orders of genes in two species can be described by a signed permutation. For example the relationship between the human and mouse  $X$  chromosomes may be encoded as (see Pevzner and Tesler (2003))

1 -7 6 -10 9 -8 2 -11 -3 5 4

In words the two  $X$  chromosomes can be partitioned into 11 segments. The first segment of the mouse  $X$  chromosome is the same as that of humans, the second segment of mouse is the 7th human segment with its orientation reversed, etc. The parsimony approach to estimation of evolutionary changes of the  $X$  chromosome between human and mouse is to ask: what is the minimum number of reversals (i.e., moves that reverse the order of a segment and therefore change its *sign*) needed to transform the arrangement above

back into  $1, \dots, 11$ ? In other words, what is the (reversal) distance between the human and mouse X chromosomes ?

Hannehalli and Pevzner (1995) developed a polynomial algorithm for answering this question. The first step in preparing to use the Hannehalli-Pevzner algorithm is to double the markers. When segment  $i$  is doubled we replace it by two consecutive numbers  $2i - 1$  and  $2i$ , e.g., 6 becomes 11 and 12. A reversed segment  $-i$  is replaced by  $2i$  and  $2i - 1$ , for example,  $-7$  is replaced by 14 and 13. The doubled markers use up the integers 1 to 22. To these numbers we add a 0 at the front and a 23 at the end. Using commas to separate the ends of the markers we can write the two genomes as follows:

mouse 0, 1 2, 14 13, 11 12, 20 19, 17 18, 16 15, 3 4, 22 21, 6 5, 9 10, 7 8, 23  
human 0, 1 2, 3 4, 5 6, 7 8, 9 10, 11 12, 13 14, 15 16, 17 18, 19 20, 21 22, 23

The next step is to construct the breakpoint graph (see Figure 1) that results when the commas are replaced by edges that connect vertices with the corresponding numbers. In the picture we have written the vertices in their order in the mouse genome. Commas in the mouse order become thick lines (black edges), while those in the human genome are thin lines (gray edges).

Each vertex has one black and one gray edge, so the connected components of the graph are easy to find: start with a vertex and follow the connections in either direction until you come back to where you start. In this example there are five components:

0 - 1 - 0      2 - 14 - 15 - 3 - 2      4 - 22 - 23 - 8 - 9 - 5 - 4  
19 - 17 - 16 - 18 - 19      13 - 11 - 10 - 7 - 6 - 21 - 20 - 12 - 13

To compute a lower bound for the distance, we take the number of commas seen when we write out one genome. In this example that is 12. In general, it is 1 plus the number of markers. We then subtract the number of components in the breakpoint graph. In this example that is 5, so the result is 7. This is a lower bound on the distance, since any reversal can at most reduce this quantity by 1, and it is 0 when the two genomes are the same. We can verify that 7 is the minimum distance by constructing a sequence of 7 moves that transforms the mouse X chromosome into the human order. There are thousands of solutions, so we leave this as an exercise for the reader. Here are some hints: (i) To do this it suffices, at each step, to choose a reversal that increases the number of cycles by 1. (ii) This never occurs if

the two chosen black edges are in different cycles. (iii) If the two black edges are in the same cycle and are  $(a, b)$  and  $(c, d)$  as we read from left to right, this will occur unless in the cycle minus these two edges  $a$  is connected to  $d$  and  $b$  to  $c$ , in which case the number of cycles will not change. For example, in the graph in Figure 1 a reversal that breaks black edges 19-17 and 18-16 will increase the number of cycles but the one that breaks 2-14 and 15-3 will not.

In general, the distance between genomes can be larger than the lower bound from the breakpoint graph. There can be obstructions called *hurdles* that can prevent us from decreasing the distance, and hurdles can be intertwined in a *fortress of hurdles* that takes an extra move to break. See Hannehalli and Pevzner (1995). In symbols, if  $\pi$  is the signed permutation that represents the relative order and orientation of segments in the two genomes, then

$$d(\pi) = n + 1 - c(\pi) + h(\pi) + f(\pi)$$

where  $d(\pi)$  is the distance from the identity,  $n$  is the number of markers,  $c(\pi)$  is the number of components in the breakpoint graph,  $h(\pi)$  is the number of hurdles, and  $f(\pi)$  is the indicator of the event  $\pi$  is a fortress of hurdles. See Section 5.2 of Durrett (2002) or Chapter 10 of Pevzner (2000) for more details.

Although  $d_0(\pi) = n + 1 - c(\pi)$  is only a lower bound on the distance, it is the right answer in most biological examples. Bafna and Pevzner (1995) consider 11 comparisons of mitochondrial and chloroplast genomes and found that this lower bound gave the right answer in all cases. This pattern has continued in more recent work, see York, Durrett, and Nielsen (2002), and Durrett, Nielsen, and York (2003). The simulations in Figure 2 will give more evidence that  $d_0(\pi)$  and  $d(\pi)$  are close in many cases.

To motivate our main question, we will introduce a second data set. Ranz, Casals, and Ruiz (2001) located 79 genes on chromosome 2 of *D. repleta* and on chromosome arm 3R of *D. melanogaster*. If we number the genes according to their order in *D. repleta* then their order in *D. melanogaster* is given in Table 1. This time we do not know the orientation of the segments, but that is not a serious problem. Using simulated annealing, one can easily find an assignment of signs that minimizes the distance, which in this case is 54. Given the large number of rearrangements relative to the number of markers, we should ask: when is the parsimony estimate reliable?

Bourque and Pevzner (2002) approached this question by taking 100

markers in order, performing  $k$  randomly chosen reversals to get a permutation  $\pi_k$ , computing the minimum number of reversals needed to return to the identity,  $d(\pi_k)$ , and then plotting the average value of  $d(\pi_k) - k \leq 0$  for 100 simulations. They concluded, based on their simulations, that the parsimony distance for  $n$  markers was a good estimate as long as the number of reversals performed was at most  $0.4n$ . In Figure 2 we have given  $-1$  times their data. We have also repeated their experiment for the approximate distance  $d_0(\pi) = n + 1 - c(\pi)$  and plotted the average value of  $k - d_0(\pi_k) \geq 0$  for 10,000 replications. Our curve is less random, but close to data of Bourque and Pevzner (2002). The smooth curve gives the result of Theorem 3 for the limiting behavior of  $(tn - d_0(\pi_{tn}))/n$  (as a function of  $t$ ).

The biological question concerns the random reversal walk. However, it is also interesting to consider the analogous problem for random transpositions. In that case the distance from the identity can be easily computed: it is the number of markers  $n$  minus the number of cycles in the permutation. For an example, consider the following permutation of 14 objects written in its cyclic decomposition:

$$(1\ 7\ 4)\ (2)\ (3\ 12)\ (5\ 13\ 9\ 11\ 6)\ (8\ 10\ 14)$$

which indicates that  $1 \rightarrow 7, 7 \rightarrow 4, 4 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 12, 12 \rightarrow 3$ , etc. There are 5 cycles so the distance from the identity is 9. If we perform a transposition that includes markers from two different cycles (e.g., 7 and 9) the two cycles merge into 1, while if we pick two in the same cycle (e.g., 13 and 11) it splits into two.

The situation is similar but slightly more complicated for reversals. There a reversal that involves edges in two different components merges them into 1, but a reversal that involves two edges of the same cycle may or may not increase the number of cycles. One can attempt to couple the components of the breakpoint graph for random reversals on  $n - 1$  markers and the cycles of random transposition of  $n$  markers as follows: number the edges between markers in the reversal chain (including the ends 0 and  $n$ ); when markers  $i$  and  $j$  are transposed, do the inversion of edges numbered  $i$  and  $j$ . The result of the coupled simulation is given in Figure 2. As expected time minus distance is smaller for reversals but the qualitative behavior is similar. Thus, we will begin by considering the biologically less relevant case of random transpositions, and ask a question that in terms of the rate 1 continuous time random walk on the symmetric group is: how far from the identity are

we at time  $cn/2$ ? We will see later that parts of the answer can be extended to the reversal random walk.

## 2.2 The coagulation-fragmentation process and the random graph process

Let  $(\sigma_t, t \geq 0)$  be the continuous-time random walk on the group of permutations, starting at the identity, in which, at times of a rate one Poisson process, we perform a transposition of two elements chosen uniformly at random, with replacement, from  $\{1, \dots, n\}$ . Choosing with replacement causes the chain to do nothing with probability  $1/n$ , but makes some of the calculations a little nicer. If we think of the permutation  $\sigma$  as being represented by numbered balls sitting on numbered locations with ball  $\sigma(k)$  sitting at  $k$ , then transposition of  $i$  and  $j$ ,  $\rho_{i,j}$ , can be implemented in two ways. We can exchange the balls at  $i$  and  $j$  or the balls numbered  $i$  and  $j$ . Algebraically these correspond to  $\rho_{i,j}\sigma$  and  $\sigma\rho_{i,j}$ . Since  $(\sigma\rho_{i,j})^{-1} = \rho_{i,j}\sigma^{-1}$  and the partition of  $\{1, \dots, n\}$  induced by the cycle decompositions of  $\sigma$  and  $\sigma^{-1}$  are equal, the results are the same for either random walk.

Define the distance to the identity  $D_t$  to be the minimum number of transpositions one needs to perform on  $\sigma_t$  to go back to the identity element. A different way of looking at  $D_t$  is the following.  $(\sigma_t, t \geq 0)$  can be viewed as a random walk on a graph  $G$ , where  $G$  is the Cayley graph of the symmetric group for the set of generators given by the set of all transpositions. Using this language, we see that  $D_t$  is nothing but the graph distance from  $\sigma_t$  to the origin, the identity element.

It is clear that if  $N_t$  is the number of transpositions distinct from the identity performed up to time  $t$  (a Poisson random variable with mean  $t(1 - 1/n)$ ), then  $D_t \leq N_t$ . As mentioned earlier  $D_t$  is given by  $D_t = n - |\sigma_t|$ , where  $|\sigma_t|$  is the number of cycles in the cycle decomposition of  $\sigma_t$ . This formula allows us to turn any question about  $D_t$  into a question about  $|\sigma_t|$ . The key to studying  $|\sigma_t|$  is that the cycles evolve according to the dynamics of a coagulation-fragmentation process. When a transposition  $\rho_{i,j}$  occurs, if  $i$  and  $j$  belong to two different cycles then the cycles merge. On the contrary, if they belong to the same cycle, this cycle is split into two cycles. From the definition it can be seen that the ranked sizes of the cycles form a coagulation-fragmentation process (see Aldous (1999) and Pitman (2002),(2003)) in which

components of size  $x$  and  $y$  merge at rate  $K_n(x, y) = 2xy/n^2$  and components of size  $x$  split at rate  $F_n(x) = x(x-1)/n^2$  and are broken at a uniformly chosen random point. Diaconis, Mayer-Wolf, Zeitouni, and Zerner (2003) have recently considered the corresponding Markov chain on partitions of the unit interval and shown that the Poisson-Dirichlet distribution is the unique invariant measure.

To study the evolution of the cycles in the random permutation, we construct a random graph process  $G_t^*$ . Start with the initial graph on vertices  $\{1, \dots, n\}$  with no edge between the vertices. When a transposition of  $i$  and  $j$  occurs in the random walk, we draw an edge between the vertices  $i$  and  $j$ , even if one is already present. Elementary properties of the Poisson process imply that if we collapse multiple edges in  $G_t^*$  into one, then the resulting graph  $G_t$  is a realization of the Erdős-Renyi random graph  $G(n, p)$ , in which edges are independently present with probability  $p = 1 - \exp(-2t/n^2)$ . The probability of picking an edge twice is  $\leq (2t/n^2)^2 = O(1/n^2)$  when  $t = cn/2$ , so the expected number of multiple edges is  $O(1)$ . Multiple edges are a nuisance but not a real problem. We have to be careful in Theorem 1 where the random variable of interest is also  $O(1)$ , but in the other cases the quantities of interest  $\rightarrow \infty$ , so multiple edges can be ignored.

It is easy to see that in order for two integers to be in the same cycle in the permutation it is necessary that they are in the same component of the random graph. To estimate the difference between cycles and components, let  $F_t$  denote the event that a fragmentation occurs at time  $t$ . It is clear that

$$D_t = N_t - 2 \sum_{s \leq t} \mathbf{1}_{\{F_s\}} \quad (2.1)$$

A fragmentation occurs in the random permutation when a transposition occurs between two integers in the same cycle, so tree components in the random graph  $G_t^*$  correspond to unfragmented cycles in the random walk. (To be precise, a tree is a connected component with no closed circuits and hence no multiple edges.) Unicyclic components in  $G_t^*$  (connected components with an equal number of vertices and edges) correspond to cycles in the permutation that have experienced exactly one fragmentation, but we need to know the order in which the edges were added to determine the resulting cycles. For more complex components, the relationship between the random graph and the permutation is less clear. Fortunately, these can be ignored in the proofs of our results. Coming back to the problem of multiple edges, the

reader should check that in the proof of Theorem 1, in particular in Lemma 4, it is indeed the number of fragmentations that is being counted, and not just the number of cycles in  $G_t$ .

## 2.3 Limit Theorems

We will now describe our results and sketch their proofs. Rigorous proofs of the results stated in this section can be found in sections 4, 5 and 6.

### 2.3.1 The subcritical regime

**Theorem 21.** *Let  $0 < c < 1$ . The number of fragmentations*

$$Z_c := \sum_{s \leq cn/2} \mathbf{1}_{\{F_s\}} \Rightarrow \text{Poisson}(\kappa(c)) \quad (2.2)$$

where  $\kappa(c) = (-\log(1-c) - c)/2$ . In fact, the convergence holds for the process  $\{Z_c : 0 \leq c < 1\}$  with the limit being a Poisson process with compensator  $\kappa(c)$ .

**Remark.** This result should be regarded as a fluctuation result for  $D_{cn/2}$  about its mean  $cn/2$  (a corollary of the Theorem). However we have chosen to formulate it in terms of fragmentations rather than the distance, since by (1)  $D_{cn/2} - cn/2 \approx N_{cn/2} - cn/2 = O(n^{1/2})$ . That is, in continuous time and in the subcritical regime the fluctuations are due to those of the Poisson process. However for the embedded discrete time chain, if  $k = \lfloor cn/2 \rfloor$ , then

$$(k - D_k)/2 \Rightarrow \text{Poisson}(\kappa(c)) \text{ as } n \rightarrow \infty \quad (2.3)$$

(We divide by 2 since a fragmentation reduces the distance by 1 instead of increasing it by 1). To deduce (2.3) from (2.2) we note that time  $k$  in the discrete walk corresponds to time  $N^{-1}(k) \approx cn/2$  in the continuous time walk.

**Sketch of the proof.** The process  $\{Z_c, 0 \leq c < 1\}$  is a càdlàg counting process. Therefore by arguments from Jacod and Shiryaev (1987), it is enough to show that its compensator  $\kappa^n$  converges to the deterministic limit  $\kappa(c)$ . If  $f_k(t)$  is the fraction of vertices that belong to cycles of size  $k$ , the rate at which fragmentations occur is just  $\sum_k f_k(t)(k-1)/n$ . Hence  $\kappa^n$  is just the

integral with respect to time of this rate. We first show that the variance converges to 0 and then, by Chebycheff's inequality, it only remains to show  $E\kappa^n(c) \rightarrow \kappa(c)$ . But by exchangeability  $E[f_k(t)] = P[|\mathcal{C}_1| = k]$  where  $|\mathcal{C}_1|$  is the size of the component that contains 1 at time  $t$ . It is not hard to see that this quantity at time  $bn/2$  converges in distribution to the total progeny  $\tau$  of a Galton-Watson branching process with offspring distribution  $\text{Poisson}(b)$ , or  $PGW(b)$ . Summing the geometric series, we see that  $E\tau = 1/(1-b)$ . Integrating with respect to  $b$  we get the desired expected value,  $\kappa(c)$ .  $\square$

To prepare for later developments, it is useful to take a second combinatorial approach to this result. We begin with Cayley's result that there are  $k^{k-2}$  trees with  $k$  labeled vertices. At time  $cn/2$  each edge is present with probability  $1 - \exp(-c/n) \sim c/n$  so the expected number of trees of size  $k$  present is

$$\sim \binom{n}{k} k^{k-2} \left(\frac{c}{n}\right)^{k-1} \left(1 - \frac{c}{n}\right)^{k(n-k) + \binom{k}{2} - k + 1} \quad (2.4)$$

since each of the  $k-1$  edges needs to be present and there can be no edges connecting the  $k$  point set to its complement or any other edges connecting the  $k$  points. (To justify replacing  $\exp(-c/n)$  by  $1 - c/n$  note that the difference is  $O(1/n^2)$ .) For fixed  $k$  (2.4) is asymptotic to

$$n \frac{k^{k-2}}{k!} c^{k-1} \left(1 - \frac{c}{n}\right)^{kn}$$

The quantity in parentheses at the end converges to  $e^{-ck}$  so we have an asymptotic formula for the number of tree components at time  $cn/2$ . As a side result we get the following known result:

**Corollary 1.** *The probability distribution of the total progeny  $T$  of a  $\text{Poisson}(c)$  branching process with  $c < 1$  is given by  $P(T = k) = \frac{1}{c} \frac{k^{k-1}}{k!} (ce^{-c})^k$*

See section 4.1 of Pitman (1999) for another proof of this result. It was first discovered by Borel (1942) and the distribution of  $T$  is called the Borel distribution. It is a particular case of the so-called Borel-Tanner distribution, see Devroye (1992) and Pitman (1998) for further references. In this context it appeared in the problem of the total number of units served in the first busy period of a queue with Poisson arrivals and constant service times. See also Tanner (1961). Of course, this becomes a branching process if we think of the customers that arrive during a person's service time as their children.

### 2.3.2 The critical regime

It is well known in the theory of random graphs that the correct time-scale to describe the critical regime is  $(n/2)(1 + \lambda n^{-1/3})$ ,  $\lambda \in (-\infty, \infty)$ . See Aldous (1997) for an interesting account that relates the growth of large clusters in the critical random graph to the multiplicative coalescent. At times  $(n/2)(1 - n^{-r})$  with  $r < 1/3$ , we are still in the subcritical regime, so the arguments in the proof of Theorem 21, when done more carefully, are still valid. More precisely, we can show that if  $c_n(r) = 1 - n^{-r/3}$  for  $0 \leq r \leq 1$ , then the expected number of fragmentations up to time  $c_n(r)n/2$  is again given by  $\kappa(c_n(r)) \sim (r/6) \log n$ . Hence define:

$$W_n(r) = \left( \frac{6}{\log n} \right)^{1/2} \left( \sum_{s \leq c_n(r)n/2} \mathbf{1}_{\{F_s\}} - \frac{r}{6} \log n \right) \quad (2.5)$$

**Theorem 22.** *As  $n \rightarrow \infty$ ,  $W_n(\cdot)$  converge weakly, with respect to the Skorokhod topology on the space of càdlàg functions on  $[0, 1]$ , to  $\{W(r), 0 \leq r \leq 1\}$ , a standard Brownian Motion on  $[0, 1]$ . Furthermore,*

$$\left( \frac{6}{\log n} \right)^{1/2} \left( \sum_{s \leq n/2} \mathbf{1}_{\{F_s\}} - \frac{1}{6} \log n \right) \Rightarrow W(1), \quad (2.6)$$

**Sketch of the proof.** Intuitively, the first result is an immediate consequence of the Poisson limit in Theorem 21 and the normal approximation to the Poisson. To prove it, we show that  $W_n(r)$  is a martingale, whose jumps are asymptotically zero, and whose quadratic variation process is  $r$  thanks to our time-change  $c_n(r) = 1 - n^{-r/3}$ . Therefore it converges to Brownian Motion.

At times  $(1 - n^{-1/3})n/2 \leq t \leq n/2$  we are in the critical range of the random graph. Results of Luczak, Pittel, and Wierman (1994) and computations with (2.4) imply that the number of fragmentations in this interval is bounded in expectation and hence can be ignored.  $\square$

**Remark.** While Theorem 22 is a nice theoretical result, it does not have much to say about any biological example. If we think of the human genome and set  $n = 3$  billion nucleotides, Theorem 2 says that after  $n/2 = 1.5$  billion transpositions there have been an average of  $(\log n)/6 = 3.63$  fragmentations, with a standard deviation of 1.91. These numbers are small so even for  $n = 3$

billion, we can't expect a very good approximation to the normal distribution. In the example that we simulated  $n = 100$  and  $(\log n)/6 = 0.767$  versus an observed average number of fragmentations = 0.662 (which translates into a value of 1.224 in Figure 2). While our estimation of the mean is not very accurate, Figure 3 shows that the distribution of the number of fragmentations is almost Poisson.

### 2.3.3 The supercritical regime

This is the most interesting case, and also the hardest one. We start by establishing a law of large numbers. For all  $c > 0$  define

$$\beta_k(c) = \frac{1}{c} \frac{k^{k-1}}{k!} (ce^{-c})^k$$

so that for  $c < 1$  it coincides with the Borel distribution of Corollary 1. When  $c > 1$ ,

$$\lim_{n \rightarrow \infty} P(|\mathcal{C}_1| = k) = \beta_k(c)$$

still holds but the  $\beta_k(c)$ 's no longer sum up to 1 because there is a probability  $\beta_\infty(c) = 1 - \sum_{k \geq 1} \beta_k(c) > 0$  that  $\mathcal{C}_1$  is the giant component.

Let us denote by  $\Upsilon(c)$  a random variable that takes the value  $1/k$  with probability  $\beta_k(c)$  when  $1 \leq k < \infty$  and the value 0 with probability  $\beta_\infty(c)$ . The motivation for this definition is that in the limit as  $n \rightarrow \infty$   $1/|\mathcal{C}_m|$  has the same distribution as  $\Upsilon(c)$  and  $\sum_{m=1}^n 1/|\mathcal{C}_m|$  gives the number of components in the random graph.

**Theorem 23.** *Let  $c > 0$  be a fixed positive number. Then the number of cycles in the random permutation at time  $cn/2$ ,  $|\sigma_{cn/2}| = g(c)n + \omega(\sqrt{n})$ , where*

$$g(c) := E\Upsilon(c) = \sum_{k=1}^{\infty} \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k \quad (2.7)$$

and the error term  $\omega(\sqrt{n})/a_n\sqrt{n} \rightarrow 0$  in probability if  $a_n \rightarrow \infty$ .

Note that the theorem is valid for all regimes and implies that the distance is given by  $D_{cn/2} = u(c)n + \omega(\sqrt{n})$  where  $u(c) = 1 - g(c)$ . Although it is not obvious from the formula,  $u(c) = c/2$  for  $c < 1$  and  $u(c) < c/2$  when  $c > 1$ . Using Stirling's formula,  $k! \sim k^k e^{-k} \sqrt{2\pi k}$ , it is easy to check that  $g'$  exists for all  $c$  and is continuous, but  $g''(1)$  does not exist. In words, there is phase

transition in the behavior of the distance of the random walk to the identity at time  $n/2$  from linear to sublinear.

*Proof.* In the supercritical regime the dynamics of the large components is quite complicated, but there can never be more than  $\sqrt{n}$  components of size  $\sqrt{n}$  or larger. On the other hand, the expected number of all fragmentations that produce (regardless of the initial size of the cycle) clusters of size smaller than  $\sqrt{n}$  by time  $cn/2$  is at most  $O(n^{1/2})$ . This follows from the following important remark, which will be used on several other occasions implicitly: suppose  $C = (x_1, \dots, x_k)$  is a cycle of the permutation  $\sigma$ . If we transpose  $x_i$  and  $x_j$  with  $x_i \neq x_j$  then  $C$  breaks into  $(x_1, \dots, x_{i-1}, x_j, x_{j+1}, \dots, x_k)$  on the one hand and  $(x_i, \dots, x_{j-1})$  on the other hand. So, to generate a fragment of size  $s$  we must transpose two elements  $(x, y)$  in the same cycle and such that  $x$  and  $y$  are separated by exactly  $s-1$  other integers in this cycle. In particular, fragmentations that produce pieces smaller than  $s$  occur with a rate smaller than  $2s/n$ . When  $s = n^{1/2}$  this gives a rate smaller than  $O(n^{-1/2})$ . Thus by time  $cn/2$  there have been no more than  $O(n^{-1/2}) \cdot cn/2 = O(n^{1/2})$  such fragmentations. From this and Chebyshev's inequality we see that up to a term  $\omega(n^{1/2})$ ,  $|\sigma_{cn/2}|$  is the number of components of the random graph, and the result follows Theorem 12 in Chapter V of Bollobás (1985).  $\square$

**Theorem 24.** *Let  $c > 1$ . As  $n \rightarrow \infty$ ,*

$$\frac{D_{cn/2} - u(c)n}{n^{1/2}} \Rightarrow \mathcal{N}(0, \sigma^2) \quad (2.8)$$

where  $\sigma = \rho[1 + \rho(c/2 - 1)]$ , and  $\rho = 1 - \theta(c)$  is the extinction probability of a supercritical PGW( $c$ ).

**Remark.** Note that the constant  $\sigma$  is different from the one given in Berestycki and Durrett (2003). We were correct in claiming that the central limit theorem in Theorem 24 is the same as the one for the number of components of the random graph, but we naively thought that the terms in  $\sum_{k=1}^n 1/|C_k|$  were sufficiently independent so that  $\sigma^2 = \text{var}(\Upsilon(c))$ .

**Sketch of Proof.** By Pittel's (1990) central limit theorem for the number of components of a random graph, it suffices to prove that the number of extra components due to fragmentation at time  $cn/2$  is  $o(\sqrt{n})$  (see his Corollary 1 and note that  $T/c = \rho$ ). Our first step is to increase the cutoff for large cycles to  $n^a$  where  $a > 1/2$ , so that the number of large cycles is at most

$n^{1-a} = o(n^{1/2})$ . The number of fragmentations that produce “small” cycles is now  $n^{-(1-a)} \cdot cn/2 = O(n^a)$  and cannot be ignored, so we need to use the fact that fragmented cycles are reabsorbed by the large components. If the fraction of mass in large cycles (“upstairs”) at time  $tn$  is  $\lambda_t$  then new fragments of size  $k$  are produced at rate  $\leq 2\lambda_t$  and each fragment of size  $k$  is reabsorbed at rate  $2k\lambda_t$ . After time change this is bounded by an  $M/M/\infty$  queue in which the expected number of customers in equilibrium is  $1/k$ . Using this, we can show that with high probability the number of small fragments at any time is at most  $(\log n)^2$ . Of course, the coagulation fragmentation process is not exactly the queuing system. Customers can split into two, coalesce with other customers, gain weight (and increase their fragmentation rate) by eating small components, etc. However,  $(\log n)^2$  is much smaller than  $n^{1/2}$  so crude but robust estimates and patience eventually lead to a proof.  $\square$

### 2.3.4 Results for Reversals.

Theorems 3 and 4 extend easily to the approximate distance for reversal chain. Recall that the main difference lies in the fact that, a reversal involving edges from different components in the breakpoint graph always yields a coagulation, but one involving two edges in the same component may or may not cause a fragmentation. The proofs of Theorems 3 and 4 for transpositions are based on showing that fragmentations can be ignored, so this difference is unimportant and these results extend to reversals. As Figure 2 shows, this is not true for the more precise results in Theorems 21 and 22. For example, the underlying data shows that up to  $c = 1$ , an average of 23% of the reversals have caused no change in the distance. Since inversions that affect an edge are much more frequent than those that involve it, it seems reasonable to guess that in the limit as  $n \rightarrow \infty$  the relative orientations of the black edges in a component of the breakpoint graph are independent. This would imply that the Poisson process of fragmentations in the reversal case is a 1/2-thinning of the one for transpositions, and Theorem 22 would hold with 6 replaced by 12.

### 2.3.5 Emergence of a giant cycle?

Since cycles in the random permutation are smaller than components of the random graph, it follows that if  $c < 1$  then the largest cycle at time  $cn/2$  has

fewer than  $\alpha(c)^{-1} \log n$  vertices, where  $\alpha(c) = (c - 1 - \log c)$ . (See Theorem 10 in Chapter V of Bollobás (1985) or Lemma 3 below.)

For  $c > 1$ , the largest component of the random graph is, as is well known, “giant,” meaning that it is of order  $n$ . In fact it is asymptotic to  $\theta(c)n$  where  $\theta(c)$  is the survival probability of a supercritical Poisson Galton-Watson with mean  $c$ . It is a natural question to ask whether the largest cycle of the random permutation is also giant in the supercritical regime.

**Conjecture.** *Let  $L_1(t)$  be the size of the largest cycle at time  $t$ . If  $c > 1$  then*

$$\frac{L_1(cn/2)}{\theta(c)n} \Rightarrow V$$

where  $V$  is a random variable with  $0 < V \leq 1$  a.s.

This problem is quite different from our original one. However our techniques enable us to prove a partial result in this direction as a corollary of the proof of Theorem 24.

**Theorem 25.** *For any  $c > 1$ , at time  $cn/2$  there are at least  $\theta(c)n - o(n)$  vertices located on large cycles (i.e., of size greater than or equal to  $n^a$ , for any  $a < 2/3$ ).*

David Aldous (private communication) conjectures that the relative sizes of the pieces of the giant cycle are in equilibrium at all times in the supercritical regime, i.e., have the Poisson-Dirichlet  $PD(0, 1)$  distribution, which gives the limiting behavior of the ordered sizes of cycles in a uniform random permutation. According to this conjecture,  $V$  would be distributed as the first coordinate of a  $PD(0, 1)$  random variable. One way to approach this conjecture would be to generalize Aldous (1997) to show that the large cycles in the critical regime converge to a coagulation-fragmentation process and to study the growth of clusters in that process.

Alternatively, one could look at the size of the cycle containing 1,  $K_1(t)$ , and try to show that

$$\frac{K_1(cn/2)}{\theta(c)n} \Rightarrow U$$

where  $U$  has a point mass of size  $1 - \theta(c)$  at 0 and is otherwise uniform on  $(0, \theta(c))$ . Figure 4 shows the average growth of  $K_1(cn/2)/n$  in 10,000 simulations of  $n = 100$ ,  $n = 1000$ , and compares the results to  $EU = \theta(c)^2/2$ .

Although this considers only one aspect of the distribution of large cycles, it agrees well with Aldous' conjecture.

Figure 5 shows a histogram of the result of 100,000 simulations of  $K_1(100)$  when  $n = 100$ . As the graph shows, the spike in the frequency of clusters of size 4 or smaller is what one would predict from the random graph cluster size distribution. The remainder of the distribution is roughly uniform except for rounding at the upper end. The latter is to be expected if Aldous' conjecture is correct, since the size of the giant component satisfies the central limit theorem.

As we were finishing this paper, we learned that Oded Schramm (2004) has proved David Aldous' conjecture.

**Remark.** The problem of the emergence of a giant cycle is closely related to Angel's (2003) work on the existence of infinite orbits for the *random stirring process*, which is the random transposition random walk on an infinite graph such as  $\mathbb{Z}^d$  or a tree, rather than the complete graph on  $\{1, \dots, n\}$  considered in this work. To explain the connection, suppose that we construct our process using a Poisson process with rate  $2/n^2$  for each  $i \neq j$ , and at these times draw an edge between  $i$  and  $j$  to indicate that  $i$  and  $j$  are to be transposed. To compute the cycles in the permutation at time  $cn/2$ , we repeat the first  $[0, cn/2]$  units of time periodically and then observe the sites that a walker starting at  $i$  visits at times  $kcn/2$ , for  $k = 1, 2, \dots$ . Angel (2003) calls this construction the *cyclic time random walk*. Its relevance to his work is that the cyclic time random walk is transient if, and only if, the cycles are infinite.

## 2.4 The subcritical regime

Let us introduce some notations for the different probability laws involved. For each  $n$ , we have the coagulation-fragmentation process, and the Erdős-Renyi random graph model. To emphasize when computations are being done for the random graph we will use  $Q_p$ , for the random graph with Bernoulli percolation parameter  $p$ , and  $Q$  for the law of the evolving random graph that at time  $s$  has  $p_s = 1 - \exp(-2s/n^2)$ . When  $s = cn/2$  this probability is  $p(c, n) = 1 - \exp(-c/n) \leq c/n$ . To simplify notation we will use  $QX$  to denote the expected value of  $X$  with respect to the probability  $Q$ .

### 2.4.1 Preliminary results : comparison with a branching process

Our first result provides a useful upper bound.

**Lemma 1.** *The cluster size  $|\mathcal{C}_1|$  in  $Q_{c/n}$  is dominated by  $Z$ , the total progeny of a branching process in which each individual has a Binomial( $n - 1, c/n$ ) number of children, i.e., we can construct these random variables on the same probability space so that  $|\mathcal{C}_1| \leq Z$  a.s. It follows from this that if  $c < 1$  then  $Q_{c/n}|\mathcal{C}_1| \leq EZ = 1/(1 - c)$ .*

*Proof.* Intuitively, this holds since a vertex in generation  $k$  may have children among all of the  $n$  vertices of the graph except those of the first  $k$  generations. To begin to prove this formally, let  $\xi_{i,j}$ ,  $1 \leq i, j \leq n$  be independent random variables, taking values 1 with probability  $c/n$  and 0 with probability  $1 - c/n$ . To start the random graph let  $Y_0 = \{1\}$  and let  $Y_1 = \{j \notin Y_0 : \xi_{1,j} = 1\}$ . To start the branching process let  $Z_0 = 1$ ,  $Z_1 = |Y_1|$ , and let  $\phi_1 : Y_1 \rightarrow \{1, 2, \dots, Z_1\}$  be 1-1 and onto.

If the first  $k$  stages of the construction have been done and we have  $Y_k \neq \emptyset$  and a  $\phi_k : Y_k \rightarrow \{1, \dots, Z_k\}$  that is 1-1 (but not onto in general), then let

$$Y_{k+1} = \cup_{i \in Y_k} \{j \notin \cup_{\ell=0}^k Y_\ell : \xi_{i,j} = 1\}$$

We let individual  $\phi_k(i)$  in the  $k$ th generation of the branching process have  $|\{j \neq i : \xi_{i,j} = 1\}|$  children. The individuals in the branching process that are not in  $\phi_k(Y_k)$  have a number of children given by independent binomials. It should be clear from the construction that we can again define  $\phi_{k+1} : Y_{k+1} \rightarrow \{1, \dots, Z_{k+1}\}$  to be 1-1, and the comparison follows by induction. The inequality follows by computing  $EZ$  (for instance by summing a geometric series).  $\square$

The next result shows that the bound in Lemma 1 is exact in the limit. Let  $\{Z_k\}_{k=0}^\infty$  be a Poisson Galton-Watson process with offspring mean  $c$  and let  $Z = \sum_{k=0}^\infty Z_k$  be its total progeny.

**Lemma 2.** *Let  $\mathcal{C}_1$  be the cluster that contains vertex 1. If  $0 \leq c < 1$  then as  $n \rightarrow \infty$*

$$Q_{p(c,n)}(|\mathcal{C}_1| = k) \rightarrow P(Z = k)$$

*Proof.* The number of children of vertex 1,  $Z_n^1 = |Y_1|$  has distribution Binomial( $n-1, p(c, n)$ ), which converges to a Poisson( $c$ ) limit. Let  $k \geq 1$  and let  $(n_1, \dots, n_{k+1}) \in \mathbb{N}^{k+1}$ . If we let  $Z_j^n = |Y_j|$  then

$$Q_{p(c,n)}(Z_{k+1}^n = n_{k+1} | Z_1^n = n_1, \dots, Z_k^n = n_k) = P\left(\sum_{i=1}^{n_k} B_i^n = n_{k+1}\right)$$

where  $B_i^n$  are i.i.d. Binomial( $n-s, p(c, n)$ ) random variables, and  $s = \sum_{i=0}^k n_k$  with  $n_0 = 1$ . From this it follows easily that the convergence of finite-dimensional distributions of  $\{Z_j^n\}_{j \geq 1}$  to those of  $PGW(c)$ . Markov's inequality and the domination result in Lemma 1 imply that

$$Q_{p(c,n)}\left(\sum_{k=K}^{\infty} Z_k^n > 0\right) \leq Q_{p(c,n)}\left(\sum_{k=K}^{\infty} Z_k^n\right) \leq c^K / (1-c)$$

and the desired conclusion follows.  $\square$

Our next ingredient is

**Lemma 3.**  $Q_{c/n}(|\mathcal{C}_1| \geq y) \leq c^{-1} \exp(-(c-1-\ln c)y)$ .

*Proof.* In view of Lemma 1, it suffices to prove the result for  $Z$ , rather than  $|\mathcal{C}_1|$ . To do this, let

$$\begin{aligned} \phi_n(\theta) &= e^{-\theta} \sum_{m=0}^{n-1} \binom{n-1}{m} \left(\frac{c}{n}\right)^m \left(1 - \frac{c}{n}\right)^{n-1-m} e^{\theta m} \\ &= e^{-\theta} \left(1 - \frac{c}{n} + \frac{c}{n} e^{\theta}\right)^{n-1} \end{aligned}$$

be the moment generating function of the distribution of the number offspring minus 1. Let  $S_m$  be a random walk that takes steps with this distribution and  $S_0 = 1$ , so that  $S_m$  explores the Galton-Watson tree. Then  $\tau = \inf\{m : S_m = 0\}$  has the same distribution as  $Z$ . Let  $R_m = \exp(\theta S_m) / \phi_n(\theta)^m$ .  $R_m$  is a nonnegative martingale. Stopping at time  $\tau$  we have  $e^{\theta} \geq E(\phi_n(\theta)^{-\tau})$ . If  $\phi_n(\theta) < 1$  it follows that

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

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

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

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

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

It follows that

$$Q_{c/n}(|\mathcal{C}_1| \geq y) \leq \frac{1}{c} \exp(-(c - 1 - \ln c)y)$$

which completes the proof of Lemma 3.  $\square$

Now recall that for  $c < 1$ ,  $Z_c = \sum_{s \leq cn/2} \mathbf{1}_{\{F_s\}}$  is the number of fragmentations up to time  $cn/2$ . The rate at which fragmentations occur is  $\sum_k f_k(t)(k-1)/n$  where  $f_k(t)$  is the fraction of vertices that belong to cycles of size  $k$ . Therefore the expected number of fragmentations is

$$EZ_c = \int_0^{cn/2} \sum_{k=1}^n E f_k(t) \frac{k-1}{n} dt$$

**Remark.** Note that this formula takes into account the case of multiple edges in the random graph, i.e. the possibility that a given transposition may be chosen twice as an increment of the random walk before time  $cn/2$ .

**Lemma 4.** *Let  $f_k(s)$  be the empirical fraction of vertices in cycles of size  $k$  at time  $s$ . If  $0 \leq c < 1$  then  $E f_k(cn/2) \rightarrow P(Z = k)$  and  $EZ_c \rightarrow \kappa(c)$ , where  $\kappa(c)$  was defined in Theorem 21.*

*Proof.* The cycle sizes at time  $s$  in the coagulation-fragmentation process are dominated by the cluster sizes in the random graph model with  $p_s = 1 - \exp(-2s/n^2) \leq 2s/n^2$ . Note that by exchangeability the expected fraction of vertices in clusters of the random graph of size  $k$  satisfies  $E(f_k) = P(|\mathcal{C}_1| = k)$ . (This also holds under  $Q$ , and we will use it below). By Lemma 3,

$$\begin{aligned} P(|\mathcal{C}_1| \text{ fragments before time } cn/2) &\leq E(\#\text{such fragmentations}) \\ &\leq \int_0^{cn/2} E \left( \frac{|\mathcal{C}_1|^2}{n^2} \right) dt \\ &\leq \int_0^{cn/2} Q_{2t/n^2} \left( \frac{|\mathcal{C}_1|^2}{n^2} \right) dt \rightarrow 0 \end{aligned}$$

Therefore  $P(|\mathcal{C}_1| = k)$  has the same asymptotics as  $Q(|\mathcal{C}_1| = k)$  and we can conclude for the first convergence by Lemma 2.

For the second convergence, first note that

$$EZ_c \leq \int_0^{cn/2} \sum_{k=1}^n Q f_k(s) \frac{k-1}{n} ds \leq \int_0^{cn/2} Q_{2s/n^2} \left( \frac{|\mathcal{C}_1| - 1}{n} \right) ds$$

Using Lemma 1  $Q_{2s/n^2} |\mathcal{C}_1| \leq 1/(1 - (2s/n))$ . Changing variables  $un/2 = s$  we have

$$EZ_c \leq -\frac{1}{2}(\log(1-c) + c) = \kappa(c) \quad (2.9)$$

For the lower bound we use Fatou's lemma (twice) and the first convergence (recall also that if  $Z(s)$  is the total progeny of a  $PGW(s)$  then  $EZ(s) = 1/(1-s)$ ):

$$\begin{aligned} \liminf_{n \rightarrow \infty} EZ_c &= \liminf_{n \rightarrow \infty} \int_0^{cn/2} \sum_{k=1}^n E f_k(t) \frac{k-1}{n} dt \\ &\geq \frac{1}{2} \int_0^c \liminf_{n \rightarrow \infty} \sum_{k=1}^n k E(f_k(sn/2)) - 1 ds \\ &\geq \frac{1}{2} \int_0^c \sum_{k=1}^{\infty} k P(Z(s) = k) - 1 ds = \frac{1}{2} \int_0^c \frac{1}{1-s} - 1 ds \\ &\geq \kappa(c) \end{aligned}$$

□

The final preparatory step is:

**Lemma 5.** *If  $c < 1$  the expected number of fragmentations that occur to cycles that have already been fragmented is  $\leq K_c(\log n)^2/n$ , and  $K_c = 9c\kappa(c)\alpha(c)^{-2}$ . (Recall  $\alpha(c) = (c - 1 - \log c)$ ).*

*Proof.* The expected number of such fragmentations is at most:

$$\begin{aligned} &\leq E \int_0^{cn/2} \frac{\#\text{vertices in fragments}}{n} \frac{L_1(bn/2)}{n} dt \\ &\quad \frac{n}{2} \int_0^c EZ_b \left( \frac{L_1(bn/2)}{n} \right)^2 db \end{aligned}$$

where  $L_1(t)$  is the size of the largest component at time  $t$ . In the event that  $L_1(cn/2) \leq 3\alpha(c)^{-1} \log n$ , the above is at most

$$(n/2)(3\alpha(c)^{-1} \log n/n)^2 \int_0^c \kappa(b) db \leq \frac{1}{2} K_c \frac{(\log n)^2}{n}$$

On the other hand by Lemma 3 the complement of this event has probability at most  $n^{-2}$ , and there can never be more than  $cn/2$  such fragmentations, so Lemma 5 is proved.  $\square$

## 2.4.2 Proof of Theorem 1

We are now ready to prove Theorem 21. Let  $\bar{Z}_c^n = \sum_{s \leq cn/2} \mathbf{1}_{\{\bar{F}_s\}}$ ,  $0 \leq c < 1$  be the counting process of fragmentations that occur to cycles which (a) have not been fragmented previously and (b) have size  $\leq n^{0.7}$ . The second condition is irrelevant in this section, but imposing it now will help in the next one. Unfragmented cycles correspond to trees in the random graph (as explained in the introduction, by tree we mean a connected components with no closed circuit, including multiple edges), so the compensator of  $\bar{Z}_c^n$  is

$$\bar{\kappa}^n(c) = \int_0^{cn/2} \bar{\psi}_s^n ds \quad (2.10)$$

where  $\bar{\psi}_s^n = \sum_{k=1}^{n^{0.7}} \bar{f}_k(s)(k-1)/n$  and  $\bar{f}_k(s)$  is the fraction of vertices that belong to tree components of size  $k$ . As noted in the sketch of the proof, it is enough to show that for each fixed  $c$ ,  $\bar{\kappa}^n(c)$  converges in probability to  $\kappa(c)$ , or, by Lemma 5, that  $\bar{\kappa}^n(c)$  converges to  $\kappa(c)$  in probability. Lemmas 3 and 4 imply that  $E[\int_0^{cn/2} \bar{\psi}_s^n ds] \rightarrow \kappa(c)$ . It remains to show that  $\text{var} \int_0^{cn/2} \bar{\psi}_s^n ds \rightarrow 0$ . Our first step will be to prove :

$$\text{var}(\bar{\psi}_s^n) \leq \frac{K}{n^3} Q_{p(c,n)}[|\mathcal{C}_1|^3] \quad (2.11)$$

for all time  $s \leq cn/2$ , where  $K$  is a constant that depends only on  $c$ .

To see this, first observe that in terms of cluster sizes

$$\bar{\psi}_s^n = \frac{1}{n^2} \sum_{i=1}^n (|\mathcal{C}_i| - 1) I_i$$

where  $I_i$  is the indicator of the event that  $\mathcal{C}_i$  is a tree. Let  $d_i = (|\mathcal{C}_i| - 1)I_i$ .

$$\text{var} \frac{1}{n^2}(d_1 + \cdots + d_n) = \frac{1}{n^4} (n \text{var}(d_1) + n(n-1) \text{cov}(d_1, d_2)) \quad (2.12)$$

Monotonicity and Lemma 3 imply,

$$\text{var}(d_1) \leq Q_{p(c,n)}[|\mathcal{C}_1|^2] \leq K \quad (2.13)$$

It remains to bound  $\text{cov}(d_1, d_2)$ . If we let

$$\pi_{n,i} = i^{i-2} (p_s)^{i-1} (1-p_s)^{i(n-i) + \binom{i}{2} - (i-1)}$$

where here  $p_s = \exp(-2s/n^2)2s/n^2$  (this is the probability that a given edge appears exactly once, since we don't want any multiple edges), then by the reasoning for (2.4) we have

$$\begin{aligned} Q_{p_s} [\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset, |\mathcal{C}_1| = j, |\mathcal{C}_2| = k, \mathcal{C}_1 \text{ and } \mathcal{C}_2 \text{ are trees}] \\ = \binom{n-2}{j-1} \pi_{n,j} \binom{n-j-1}{k-1} \pi_{n-j,k} \\ Q_{p_s} [\mathcal{C}_1 = \mathcal{C}_2, |\mathcal{C}_1| = k, \mathcal{C}_1 \text{ is a tree}] = \binom{n-2}{k-2} \pi_{n,k} \end{aligned}$$

From this it follows that  $\text{cov}(d_1, d_2)$

$$\begin{aligned} &= \sum_{j,k} \left[ \binom{n-2}{j-1} \binom{n-j-1}{k-1} (1-p_s)^{-j} - \binom{n-1}{j-1} \binom{n-1}{k-1} \right] \\ &\quad \times (j-1)(k-1) \pi_{n,j} \pi_{n,k} \\ &+ \sum_k \binom{n-2}{k-2} (k-1)^2 \pi_{n,k} \end{aligned}$$

For the first term in the right-hand side,

$$\begin{aligned} &\left[ \binom{n-2}{j-1} \binom{n-j-1}{k-1} (1-p_s)^{-j} - \binom{n-1}{j-1} \binom{n-1}{k-1} \right] \\ &\leq \frac{(n-2)! e^{2c}}{(j-1)!(k-1)!(n-j-k)!} - \frac{(n-1)!}{(j-1)!(n-j)!} \frac{(n-1)!}{(k-1)!(n-k)!} \\ &\leq 0 \end{aligned}$$

where we have used for the first inequality that  $-\log(1-x) \leq 2x$  if  $|x|$  is small enough, and for the second one,  $(n-2)!e^{2c} \leq (n-1)!$  for large  $n$  and  $(n-j)!/(n-j-k)! \leq (n-1)!/(n-1-k)!$ .

For the second term,

$$\sum_k \binom{n-2}{k-2} (k-1)^2 \pi_{n,k} \leq \frac{1}{n-1} \sum_k k^3 \binom{n-1}{k-1} \pi_{n,k} \leq \frac{1}{n-1} Q_{p(c,n)}[|\mathcal{C}_1|^3]$$

Combining this with (2.12) and (2.13) gives (2.11).

Hence by the Cauchy-Schwarz inequality we get:

$$\begin{aligned} \text{var} \left( \int_0^t \bar{\psi}_s^n ds \right) &= Q \left[ \left( \int_0^t (\bar{\psi}_s^n - Q[\bar{\psi}_s^n]) ds \right)^2 \right] \leq t \int_0^t \text{var}(\bar{\psi}_s^n) ds \\ &\leq \frac{cn}{2} \int_0^{cn/2} \frac{K}{n^3} ds = \frac{c^2 K}{4n} \rightarrow 0 \end{aligned}$$

where we have used both (2.11) and Lemma 3. This completes the proof of Theorem 21.

## 2.5 The critical regime

The first step in the proof of Theorem 22 is to argue that fragmentations of previously fragmented cycles can be ignored. The number of such fragmentations is smaller than the total number of cycles in multicyclic components (i.e., components with at least 2 cycles) in the random graph. Theorem 1 and Corollary 3 in Luczak, Pittel, and Weirman (1994) imply that the total number of cycles in multicyclic components in the critical regime is bounded in probability.<sup>1</sup> In particular, divided by  $(\log n)^{1/2}$  it converges to 0 in probability. As a result, by the converging together lemma (see e.g., Durrett (1996), Chap.2, Ex. 2.10), it suffices to prove the central limit theorem for the number of fragmentations on tree components.

As in the previous section, we will in addition restrict our attention to fragmentations of tree components of size at most  $n^{0.7}$ , and continue to use the notation introduced there. (Indeed, classical results from the theory of random graphs, or Aldous (1997), show that asymptotically almost surely all clusters are smaller than  $n^{0.7}$ ).

---

<sup>1</sup>This result could also be derived from the Folk Theorem 1 in Aldous (1997) which gives the limit for the joint distribution of the component sizes and the number of cycles they contain. See the discussion page 850 of his paper.

Let  $\bar{W}_n(r) := (6/\log n)^{1/2}(\bar{Z}^n(r) - \bar{\kappa}^n(r))$ . By standard methodology in the theory of stochastic processes (see Jacod and Shiryaev (1987) or Revuz and Yor(1999) for instance), to prove convergence of  $\bar{W}_n(\cdot)$  to Brownian Motion, the two things we need to check are: (i)  $E[\sup_{0 \leq r \leq 1} |\bar{W}_n(r) - \bar{W}_n(r^-)|] \rightarrow 0$  and (ii) The quadratic variation of  $\bar{W}_n$ , i.e. the increasing process associated with  $\bar{W}_n(\cdot)^2$ , must converge to  $r$  at time  $r$ . (i) is obvious because  $\bar{Z}^n$  is a counting process, and (ii) turns into  $E(6\bar{\kappa}^n(r)/\log n) \rightarrow r$  and  $\text{var}(6\bar{\kappa}^n(r)/\log n) \rightarrow 0$ . These two steps are dealt with respectively in lemmas 7 and 8.

But first, we need a technical lemma that will be useful on several occasions (e.g., for computing precise asymptotics of the number of trees of a given size).

**Lemma 6.** *If  $k \rightarrow \infty$  and  $k = o(n^{3/4})$  then*

$$\begin{aligned} \gamma_{n,k}(c) &\equiv \binom{n}{k} k^{k-2} \left(\frac{c}{n}\right)^{k-1} \left(1 - \frac{c}{n}\right)^{kn-k^2/2-3k/2+1} \\ &\sim \frac{nk^{-5/2}}{c\sqrt{2\pi}} \exp\left(-\alpha(c)k + (c-1)\frac{k^2}{2n} - \frac{k^3}{3n^2}\right) \equiv \lambda_{n,k}(c) \end{aligned}$$

where  $\alpha(c) = c - 1 - \log(c)$ . There is a constant  $K$  so that if  $1 \leq k \leq n^{0.7}$  and  $c \leq 1$  then  $\gamma_{n,k}(c) \leq K\lambda_{n,k}(c)$ .

*Proof.* Stirling's formula implies  $k! \sim k^k e^{-k} \sqrt{2\pi k}$ . Using this we have that

$$\gamma_{n,k}(c) \sim \frac{nk^{-5/2}}{c\sqrt{2\pi}} \left[ \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) \right] e^k c^k \left(1 - \frac{c}{n}\right)^{kn-k^2/2-3k/2+1}$$

Using the expansion  $\log(1-x) = -x - x^2/2 - x^3/3 - \dots$  we see that if  $k = o(n)$  then

$$\left(1 - \frac{c}{n}\right)^{kn-k^2/2-3k/2+1} \sim \exp(-ck + k^2/2n)$$

while if  $k = o(n^{3/4})$  we have

$$\begin{aligned} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) &= \exp\left(-\frac{1}{n} \sum_{j=1}^{k-1} j - \frac{1}{n^2} \sum_{j=1}^{k-1} j^2 + O\left(\frac{j^4}{n^3}\right)\right) \\ &\sim \exp\left(-\frac{k(k-1)}{2n} - \frac{k(k-1)(2k-1)}{6n^2}\right) \sim \exp\left(-\frac{k^2}{2n} - \frac{k^3}{3n^2}\right) \end{aligned}$$

Combining the last three formulas gives the asymptotic formula. To prove the bound we note that Stirling's formula implies  $k! \geq \delta k^k e^{-k} \sqrt{2\pi k}$  for some  $\delta > 0$ . Using the bounds  $\log(1-x) \leq -x$  and  $\log(1-x) \leq -x - x^2/2$  in the last two calculations gives the upper bound.  $\square$

**Lemma 7.**

$$E \left[ \frac{6}{\log n} \int_0^{c_n(r)n/2} \bar{\psi}_s^n ds \right] \rightarrow r$$

*Proof.* The upper bound follows from (2.9) which holds for all  $c < 1$ . In the other direction, we will use Fatou's lemma so it is enough to know the asymptotic behavior of the integrand. Changing variables  $s = c_n(v)n/2$  where  $c_n(v) = 1 - n^{-v/3}$  and noting  $c'_n(v) = (1/3)(\log n)n^{-v/3}$  gives

$$\begin{aligned} E \left[ \frac{6}{\log n} \bar{Z}^n(r) \right] &= n \int_0^r Q[\bar{\psi}_{c_n(v)n/2}^n] n^{-v/3} dv \\ &= \int_0^r \sum_{k=1}^{n^{0.7}} \frac{k-1}{n} Q_{p(c_n(v),n)}[kT_k] n^{-v/3} dv \end{aligned} \quad (2.14)$$

where  $T_k$  is the number of tree components of size  $k$ , and  $p(c,n) = 1 - \exp(-c/n)$ .

We can take the limit of the last expression by using formula (2.4), combined with Lemma 6. Indeed formula (2.4) shows that  $ET_k = \gamma_{n,k}(c)$ , and  $k \leq n^{0.7} = o(n^{3/4})$ , so that the use of Lemma 6 is justified. Hence

$$E \left[ \frac{6}{\log n} \bar{Z}^n(r) \right] = \int_0^r \sum_{k=1}^{n^{0.7}} \frac{k(k-1)}{n} \gamma_{n,k}(c_n(v)) n^{-v/3} dv$$

Setting  $c = 1 - b$  with  $b = n^{-v/3} \rightarrow 0$  and using Taylor's theorem

$$-(c-1-\log(c))k - b \frac{k^2}{n} = -\frac{b^2}{2}k - b \frac{k^2}{2n} + o(b^2k) \quad (2.15)$$

The first term becomes significantly negative when  $k \approx 1/b^2 = n^{2v/3}$ , the second when  $k \approx \sqrt{n/b} = n^{(1+v)/3}$ . Let  $2v/3 < w < (1+v)/3$ . The terms with  $k \geq n^w$  tend to 0 exponentially fast in  $n$  and there are less than  $n^{0.7}$  of them so their sum goes to 0. In the range  $k \leq n^w$  the second term can be ignored, so Lemma 6 implies that if  $v < 1$

$$\sum_{k=1}^{n^{0.7}} \frac{k^2}{n} Q_{p(c_n(v),n)}[T_k] \sim \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{n^w} k^{-1/2} \exp(-n^{-2v/3}k/2) \quad (2.16)$$

Here we have used the asymptotic formula of Lemma 6 for all  $k$ . However, the next computation will show that the sum grows like  $n^{v/3}$  so the contributions from small  $k$  can be ignored.

To recognize (2.16) as a Riemann sum with spacing  $n^{-2v/3}$ , we rewrite it as

$$n^{v/3} \sum_{k=1}^{n^{0.7}} n^{-2v/3} (n^{-2v/3} k)^{-1/2} \exp(-n^{-2v/3} k/2)$$

Since  $x^{-1/2}e^{-x}$  is decreasing it is straightforward to estimate the difference between the sum and the limiting integral and we conclude that

$$n^{-v/3} \sum_{k=1}^{n^{0.7}} \frac{k(k-1)}{n} Q_{p(c_n(v),n)}[T_k] \rightarrow \int_0^\infty \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2} dx$$

Changing variables  $x = y^2$ ,  $dx = 2y dy$  the integral becomes

$$(2\pi)^{-1/2} \int_0^\infty 2e^{-y^2/2} dy = 1$$

Therefore, by Fatou's lemma:

$$\liminf_{n \rightarrow \infty} E \left[ \frac{6}{\log n} \bar{Z}^n(r) \right] \geq \int_0^r n^{-v/3} \cdot n^{v/3} dv = r$$

□

We turn now to the analysis of the variance.

**Lemma 8.**  $\text{var} \left( \frac{6}{\log n} \bar{\kappa}_n(r) \right) \rightarrow 0$

*Proof.* Changing variables as in (2.14) and using Cauchy-Schwarz inequality as in the proof of the subcritical regime,

$$\begin{aligned} \text{var} \left( \frac{6}{\log n} \int_0^{c_n(r)n/2} \bar{\psi}_s^n ds \right) &= \text{var} \left( n \int_0^r \bar{\psi}_{c_n(v)n/2}^n n^{-2v/3} dv \right) \\ &\leq n^2 \int_0^r n^{-2v/3} \text{var} (\bar{\psi}_{c_n(v)n/2}^n) dv \\ &\leq \frac{2}{n} \int_0^r n^{-2v/3} Q_{p(c_n(v),n)}[|C_1|^3 I_1] dv \end{aligned}$$

Reasoning as in (2.16) but using the bound in Lemma 6

$$\sum_{k=1}^n k^3 Q_{p(c_n(v),n)}[kT_k] \leq K \sum_{k=1}^n k^{3/2} \exp(-n^{-2v/3}k/2)$$

To check the right-hand side note that the power of  $k$  has increased by 2, from the previous calculation. If we view the last sum as a Riemann sum with spacing  $n^{-2v/3}$ , we can rewrite it as

$$n^{5v/3} \sum_{k=1}^n n^{-2v/3} (n^{-2v/3}k)^{3/2} \exp(-n^{-2v/3}k/2)$$

Now  $x^{3/2}e^{-x/2}$  has derivative  $((3/2)x^{1/2} - (1/2)x^{3/2})e^{-x/2}$  so it is increasing on  $[0, 3]$  and then decreasing on  $[3, \infty)$ . Thus if we discard the term with the largest  $k$  so that  $n^{-2v/3}k \leq 3$  we have a lower bound on the integral.

$$n^{-2v/3} \sum_{k=1}^n k^3 Q_{p(c_n(v),n)}[kT_k] \leq n^v \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{3/2} e^{-x/2} dx + n^{v/3} 3^{3/2} e^{-3/2}$$

Using this it follows that

$$\text{var} \left( \frac{6}{\log n} \bar{\kappa}^n(r) \right) \leq \frac{K}{n} \int_0^r n^u du$$

Writing  $n^u = \exp(-u \log n)$  and integrating we have that the right-hand side is  $\leq K/(\log n) \rightarrow 0$ . This concludes the proof of the first result in Theorem 22.  $\square$

The final step is to estimate the number of fragmentations that occur to tree components of size  $\leq n^{0.7}$  at times between  $(1 - n^{-1/3})n/2$  and  $n/2$ :

$$\int_{(1-n^{-1/3})n/2}^{n/2} Q \bar{\psi}_s^n ds$$

For each  $s$  in the interval the integrand is smaller than  $\sum_{k=1}^{n^{0.7}} \frac{k^2}{n} Q_{1/n} T_k$ . Using Lemma 6, the last quantity is smaller than

$$\frac{K}{n} \sum_{k=1}^{\infty} k^{-1/2} \exp(-k^3/3n^2)$$

which we can rewrite as

$$\frac{K}{n} n^{1/3} \sum_{k=1}^{\infty} n^{-2/3} (kn^{-2/3})^{-1/2} \exp(-(kn^{-2/3})^3/2)$$

The above sum is a Riemann sum so it converges to  $\int_0^{\infty} x^{-1/2} e^{-x^3/2} dx$ . Therefore,  $Q\bar{\psi}_s^n \leq Kn^{-2/3}$ . Since the duration of the critical regime is  $n^{2/3}/2$ , the expected number of fragmentations is bounded and the proof of Theorem 22 is complete.

## 2.6 The supercritical regime

By Pittel's (1990) central limit theorem for the number of components of a supercritical random graph, it is enough to show that, with probability going to 1 as  $n \rightarrow \infty$ , at time  $cn/2$  there are fewer than  $o(n^{1/2})$  extra components due to fragmentation. (This was already indicated in the sketch of the proof of Theorem 24).

Let  $a = 0.55$ . (In fact the results stated in this section would also be valid for any  $1/2 < a < 2/3$  but making this choice makes some proofs slightly easier). We call cycles of size  $k \geq n^a$  large. These can be ignored since there cannot be more than  $n^{1-a} = o(n^{1/2})$  such components. We define the amount of mass "upstairs" by

$$N_t^{\uparrow} = \sum_{k > n^a} kX_k(t)$$

where  $X_k(t)$  is the number of cycles of size  $k$  at time  $n/2+t$ . (It is convenient in this section to shift the time so that  $t = 0$  corresponds to critical time  $n/2$ .) If all of the mass was upstairs, then the expected number of cycles of size less than  $n^a$  produced by fragmentation would be  $2n^{a-1}(cn/2) = O(n^a)$ . It is overly pessimistic to think that all of the mass will be upstairs, but by analogy with the random graph, we expect (and will eventually prove in Theorem 25) that at times  $c > 1$  a positive fraction of the total mass  $n$  will be there, so this estimate of the number of fragmentations is too large to ignore.

To improve this crude estimate, we take advantage of the fact that fragmented pieces are reabsorbed upstairs. Let  $X_k^{\downarrow}(t)$  be the number of cycles of size  $k$  produced by fragmentation of cycles upstairs.  $X_k^{\downarrow}(t)$  can only increase

when a transposition is performed, and only if it is made of one of the  $N_t^\uparrow$  vertices upstairs and of one of the 2 points located  $k$  steps away when writing the corresponding cycle of the current permutation. This gives a rate at most  $2N_t^\uparrow/n^2$ . As for the death rate, one way to get rid of a component of size  $k$  is by picking one of the  $k$  vertices of one of the  $X_k^\downarrow(t)$  components and one of the  $N_t^\uparrow$  vertices upstairs. This happens with rate  $2kX_k^\downarrow(t)N_t^\uparrow/n^2$ . For the moment we are ignoring the fact that cycles may experience coalescence or fragmentation while downstairs. We will deal with these complexities once we have an understanding of the basic birth and death process of fragments of large clusters.

### 2.6.1 The cluster queuing system

It is fortunate that the unknown quantity  $N_t^\uparrow \leq n$  appears in both rates, so that as long as  $N_t^\uparrow > 0$  we can remove it by time change. Once this is done, we have a system of stochastic processes  $\xi_t^k$ , for  $1 \leq k \leq n^a$  that we call a *cluster queuing system*: let  $\xi_t^k$  be independent birth-and-death chains with birth rate 1 and death rate  $k\xi_t^k$ , that begin with  $\xi_0^k = 0$ .

**Lemma 9.** *With probability  $\rightarrow 1$  as  $n \rightarrow \infty$  we have*

$$\sum_{k=1}^{n^a} \xi_t^k \leq (\log n)^2 \quad \text{and} \quad \sum_{k=1}^{n^a} k\xi_t^k \leq n^a(\log n)^2$$

for all  $t \leq c$  ( $c > 0$ ).

**Remark.** Although this system of stochastic processes can be defined without any reference to our random walk problem, it is useful to bear in mind that the state of this cluster queuing system at time  $t$  describes the number of fragments of large cycles at time

$$\frac{n}{2} + \int_0^t \frac{n^2}{2N_s^\uparrow} ds \geq \frac{n}{2}(1+t)$$

since  $N_s^\uparrow \leq n$ . Thus the control obtained in the above lemma for all  $t \leq c$ , will provide useful information for the random walk between times  $n/2$  and  $(1+c)n/2$  for any  $c > 0$ . On our original time-scale, this corresponds exactly to the supercritical regime, i.e. up to time  $cn/2$  for any  $c > 1$ .

*Proof.* The second result is a trivial consequence of the first. The key idea to handle the processes  $\xi_t^k$  is to consider strips  $2^j \leq k < 2^{j+1}$ . Because there are no simultaneous jumps, we can prove that the queues  $\xi_t^k$  at each level  $k$  are independent processes (see e.g. Revuz-Yor (1999), chap. XII, prop. (1.7), for a proof of this fact in the case of Poisson processes). Therefore, for each  $1 \leq j \leq \log_2 n^a$ , the number of cycles with sizes in  $[2^j, 2^{j+1})$ ,  $\zeta_t^j$ , is dominated by a birth and death chain with birth and death rates respectively  $2^j$  and  $2^j \zeta_t^j$ . To analyze these processes, we consider the successive excursions away from 0. Their embedded discrete time processes  $Y_s$  jump from  $m$  to  $m-1$  with probability  $m/(m+1)$  and from  $m$  to  $m+1$  with probability  $1/(m+1)$ . Let us try to find a function  $\phi$  such that  $\phi(0) = 0$ ,  $\phi(1) = 1$  and  $\phi(Y_s)$  is a martingale. The latter implies

$$\frac{1}{m+1}[\phi(m+1) - \phi(m)] = \frac{m}{m+1}[\phi(m) - \phi(m-1)]$$

so  $\phi(x) = \sum_{k=1}^x (k-1)!$ . Since  $\phi(1) = 1$  and  $\phi(0) = 0$ , it follows by optional sampling that the maximum level reached during an excursion of  $\zeta^j$ ,  $M$ , satisfies

$$P(M > x) = 1/\phi(x+1) \leq 1/x! \quad (2.17)$$

To bound the number of excursions for the process in the  $j^{\text{th}}$  strip before time  $c$ ,  $N_j(c)$ , we note that jumps from 0 to 1 occur at rate  $2^j$  so ignoring the amount of time it takes to return to 0 from 1, the number of excursions by time  $c$  is bounded by a Poisson random variable with mean  $2^j c \leq cn^a$ . Markov's inequality implies that  $P(N_j(c) > n^2) \leq cn^{a-2}$  so

$$P\left(\max_{1 \leq j \leq a \log_2 n} N_j(c) > n^2\right) \rightarrow 0 \quad (2.18)$$

To estimate the probability that the maximum of  $n^2$  excursions is  $> \log n$  we recall (2.17) and that Stirling's formula implies  $k! \geq \delta_0 k^k e^{-k} / \sqrt{2\pi k}$  for some  $\delta_0 > 0$ , so

$$(\log n)! \geq \delta_1 (\log n)^{\log n} n^{-1} (\log n)^{-1/2} = \delta_1 n^{\log \log n - 1} (\log n)^{-1/2}$$

The right-hand side goes to  $\infty$  faster than  $n^2 \log_2 n$  so using (2.18) we have

$$P\left(\max_{1 \leq j \leq a \log_2 n} \max_{0 \leq t \leq c} \zeta_t^j > \log n\right) \rightarrow 0$$

When the last event does not occur we have

$$\sum_{k=1}^{n^a} \xi_t^k \leq a(\log_2 n) \log n = \frac{a}{\log 2} (\log n)^2$$

Since  $a < 2/3 < \log 2 \approx 0.69$ , this gives the desired result.  $\square$

## 2.6.2 Completion of the proof of Theorem 24

The cluster queuing system is the first approximation to the analysis of the dynamics of the supercritical regime. However, it ignores customer fragmentation and a number of “bad events” that we need to consider in order to give a rigorous proof of Theorem 24. Though *a priori* one might expect it to be difficult to take account of corrections of second order, third order, . . . , and have nightmares about adding up infinitely many terms, we were pleasantly surprised to see that the proof could be completed with a few simple estimates.

The first technical problem to confront is to show that the total amount of mass upstairs stays positive at any given time so we can apply our time change. This is done in section 2.6.3.

The more difficult problem is to control the difference between the CQS and the real system of clusters. To do this, we need a notational scheme to verify that we have indeed taken care of all of the relevant events. We call clusters of size larger than  $n^a$  *large*, those in the CQS (i.e., those that were generated by a fragmentation of some large cycle), *medium*, and non-giant clusters in the random graph *small*. Writing *frag* and *coag* as shorthand for fragmentation and coagulation, we have three *frag* and six *coag* events to handle:

*coag(small,small)* is a natural part of the random graph so these events are not errors. The fragmentation of small clusters involves  $o(n^{1/2})$  clusters and hence does not significantly alter this process (see *frag(small)* and Lemma 12).

*coag(small,large)* eliminates a small component, but in the random graph these correspond to the small cluster being absorbed into the giant component, so this is not an error.

*frag(small)* is easy to take care of due to the duality principle which asserts that finite clusters in the random graph at time  $c > 1$  have the same distribution as clusters at time  $c\rho < 1$  where  $\rho$  is the probability of no percolation.

This allows us to use our subcritical estimates for fragmentation of small supercritical clusters. More details are given in Lemma 12.

*coag(large, large)* We do not care about these events since we do not need to keep track of the number of cycles upstairs.

*frag(large)* These are the arrivals in the cluster queuing system

*coag(medium, large)* are (almost) the departures in the cluster queuing system. The problem is that the next three events can cause clusters to gain weight or split into two.

*coag(medium, medium)* are helpful events since they reduce the number of customers in the CQS. This does make the fragmentation rate for the new cluster larger than the sum of the two previous clusters but Lemma 11 will take care of this. More importantly, it makes the departure rate of the new cluster larger. This, applied to *coag(medium, medium)* and *coag(medium, small)*, shows that the number of medium clusters is stochastically bounded by the CQS of section 2.6.1, and is the content of Lemma 10.

*coag(medium, small)* eliminates a small component, but in the random graph these correspond to the small cluster being absorbed into the giant component. Again, this also makes the fragmentation rate larger for the cluster that gained weight but Lemma 11 will take care of this.

*frag(medium)* is taken care of by Lemma 11.

To complete the proof it remains to prove the three promised lemmas.

**Lemma 10.** *The number of medium clusters is dominated by that of the CQS. Therefore there are never more than  $(\log n)^2$  medium clusters, and never more than  $n^a(\log n)^2$  vertices in medium clusters.*

*Proof.* As was just mentioned, the only differences between the CQS and the medium clusters are generated by events of type *coag(medium, medium)* and *coag(small, medium)*. However both those events do not increase the number of medium clusters, and both those events make the death rate of the clusters concerned higher. Hence we can construct the CQS and the medium clusters process on the same probability space, in such a way that the *total* number of medium clusters is smaller than that of the CQS.  $\square$

**Lemma 11.** *The expected number of fragmentations of medium clusters is at most  $O(n^{2a-1}(\log n)^2)$ .*

*Proof.* There are never more than  $(\log n)^2$  medium clusters. Since there are at most  $n^a$  vertices per medium clusters the total number of vertices is at most  $n^a(\log n)^2$ . The rate at which those fragmentations happen is thus bounded by

$$\left(\frac{n^a(\log n)^2}{n}\right) \frac{n^a}{n}$$

so that the expected number of such fragmentations is indeed  $O(n^{2a-1}(\log n)^2)$ .  $\square$

**Lemma 12.** *The number of fragmentations of small components is  $o(n^{1/2})$ .*

*Proof.* By a now familiar estimate, the expected number of fragmentations that produce clusters of size smaller than  $n^p$  at times between  $n$  and  $n+t$  is at most  $2n^{p-1}t$ . So we can ignore fragmentations that (a) produce clusters of size smaller than  $n^{0.45}$  before time  $cn/2$  and (b) produce clusters of size smaller than  $n^{0.55}$  at times between  $n$  and  $n+n^{0.9}$ .

If  $c > 1$  the distribution of nongiant components in the random graph is given by progeny of a Poisson Galton Watson process with mean  $c$  on the event of its extinction. If we let  $\rho$  denote its extinction probability, then the offspring distribution conditional on extinction is given by

$$\frac{1}{\rho} e^{-c} \frac{(c\rho)^k}{k!} = e^{-c\rho} \frac{(c\rho)^k}{k!}$$

since  $\rho = e^{-c(1-\rho)}$ . In short,  $PGW(c)$  conditioned on extinction is  $PGW(c\rho)$ . The last observation implies that results for finite supercritical clusters can be derived from those for subcritical clusters. In particular, by Lemma 3, the largest nongiant components seen after time  $n+n^{0.9}$ , are smaller than  $n^{0.2}$ . Since fragmentations of such clusters necessarily produce pieces smaller than  $n^{0.2}$  these fragmentations can be ignored by (a).  $\square$

### 2.6.3 The initial mass upstairs

The last step in the proof of Theorem 24 is to ensure that upstairs never becomes empty in this process. In other words we must prove that  $N_t^\uparrow > 0$  for all  $t > 0$  with high probability, so that we can indeed time-change the queues by  $(N_t^\uparrow)^{-1}$ , and use rigorously all the analysis carried out on (CQS) in section 2.6.1. This will be done by showing that initially there are already more vertices upstairs than will ever (with high probability) be taken away by fragmentation in the cluster queuing system.

**Lemma 13.** *Initially, upstairs contains at least  $N_0^\uparrow \geq Kn^{1-a/2}$  vertices. In particular  $N_0^\uparrow > n^a(\log n)^2$  and it never becomes empty during the supercritical regime.*

*Proof.* Lemma 6 implies that when  $c = 1$  the expected number of trees of size  $k$

$$ET_k \sim \frac{nk^{-5/2}}{\sqrt{2\pi}} \exp(-k^3/3n^2)$$

If we let  $|\mathcal{C}_{\geq a}| = \sum_{k=n^a}^{\infty} T_k$  then it follows that

$$E|\mathcal{C}_{\geq a}| \sim \frac{n}{\sqrt{2\pi}} \sum_{k=n^a}^{\infty} k^{-5/2} \sim \frac{2}{3\sqrt{2\pi}} n^{1-3a/2}$$

Bollobás (1985) has calculated (see page 107) that the expected number of ordered pairs of trees of sizes  $j$  and  $k$ ,

$$E(T_j, T_k) \leq ET_j ET_k$$

When  $j \neq k$  this implies  $\text{cov}(T_j, T_k) \leq 0$  and for  $j = k$  that  $ET_k(T_k - 1) \leq (ET_k)^2$  or  $\text{var}(T_k) \leq ET_k$ . Summing we have

$$\text{var}(|\mathcal{C}_{\geq a}|) \leq E|\mathcal{C}_{\geq a}|$$

and it follows from Chebyshev's inequality that  $|\mathcal{C}_{\geq a}|/E|\mathcal{C}_{\geq a}| \rightarrow 1$  in probability. These trees have not experienced fragmentation so their size is always at least  $n^a$  and the total mass in large components is at least  $Kn^{1-a/2}$ . When  $a < 2/3$  and  $n$  is large, this is much larger than the  $n^a(\log n)^2$  upper bound on the missing mass due to fragmentations.

At this point the proof of Theorem 24 is complete. □

## 2.6.4 A sharper estimate for the mass upstairs

In section 2.6.3 above, we have just proved that upstairs never becomes empty in the supercritical regime (Lemma 13). But, as was already mentioned earlier, we expect by analogy with the random graph that in fact a positive fraction of all  $n$  vertices stay upstairs. This is the content of Theorem 25, which we restate here for convenience and then prove.

**Theorem 25** *For any  $c > 1$ , at time  $cn/2$  there are at least  $\theta(c)n - o(n)$  vertices located on large cycles (i.e., of size greater than or equal to  $n^a$ , for any  $a < 2/3$ ).*

*Proof.* In fact it is a simple consequence of Lemmas 10 and 11. Indeed, the mass missing upstairs must be a piece of the random graph's giant component fallen downstairs by fragmentation. Therefore either it is a medium cluster or it has experienced a consecutive fragmentation. But we now know that there are never more than  $n^a(\log n)^2$  vertices in medium clusters by Lemma 10. On the other hand, by Lemma 11, the expected number of vertices in clusters having experienced multiple fragmentation has to be smaller than

$$n^a \cdot Kn^{2a-1}(\log n)^2 = o(n)$$

as long as  $a < 2/3$ .

□

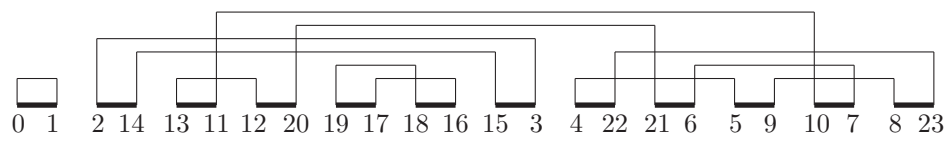


Figure 2.1: Breakpoint graph for human-mouse X chromosome comparison

36	37	17	40	16	15	14	63	10	9
55	28	13	51	22	79	39	70	66	5
6	7	35	64	33	32	60	61	18	65
62	12	1	11	23	20	4	52	68	29
48	3	21	53	8	43	72	58	57	56
19	49	34	59	30	77	31	67	44	2
27	38	50	26	25	76	69	41	24	75
71	78	73	47	54	45	74	42	46	

Table 2.1: Order of the genes in *D. repleta* compared to their order in *D. melanogaster*

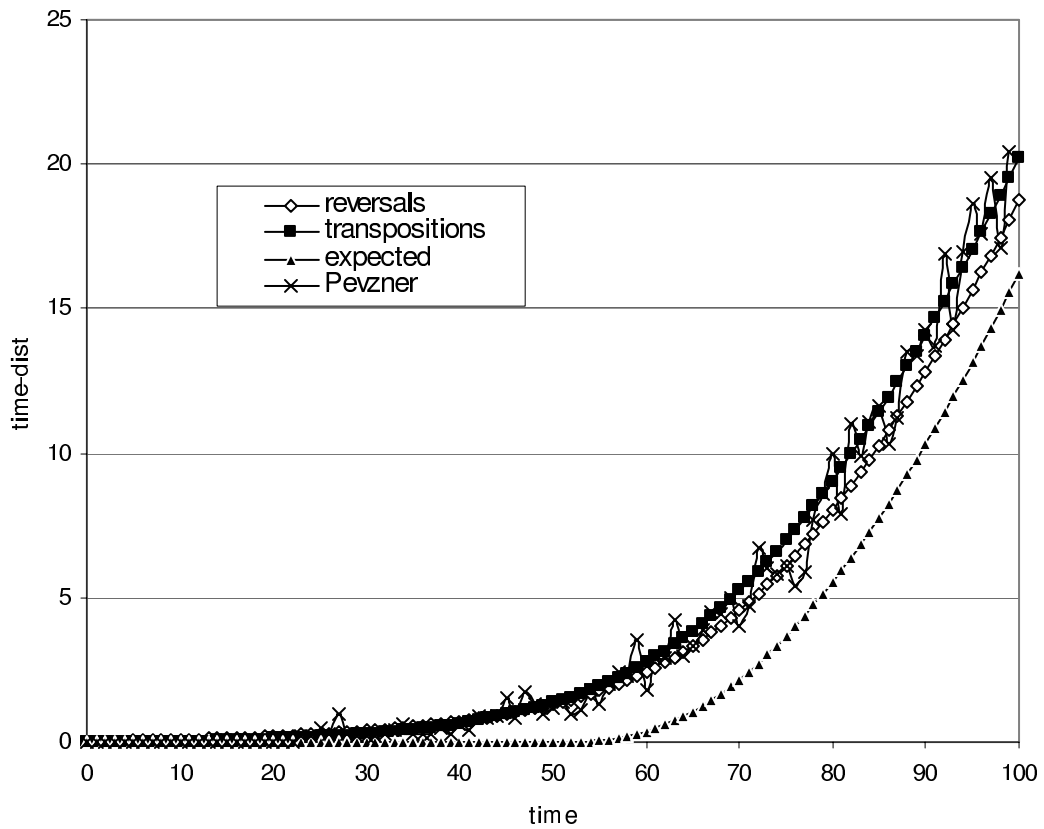


Figure 2.2: Simulations with  $n = 100$  markers. Average values of  $k - D_k$  for 10,000 simulations of the random transposition chain, 10,000 simulations of  $k - d_0(\pi_k)$  and Pevzner's 100 simulations of  $k - d(\pi_k)$  for the reversal chain. The smooth curve of small triangles gives the limiting behavior as  $n \rightarrow \infty$  of  $(cn/2 - D_{cn/2})$  from Theorem 3.

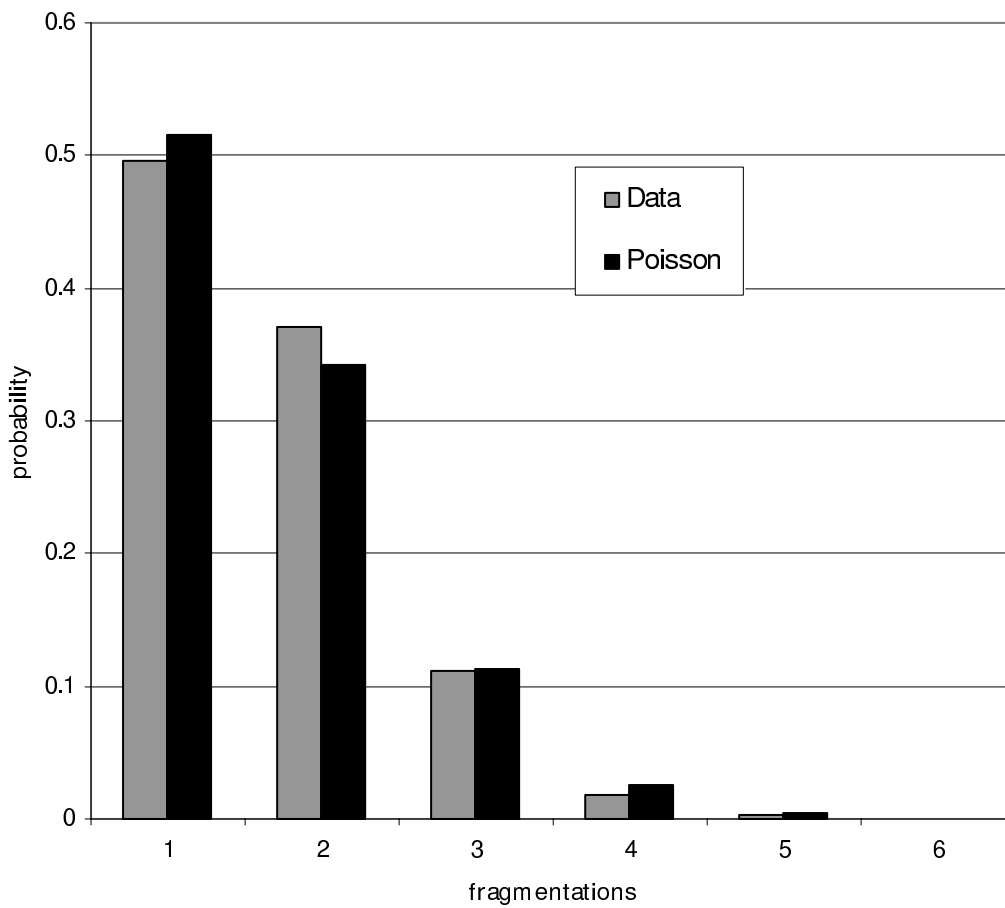


Figure 2.3: Comparison of the distribution of the number of fragmentations in 10,000 simulations of the random transposition chain with the Poisson distribution with the same mean.

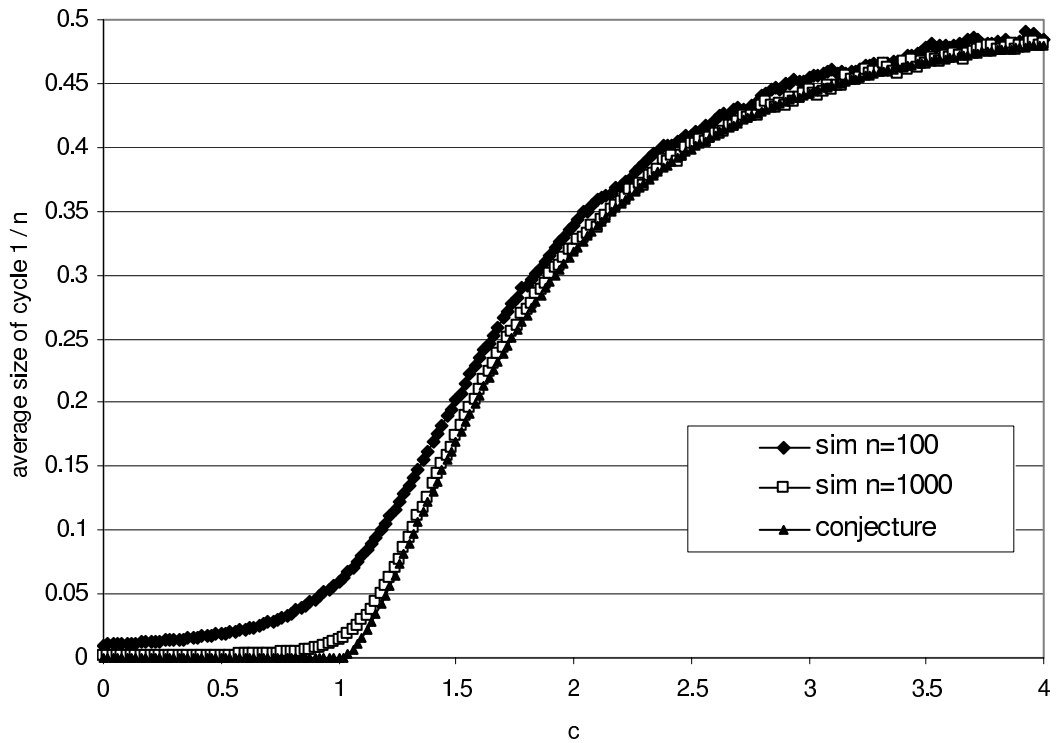


Figure 2.4: Growth of the average fraction of vertices in the cycle containing 1 at time  $cn/2$  in 10,000 simulations of the random transposition chain with  $n = 100$  and  $n = 1000$  compared to  $\theta(c)^2/2$  ( $\theta(c)$  being the percolation probability of the corresponding random graph with  $p = c/n$ ).

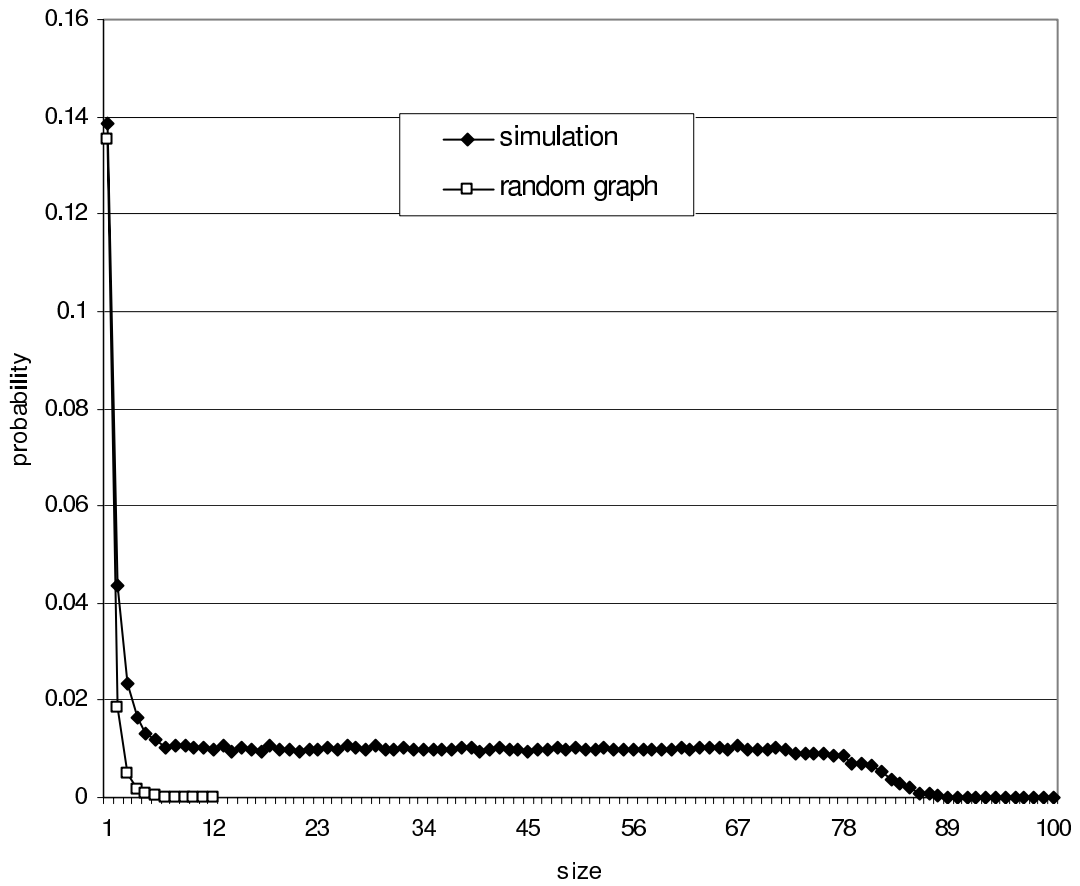


Figure 2.5: Histogram of the size of the cycle containing 1 in 100,000 simulations of the random transposition chain with  $n = 100$  at time 100 ( $c = 2$ ). The open squares give the distribution of the size of finite clusters in the corresponding random graph.

# Chapter 3

## The hyperbolic geometry of random transpositions

by Nathanaël Berestycki

### 3.1 Introduction

Let  $\mathcal{S}_n$  be the set of permutations of  $\{1, 2, \dots, n\}$ , and let  $\sigma_t$  be the continuous time random walk on  $\mathcal{S}_n$  that results when randomly chosen transpositions are performed at rate 1. Let  $d(\sigma_t)$  be the distance from the identity  $I$  at time  $t$ , i.e., the minimum number of transpositions needed to return to  $I$ . In a previous paper, Berestycki and Durrett (2004) showed

**Theorem 0.** *As  $n \rightarrow \infty$ ,  $d(\sigma_{nc/2})/n \rightarrow u(c)$  where*

$$u(c) = 1 - \sum_{k=1}^{\infty} \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k \quad (3.1)$$

Although it is not easy to see from the formula, the function  $u(c) = c/2$  for  $c \leq 1$  and is  $< c/2$  for  $c > 1$ .

We can think of  $\sigma_t$  as a random walk on the graph  $G_n$  with vertices  $\mathcal{S}_n$  and edges connecting two permutations that differ by one transposition, so that  $G_n$  is the Cayley graph of  $\mathcal{S}_n$  associated with the set of generators  $S = \{\text{all transpositions}\}$ . Theorem 0 was proved by establishing a connection with

Erdős-Renyi random graphs. The phase transition observed for  $\sigma_t$  is then related to the well-known double jump of the size of connected components of  $G(n, c/n)$  at  $c = 1$ .

In this paper, we try to investigate some of the geometric implications of Theorem 0. We find a new connection between the speed of a random walk and the Gromov hyperbolicity of the space in which the random walk is evolving.

**Organization of the paper.** In sections 1.1, 1.2, 1.3 we present our results. The proofs of these results can be found successively in section 2–8. Each proof is preceded by a restatement of the corresponding theorem for convenience, and by an informal proof which outlines the main ideas used.

### 3.1.1 Asymptotic hyperbolicity

The notion of hyperbolicity for a discrete structure such as a group is a notion that goes back to Gromov (1987). As there is no derivative, and thus no curvature available in a discrete space, the idea is to define what hyperbolic means using only elementary properties of the space.

One way to do is as follows. Let  $(X, |\cdot|)$  be a metric space, where  $|x - y|$  denotes the distance between  $x$  and  $y$ . For points  $x, y$  and  $p$  in  $X$ , define the Gromov inner product by

$$2(x|y)_p = |x - p| + |y - p| - |x - y|$$

$(x|y)_p$  thus measures how well the union of the geodesic segments  $[p, x] \cup [p, y]$  approximates a geodesic between  $x$  and  $y$ . Gromov's original definition of hyperbolic spaces is as follows. Call  $X$   $\delta$ -hyperbolic if

$$(x|z)_p \geq (x|y)_p \wedge (y|z)_p - \delta \tag{3.2}$$

for all  $x, y, z$  and  $p$ . This definition is not very intuitive at first, but fortunately there is an equivalent definition, which can be formulated using the notion of  $\delta$ -thin triangle. A triangle  $(x, y, z)$  with geodesic sides  $s_1, s_2, s_3$  is said to be  $\delta$ -thin if any side, say  $s_1$ , lies entirely within distance at most  $\delta$  of the two remaining sides:

$$s_1 \subset \{x \in X, d(x, s_2 \cup s_3) \leq \delta\}$$

The space is called  $\delta$ -hyperbolic if all geodesic triangles are  $\delta$ -thin, and it is simply called hyperbolic if it is  $\delta$ -hyperbolic for some  $\delta \geq 0$  (when  $\delta = 0$ , the space isometrically embeds into a tree). It is not immediate, but not hard to check, that if *all* triangles  $(x, y, z)$  are  $\delta$ -thin, then (3.2) is satisfied for some number  $\delta'$  that may differ by a constant factor from  $\delta$ . Conversely, in a space where (3.2) is satisfied for *all* points  $(p, x, y, z)$ , all triangles are  $\delta'$ -thin, where  $\delta'$  may differ from  $\delta$  by a constant factor.

Of course a bounded space (in particular, a finite space such as  $\mathcal{S}_n$ ) is trivially hyperbolic, but we will be interested in situations where the constant  $\delta$  may or may not stay bounded as the size of the space tends to  $\infty$ .

Our first result makes the connection between Theorem 0 and Gromov hyperbolic spaces, where we look at the two definitions of hyperbolic constants suitably weakened. For  $0 < a < 1$ , let  $\partial B(an)$  be the sphere of radius  $an$ , i.e. the set of points at distance  $\lfloor an \rfloor$  from the origin. We let  $\nu$  be the hitting distribution of  $\partial B(an)$  by  $\sigma_t$ , i.e.  $\nu$  is the law on  $\partial B(an)$  of  $\sigma_T$  where  $T = \inf\{t > 0, d(\sigma_t) = \lfloor an \rfloor\}$ .

**Theorem 26.** *Let  $x, y$  be sampled from  $\nu$  independently, and set  $p = I$ , the identity element.*

1. *If  $a < 1/4$ , then there is some  $\delta < \infty$  (depending only on  $a$ ), such that*

$$E(x|y)_p \leq \delta$$

*Moreover, with probability asymptotically 1, there is a geodesic between  $x$  and  $y$  that comes within expected distance  $\delta' < \infty$  of  $p$ .*

2. *If  $a > 1/4$ , then*

$$E(x|y)_p \sim \delta n$$

*for some  $0 < \delta < \infty$ . Moreover, no geodesic between  $x$  and  $y$  can approach  $p$  closer than  $\delta'n$  with probability asymptotically 1, where  $0 < \delta' < \infty$*

In the statement of the theorem and in the rest of the paper,  $a_n \sim b_n$  means that  $a_n/b_n \rightarrow 1$ .

**Remark.** It follows immediately from Theorem 1 that when  $a < 1/4$ , with probability asymptotically 1

$$(x|z)_p \geq (x|y)_p \wedge (y|z)_p - \delta$$

for independent  $x, y, z$  sampled from  $\nu$ , hence the idea that definition 1 of hyperbolicity is satisfied "asymptotically  $\nu$ -almost surely". The statement about the geodesics shows that definition 2 is satisfied "asymptotically  $\nu$ -almost surely" when  $a < 1/4$ .

At this point we should emphasize that the result in Theorem 26 involves hyperbolic constants that are different from the standard definitions discussed above in several important ways. The most obvious difference comes from the randomness of  $x$  and  $y$ , and from the fact that the role played by  $x, y$  and  $p$  are somewhat different. Here  $p$  is a fixed reference point, whereas Gromov's definition requires that every triangle should be thin. Another issue is that, corresponding to the second definition of hyperbolicity with thin triangles, we show that there exists a certain geodesic between  $x$  and  $y$  having the desired properties. As we will see below in Theorem 31, there may be a great many geodesics between two given points in  $\mathcal{S}_n$ . More importantly, these geodesics can be far apart, as will show the following concrete example:

$$\begin{array}{ll} \sigma & (1\ 145\ 11)(2)(39)(4\ 136)(7\ 128)(10) \\ \pi_1 & (1)(14)(5)(11)(2)(3)(9)(4\ 136)(7\ 128)(10) \\ \pi_2 & (1\ 145\ 11)(2)(39)(4)(13)(6)(7)(12)(8)(10) \\ \pi_1\pi_2^{-1} & (11\ 5\ 14\ 1)(2)(93)(4\ 136)(7\ 128)(10) \end{array}$$

Since for any permutation  $\pi$  we have  $d(\pi) = n - \#$  cycles of  $\pi$ ,  $d(\sigma) = 8$ .  $\pi_1$  and  $\pi_2$  are on two geodesics from  $I$  to  $\sigma$ , but  $d(\pi_1, \pi_2) = d(\pi_1\pi_2^{-1}) = 8$ . In general if  $d(\sigma) = cn/2$  with  $c < 1$ , and we divide the cycles at random into two groups, we can define  $\pi_1$  to have cycle structure given by the first group of  $\sigma$  staying as it is and the second completely broken in cycles on lengths 1. If we define  $\pi_2$  by the exchanging the two groups, then we will have  $d(\sigma, \pi_i) = cn/4$  and  $d(\pi_1, \pi_2) = cn/2$ .

### 3.1.2 The geometry of $G_n$

How much can we learn from Theorem 26 about the global geometry of  $G_n$ ? To answer this question, we need to see how special a choice it is to sample the points  $x$  and  $y$  according to the hitting distribution  $\nu$ . (The fact that  $p = I$  is a fixed reference point is not too important, due to the transitivity of  $G_n$ ). We begin by an apparently unrelated question, which is to ask how large is a ball of radius  $an$ .

**Theorem 27.** *If  $0 \leq a \leq 1$  then as  $n \rightarrow \infty$ , we have  $|B(I, an)| \approx (n!)^a$  in a logarithmic sense, i.e.,*

$$\lim_{n \rightarrow \infty} \frac{\log |B(I, an)|}{n \log n} = a$$

This result is probably not new, but we have not found it in the literature. Our original motivation for studying the volume growth in  $G_n$  was to try to understand the phase transition of Theorem 0 in terms of the geometry of  $G_n$ . Our first thought was that since the speed was non-smooth we might see a change in the volume growth. The above result contradicts this idea.

To put our next two results into perspective it is useful to contrast them with Brownian motion  $B_t$  on a  $d$ -dimensional manifold of constant negative curvature  $-1$ . In that case as  $t \rightarrow \infty$ , if  $d(B_t)$  is the distance from the origin then (see Prat (1971) for instance) there is a constant  $v$  so that

$$d(B_t)/t \rightarrow v \quad \text{as } t \rightarrow \infty$$

In the case of Brownian motion on hyperbolic space, rotational symmetry implies that the hitting distribution is uniform. In contrast for the random transposition random walk, we will see in Theorem 28 that the hitting distribution is asymptotically singular with respect to the uniform distribution on  $\partial B(I, an)$ .

**Theorem 28.** *Let  $|\mathcal{C}_1|$  be the length of the cycle that contains 1. Under  $\mu$ , the uniform distribution on  $\partial B(I, an)$ ,*

$$|\mathcal{C}_1| \Rightarrow \mathbf{G}$$

*where  $\mathbf{G}$  is a geometric r.v. with  $P(\mathbf{G} > k) = (b/(1+b))^k$  and  $b$  satisfies  $\log(1+b)/b = 1-a$ .*

To describe the hitting distribution  $\nu$ , we note that (3.1) suggests that it will be the same as the distribution of  $\sigma_{cn/2}$  where  $c = u^{-1}(a)$ . When  $a > 1/2$  this is much different from the distribution in Theorem 28 since in this case  $c > 1$  and Schramm (2004) has shown that  $\sigma_{cn/2}$  has cycles of lengths of order  $n$ .

Here we will concentrate on what happens when  $a < 1/2$  and  $c = 2a$ . In this case results in Berestycki and Durrett (2004) show that as  $n \rightarrow \infty$ ,

the number of fragmentations before time  $cn/2$  is asymptotically a Poisson random variable with mean  $\kappa(c) = -(\log(1-c) + c)/2$ . In particular,

$$P(d(\sigma_{cn/2}) = cn/2) \rightarrow e^{-\kappa(c)} = e^{c/2}\sqrt{1-c}$$

It will be convenient to approach the hitting distribution  $\nu$  by the distribution  $\nu_0$  of  $\sigma_{cn/2}$  conditioned on no fragmentation. More generally, if  $\nu_k = \nu$  conditioned on exactly  $k$  fragmentations before the hitting time,

$$\nu = e^{-\kappa(c)} \sum_{k=0}^{\infty} \nu_k \frac{\kappa(c)^k}{k!} + o(1)$$

To study  $\nu_0$ , we recall the connection with random graphs developed in Berestycki and Durrett (2004): when we transpose  $i$  and  $j$  we draw an edge between  $i$  and  $j$ . In order for the distance from the identity to increase by one at each time, each transposition must involve indices from two different cycles and will merge them into 1. In terms of the random graph, this means that all components are trees. Using results from Berestycki and Durrett (2004), it is straightforward to show

**Theorem 29.** *Let  $\mathcal{C}_1$  be the length of the cycle that contains 1. Let  $c < 1$ . Under  $\nu_0$ ,*

$$P(|\mathcal{C}_1| = k) \rightarrow \frac{1}{c} \frac{k^{k-1}}{k!} (ce^{-c})^k \quad \text{for all } k \geq 1$$

Theorems 28 and 29 show that the uniform distribution  $\mu$  and the hitting distribution  $\nu_0$  concentrate on different permutations. In the first case the number of fixed points will be close to its expected value  $nP(|\mathcal{C}_1| = 1) = n/(1+c)$ . In the second it will be close to  $ne^{-c}$  by Theorem 29. This is made precise by the following theorem.

**Theorem 30.** *As  $n \rightarrow \infty$ , the hitting distribution  $\nu$  and the uniform distribution  $\mu$  on a sphere of radius  $n$  are asymptotically singular:*

$$d_{TV}(\mu, \nu) \rightarrow 1$$

Let  $t = [cn/2]$  with  $c < 1$ . To understand why  $\nu$  is different from  $\mu$  we will examine the Radon-Nikodym derivative  $r(\sigma) = d\nu_0/d\mu$ . It is not hard to show that

**Theorem 31.** *Suppose  $d(\sigma) = t$  and  $m_1, \dots, m_j$  are the cycle lengths of  $\sigma$ . The number of paths of length  $t$  from  $I$  to  $\sigma$  is*

$$t! \prod_{i=1}^j \frac{m_i^{m_i-2}}{(m_i-1)!}$$

If  $t = \lfloor cn/2 \rfloor$  with  $c < 1$  then

$$r(\sigma) = K_{n,t} \prod_{i=1}^j \frac{m_i^{m_i-2}}{(m_i-1)!}$$

where  $K_{n,t}$  is a constant that only depends on  $n$  and  $t$ .

The last result enables us to prove a stronger version of Theorem 30 : it tells us that the "support" of  $\nu$  is concentrated on a set that is exponentially smaller than the size of  $\partial B(an)$ .

**Theorem 32.** *Suppose  $a < 1/2$ . There exists a set  $S_n \in \partial B(an)$  such that  $\nu(S_n) \rightarrow 1$  as  $n \rightarrow \infty$  and*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{|S_n|}{|\partial B(an)|} = \gamma < 0$$

### 3.1.3 The hyperbolic constant under the uniform measure

In Theorem 26, we learn that if  $x$  and  $y$  are sampled from  $\nu$ , roughly speaking, the Gromov hyperbolicity of the "support" breaks down at  $a = 1/4$ , i.e. the hyperbolic constant increases suddenly from  $O(1)$  to  $O(n)$  at this point.

However, the results from the previous section tell us that this "support" is (exponentially) small with respect to the ambient space. It is therefore natural to ask what happens to Theorem 26 when we replace  $\nu$  with the uniform measure  $\mu$  on  $\partial B(an)$ . Theorem 33 will show that the qualitative behavior of the hyperbolic constant remains the same. We prove that there is a threshold where the expected Gromov inner product  $E(\sigma|\pi)_p$  jumps from  $O(1)$  to  $O(n)$ , but this time the critical value is  $a = 1 - \log 2 \approx 0.31$ , rather than  $a = 1/4$ .

When  $\sigma$  and  $\pi$  are independent uniform permutations on  $\partial B(an)$ , by the transitivity of  $G_n$ , it is enough to analyze  $d(\sigma, \pi)$  to understand  $(\sigma|\pi)_p$ , the inner Gromov product. Since  $d(\sigma, \pi) = d(I, \sigma^{-1}\pi)$ , which has the same law as  $d(I, \sigma\pi)$ , it will be enough to characterize the values of  $a$  for which  $d(I, \sigma\pi) = 2an + o(n)$  and those for which it is  $< 2an$ .

**Theorem 33.** *Let  $0 < a < 1$  and let  $\sigma, \pi$  be two random independent points chosen uniformly from  $\partial B(an)$ . Then*

1. *If  $a < 1 - \log 2$*

$$E(\sigma|\pi)_p \leq \delta(\log n)^2$$

*for some  $0 < \delta = \delta(a) < \infty$ . Moreover, with probability asymptotically 1, there is a geodesic between  $\sigma$  and  $\pi$  that comes within distance at most  $\delta(\log n)^2$  of  $p$ .*

2. *If  $a > 1 - \log 2$*

$$E(\sigma|\pi)_p \sim \delta n$$

*for some  $\delta = \delta(a) > 0$ . Moreover, no geodesic can approach  $p$  closer than  $\delta'n$  for some  $0 < \delta' < \infty$ .*

**Remark.** The  $O((\log n)^2)$  bound in part 1 of the Theorem could probably be improved into a  $O(1)$  bound (just like in Theorem 26) with some more work, but we have not tried to do so. In part 2, by analogy with Berestycki and Durrett (2004), we conjecture that the fluctuations are of order exactly  $n^{1/2}$  in the supercritical regime. More precisely it should be true that when  $a > 1 - \log 2$

$$n^{-1/2}(E(\sigma|\pi)_p - \delta n) \Rightarrow \mathcal{N}(0, \kappa)$$

where  $\delta$  is the limit in part 2 of the theorem, and  $\kappa$  is some parameter.

## 3.2 Asymptotic hyperbolicity under $\nu$

The first result we prove is Theorem 26.

**Theorem 26** *Let  $x, y$  be sampled from  $\nu$  independently, and set  $p = I$ , the identity element.*

1. If  $a < 1/4$ , then there is some  $\delta < \infty$  (depending only on  $a$ ), such that

$$E(x|y)_p \leq \delta$$

Moreover, with probability asymptotically 1, there is a geodesic between  $x$  and  $y$  that comes within expected distance  $\delta' < \infty$  of  $p$ .

2. If  $a > 1/4$ , then

$$E(x|y)_p \sim \delta n$$

for some  $0 < \delta < \infty$ . Moreover, no geodesic between  $x$  and  $y$  can approach  $p$  closer than  $\delta'n$  with probability asymptotically 1, where  $0 < \delta' < \infty$

**Sketch of the proof** of Theorem 26. Let  $X_t$  and  $Y_t$  be two independent random walks starting at the origin. Let them run until the times  $T$  and  $T'$  where they respectively hit the sphere  $\partial B(an)$ . Then the transitivity of the Cayley graph of  $\mathcal{S}_n$ , and the reversibility of the increments of the random walk, imply that  $(X_T, X_{T-1}, \dots, p, Y_1, \dots, Y_{T'})$  is a random walk path of length  $T + T'$ . Hence the distance between  $X_T = x$  and  $Y_{T'} = y$  is the same as  $d(\sigma_{T+T'})$ . By Theorem 0,  $T$  and  $T' \approx \frac{1}{2}u^{-1}(a)n$ , so applying Theorem 0 again, when  $a < 1/4$ ,  $|x - y| \approx 2an = |x| + |y|$  (the random walk runs for a time  $2an < n/2$  and there are only  $O(1)$  fragmentations). For  $a > 1/4$ , the random walk is run for time  $u^{-1}(a)n$  which, in view of Theorem 0, means that  $c = 2u^{-1}(a)$ , and  $|x - y| = nu(2u^{-1}(a)) \ll 2an$ .

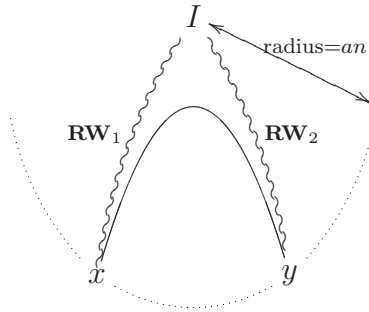


Figure 3.1: Two independent random walks run until they hit the sphere of radius  $an$ .

The claim about the existence of a geodesic that makes the triangle  $(x, p, y)$  thin involves necessarily another argument, since geodesics may be far apart.

However, it is not very hard to construct by hand a geodesic between the identity and  $x$  such that each point of the random walk path is within  $O(1)$  of this geodesic. Applying this construction to the two random walk paths gives the result of Theorem 26.  $\square$

## Proof of Theorem 26

Let us first deal with case  $a < 1/4$  and prove that in this case  $E(x|y)_p \leq \delta$ . Keeping the same notations as above, note that  $an + O(1)$  steps are sufficient for  $X$  to reach distance  $an$ . Indeed, after  $an$  steps,  $X_{an}$  is at distance  $an - X_1$  where  $X_1$  is twice a Poisson random variable by Theorem 1 in Berestycki and Durrett (2004). It is immediate that in  $X_1$  steps the probability that  $X$  has a fragmentation converges to 0. Therefore  $T - an \Rightarrow X_1$  (remember that here time is measured discretely). Similarly  $T' - an \Rightarrow X_2$  where  $X_1, X_2$  are i.i.d. Hence  $(x = X_t, X_{T-1}, \dots, X_1, I, Y_1, \dots, Y_t = y)$  is a random walk of  $2an + X_1 + X_2$  steps. In the worst case possible all  $X_1 + X_2$  steps represent "backward" steps (meaning, toward  $x$  rather than  $y$ ). Hence if  $X_3 = an - d(I, \sigma_{2an})$ , (so that  $X_3$  is also twice a Poisson random variable, but with a different parameter),

$$\begin{aligned} 2E(x|y)_p &= 2an - |x - y| \\ &\leq 2an - Ed(I, \sigma_{2an}) + E(X_1 + X_2) \\ &\leq E(X_1) + E(X_2) + E(X_3) < \infty \end{aligned}$$

It is slightly simpler to prove that when  $a > 1/4$ ,  $E(x|y)_p \sim \delta n$ . Indeed, in this case, by Theorem 0, we have that

$$\frac{1}{2}u^{-1}(a) - \varepsilon \leq T/n \leq \frac{1}{2}u^{-1}(a) + \varepsilon$$

Therefore

$$\inf_{|t/n - u^{-1}(a)| \leq 2\varepsilon} d(I, \sigma_t) \leq |x - y| \leq \sup_{|t/n - u^{-1}(a)| \leq 2\varepsilon} d(I, \sigma_t)$$

An easy estimate shows that we are never off by more than  $O(n^{1/2})$  if we evaluate the distance of the random walk by counting the number of clusters of the random graph rather than the number of cycles of  $\sigma_t$ . But for the

random graph, the number of clusters in monotone increasing. Hence, if  $\alpha$  denotes  $u(2u^{-1}(a))$ , we have by continuity of  $u$  that

$$\alpha - \varepsilon' + o(1) \leq \frac{E|x - y|}{n} \leq \alpha + \varepsilon' + o(1)$$

and  $\varepsilon'$  can be made as small as desired by continuity of  $u$ . Therefore

$$\frac{E|x - y|}{n} \rightarrow \alpha$$

It suffices now to prove that  $\alpha < 2a$ , i.e.  $u(2u^{-1}(a)) < 2a$  or, after change of variable  $c = u^{-1}(a)$ , it suffices to prove  $u(2c) < 2u(c)$  for all  $c > 1/2$ . This fact is a consequence of the sublinearity of  $u$ : it will be proved later that  $u$  is strictly concave on  $[1, \infty)$ , from which it follows that  $u(c) > u(2c)/2$ .

We now turn to the part of the theorem that concerns geodesics, and prove that for a random walk  $(X_t, t \leq cn/2)$  of time-duration  $cn/2$  with  $c < 1$ , there is a geodesic between  $\sigma = X_{cn/2}$  and  $I$ , that we call  $\gamma$ , such that

$$E \sup_{t \leq cn/2} d(X_t, \gamma) = O(1) \tag{3.3}$$

This shows that when  $c < 1$  there is a geodesic that stays close to the random walk path. When  $a < 1/4$ ,  $p = I$  is on the random walk path that leads from  $x$  to  $y$ , so this shows that  $E(d(I, \gamma)) = O(1)$ , as claimed in the Theorem. The case  $a > 1/4$  is trivial by the triangle inequality.

Let  $\tau_1, \dots, \tau_N$  be the sequence of transpositions that are the increments of the random walk path leading to  $\sigma$ , so that  $\sigma = \tau_1 \dots \tau_N$ . Let  $\gamma$  be the geodesic between  $\sigma$  and  $I$  defined by  $\gamma_0 = \sigma$ ,  $\gamma_1 = \sigma\tau_N$ ,  $\gamma_2 = \gamma_1\tau_{N-1}, \dots$ , until the first time  $t$  such that multiplying  $\gamma_t$  by  $\tau_{N-t}$  would result in a coagulation of two cycles of  $\gamma_t$ . We do not allow this possibility (otherwise  $\gamma$  would not be a geodesic), and simply skip  $\tau_{N-t}$ :  $\gamma_{t+1} = \gamma_t\tau_{N-t-1}$ . We will see in a moment that this path never backtracks and that it ends at a bounded distance from  $I$ , to which it will be necessary to add a (bounded) number of steps so that it actually ends at  $I$ .

Let  $n(t)$  be the index of the transposition to be performed at time  $t$  on  $\gamma_t$ . Note that we can always write

$$\gamma_t = \tau_1\tau_2 \dots \tau_{n(t)} \prod_{i \in K_t} \tau_i$$

where  $K_t$  is a set whose size we will show is bounded. Indeed, even when we skip  $\tau_{n(t)}$  in  $\gamma_t$ , (so that  $n(t) \in K_{t+1}$ ), the following transpositions  $\tau_{n(t)-1}, \dots$  commute with the members of  $K_t$  with high probability and they can "jump above" the terms in  $K_t$  and cancel the rest of the transpositions  $(\tau_1 \dots \tau_{n(t)-1})$ .

**Lemma 14.** *For all  $t$ ,  $E(|K_t|) \leq O(1)$ , where  $O(1)$  is a constant that depends only on  $c < 1$ . As a consequence, the path ends at bounded distance from the identity and the distance  $E \sup_{t \leq cn/2} d(X_t, \gamma) = O(1)$ .*

*Proof.* There are two ways to add a member to  $K_{t-1}$  at time  $t$ . The first one is that performing  $\tau_{n(t)}$  will result in a coagulation, so that it is skipped by  $\gamma$ . The other way is if  $\tau_{n(t)}$  does not commute with one of the members of  $K_{t-1}$ , it stays stuck somewhere in  $K_t$ .

If  $\tau_{n(t)} = (i, j)$ , we claim that in order for  $i$  and  $j$  to be in the same cycle of  $\gamma_t$ ,  $i$  and  $j$  must belong to a component of the Erdős-Renyi graph associated with the random walk that contains a cycle at time  $cn/2$ . We will prove this in a moment, but if we admit this, then it follows that all transpositions in  $K_t$  act on vertices that belong to  $U(cn/2)$ , the unicyclic components of the random graph at time  $cn/2$ : if  $i \in K_t$ , then either  $\tau_i = (i, j)$  yields a coagulation in  $\gamma_t$ , or it doesn't commute with some member of  $(k, l)$  of  $K_{t-1}$ , in which case  $(i, j)$  overlaps with  $(k, l)$ . By induction,  $k, l \in U(cn/2)$ , therefore so are  $i$  and  $j$ .

Let us prove our claim that if  $(i, j)$  would yield a coagulation in  $\gamma_t$ , then  $i, j \in U(cn/2)$ . Let us observe first that  $i$  and  $j$  must already be in the same component of the random graph: because  $\tau_{n(t)}$  was performed on the random walk,  $i$  and  $j$  were connected at that point in the random graph and they remain so. If  $i$  and  $j$  are in different cycles of  $\sigma$ , then there must have been some ulterior fragmentation in their cycles, so the claim holds. When they are in the same cycle of  $\sigma$ , then there must be some transposition  $\tau_i$  with  $i \in K_t$  such that  $i$  and  $j$  are in different cycles of  $\gamma$  after  $\tau_i$ . Call those cycles  $C_1$  and  $C_2$ .  $\tau_i$  involves two members  $k$  and  $l$  of  $C_1 \cup C_2$ . Moreover the cycle structure of  $\gamma$  before  $(k, l)$  is performed must be of the form:

$$(k, \dots, i, \dots, j, \dots, l, \dots)$$

otherwise  $(k, l)$  cannot separate  $i$  and  $j$  at the next step. Unless  $i$  and  $j$  belong to a complex component, this implies that the cycle structure of  $\sigma$  has the same form. However this can only happen if  $k$  and  $l$  were connected to the

component of  $i$  and  $j$  at different times: otherwise the cycle structure would be of the form  $(i, \dots, j, \dots, k, \dots, l)$  or  $(i, \dots, k, \dots, l, \dots, j)$ . This implies in turn the existence of a cycle in the random graph component of  $i$  and  $j$  at time  $cn/2$ .

From there it follows in a straightforward way that  $|K_t| \leq |U(cn/2)|$  (in a unicyclic component there are as many edges as vertices). It is now standard in the theory of random graphs to show that  $|U(cn/2)|$  is bounded:

$$\begin{aligned} E|U(cn/2)| &\sim \left(\frac{\pi}{8}\right)^{1/2} \sum_{k=2}^{\infty} \binom{n}{k} \frac{k^{k-1/2}}{k!} \left(\frac{c}{n}\right)^k \left(1 - \frac{c}{n}\right)^{k(n-k) + \binom{k}{2} - k} k \\ &\sim \left(\frac{\pi}{8}\right)^{1/2} \sum_{k=2}^{\infty} \frac{k^{k+1/2}}{k!} (ce^{-c})^k < \infty \end{aligned}$$

which completes the proof of the lemma.

Now let  $X_t$  be a point on the random walk path. Since  $\gamma$  tries to perform all  $\tau_i$  (at reverse), there is a time  $s$  such that  $n(s) = t$ , i.e. the next transposition to be examined by  $\gamma_s$  is  $\tau_t$ . At this time,

$$\gamma_s = \tau_1 \dots \tau_t \prod_{i \in K_s} \tau_i$$

so that  $|K_s|$  steps is enough to reach  $\gamma_s$  from  $X_t$ . Since  $E(|K_s|) < O(1)$  by the lemma, we have proved that

$$E \sup_{t \leq cn/2} d(X_t, \gamma) \leq O(1)$$

and Theorem 26 is proved. □

### 3.3 Large deviations and volume growth

The goal of this section is to prove Theorem 2, which we restate here for convenience.

**Theorem 2.** *If  $0 \leq a \leq 1$  then as  $n \rightarrow \infty$ , we have  $|B(I, an)| \approx (n!)^a$  in a logarithmic sense, i.e.,*

$$\lim_{n \rightarrow \infty} \frac{\log |B(I, an)|}{n \log n} = a$$

**Sketch of the proof.** The proof of the result is more interesting than the limit. We begin by recalling the dynamics of the Chinese restaurant process (see e.g., Pitman (2002b)). Customer 1 enters and sits at table 1. At step  $i$ , customer  $i$  enters and starts a new table with probability  $1/i$  or sits to the left of customer  $k$  where  $k$  is chosen uniformly at random in  $\{1, \dots, i\}$ . From the tables we define a permutation  $\sigma$  by  $\sigma(i) = i$  if customer  $i$  is sitting by himself on his table and  $\sigma(i) = k$  if  $k$  sits to the right of  $i$ . It is easy to see that this defines a uniform random permutation on  $\mathcal{S}_n$ , and that the cycle structure is given by listing the individuals at the tables in clockwise order.

It is well-known that if  $\sigma \in \mathcal{S}_n$ , then  $d(\sigma) = n -$  the number of cycles of  $\sigma$ . In the Chinese restaurant process construction, let  $\zeta_i$  be the random variables taking the value 1 if customer  $i$  sits at an existing table (and 0 otherwise). The  $\zeta_i$ 's are independent Bernoulli  $(1 - 1/i)$  random variables. Recall that if  $\sigma$  is a permutation, then  $d(\sigma, I) = d(\sigma)$  is  $n - \#$ cycles of  $\sigma$ . Hence, if  $\sigma$  is uniformly distributed over  $\mathcal{S}_n$ , then  $d(\sigma)$  has the same distribution as  $S_n = \sum_{i=1}^n \zeta_i$ .

The  $i$ 's where a new cycle starts (i.e.  $\zeta_i = 0$ ) are distributed with the same law as that of the occurrences of records for i.i.d variables with continuous distribution function (cf. Durrett, example 6.2 of chapter 1). From calculations in that example it follows that  $(n - S_n)/\log n \rightarrow 1$  in probability.

Returning to our calculation of the volume of the ball,

$$|B(I, an)| = n! P(S_n \leq an)$$

for all  $0 < a < 1$ . It is straightforward to generalize large deviations results for i.i.d. random variables (see e.g., Durrett (2004), Section 2.9) to prove Theorem 1. One begins with the observation that for  $\lambda > 0$

$$P(S_n \leq an) \leq e^{\lambda an} E e^{-\lambda S_n} \tag{3.4}$$

optimizes the upper bound over  $\lambda$  and uses a change of measure argument to prove a corresponding lower bound.

## Proof of Theorem 2

Let  $\{\zeta_i, i \geq 1\}$  be independent with  $P(\zeta_i = 1) = 1 - 1/i$ , and let  $S_n = \sum_{i=1}^n \zeta_i$ . Since  $(\log n!)/(n \log n) \rightarrow 1$  it suffices to show:

**Lemma 15.** *Let  $0 < a < 1$ . As  $n \rightarrow \infty$ ,*

$$\lim_{n \rightarrow \infty} \frac{\log \mathbf{P}[S_n \leq an]}{n \log n} = a - 1$$

*Proof.* Let  $\varphi_n(\lambda) = \mathbf{E}[e^{-\lambda S_n}]$ . Using the definition we have

$$\varphi_n(\lambda) = \prod_{i=1}^n \left[ \left(1 - \frac{1}{i}\right)e^{-\lambda} + \frac{1}{i} \right] \equiv \prod_{i=1}^n q_i$$

where  $\equiv$  indicates that the last equation is the definition of  $q_i$ . By Markov's inequality we have

$$\log \mathbf{P}[S_n \leq an] \leq n(\lambda a + \frac{1}{n} \log \varphi_n(\lambda)), \quad \text{for all } \lambda \quad (3.5)$$

If we define

$$F_\lambda(x) = \frac{1}{\varphi_n(\lambda)} \int_{-\infty}^x e^{-\lambda y} dF^n(y)$$

Then  $F_\lambda$  is a distribution function such that

$$\text{mean}(F_\lambda) = -\frac{\varphi_n'(\lambda)}{\varphi_n(\lambda)} \quad \text{and} \quad \text{var}(F_\lambda) = \frac{d}{d\lambda} \frac{\varphi_n'(\lambda)}{\varphi_n(\lambda)} \geq 0$$

To optimize (3.5), we want to choose  $\lambda$  so that

$$a + \frac{1}{n} \frac{\varphi_n'(\lambda)}{\varphi_n(\lambda)} = 0$$

This says that the mean of the transformed distributions is  $na$ , so

$$a = \frac{1}{n} \sum_{i=1}^n \frac{(1 - 1/i)e^{-\lambda}}{q_i} = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)e^{-\lambda}}{(i-1)e^{-\lambda} + 1}$$

We guess that the optimal  $\lambda$  must be given by (asymptotically) by  $e^{-\lambda_{\text{opt}}} = b/n$ . Plugging this in the above gives,

$$a = \frac{1}{n} \sum_{j=1}^{n-1} \frac{jb/n}{(jb/n) + 1} \rightarrow \int_0^1 \frac{bx}{bx + 1} dx = 1 - \frac{1}{b} \log(b+1)$$

From this we see that we should choose  $b$  so that  $\log(b+1)/b = 1 - a$ .

*Upper bound:* Let us calculate what (3.5) gives with this choice of  $\lambda$ .

$$\frac{1}{n} \log \varphi_n(\lambda) = -\log n + \frac{1}{n} \log \prod_{i=1}^n \left( \left(1 - \frac{1}{i}\right)b + \frac{n}{i} \right) \quad (3.6)$$

$$\rightarrow -\log n + \int_0^1 \log(b + 1/x) dx \quad (3.7)$$

Since the last integral is finite it follows from (3.5) and  $\lambda_{\text{opt}} = -\log b + \log n$  that

$$\limsup_{n \rightarrow \infty} \frac{1}{n \log n} P(S_n \leq na) \leq a - 1$$

proving the upper bound half of Lemma 15.

*Lower bound.* The argument is similar to Durrett (2004, p.73). Fix any  $\nu < a$  and  $\nu < \nu' < a$ . Define a real number  $b'$  by

$$\frac{\log(1+b')}{b'} = 1 - \nu'$$

For any  $\lambda$ ,

$$\begin{aligned} \mathbf{P}[S_n \leq an] &\geq \int_{\nu n}^{an} dF^n(x) \geq \int_{\nu n}^{an} e^{\lambda x} \varphi_n(\lambda) dF_\lambda(x) \\ &\geq \varphi_n(\lambda) e^{\lambda n \nu} [F_\lambda(na) - F_\lambda(n\nu)] \end{aligned}$$

First, we prove that we can choose  $\lambda$  such that  $[F_\lambda(na) - F_\lambda(n\nu)] \rightarrow 1$ . Recall that the mean of  $F_\lambda$  is  $-\frac{\varphi'_n(\lambda)}{\varphi_n(\lambda)}$ , and that the latter function starts at  $n - \log n$  for  $\lambda = 0$ , is strictly decreasing and equals  $na$  when  $\lambda = \lambda_{\text{opt}} = -\log b + \log n$  i.e.  $e^{-\lambda_{\text{opt}}} = b/n$ . Thus if we pick  $\lambda = \lambda'$  such that  $e^{-\lambda'} = b'/n$ , the mean of  $F_{\lambda'}$  is by the lower bound calculation exactly  $n\nu'$ , and we have chosen  $\nu < \nu' < a$ . To conclude that  $F_{\lambda'}(na) - F_{\lambda'}(n\nu) \rightarrow 1$ , instead of using a law of large number arguments such as in the i.i.d. case, we simply compute the variance of  $F_{\lambda'}$  directly. Anticipating on the calculations of the next section, breaking the factor  $e^{-\lambda x}$  in the Radon-Nikodym derivative of  $F_\lambda$  into  $e^{-\lambda \sum x_i}$  means that we can see  $F_\lambda$  as a sum of independent Bernoulli random variables with parameter  $\beta_i$  so that the variance is

$$\text{var } F_{\lambda'} = \sum_{i=1}^n \beta_i(1 - \beta_i) \leq \sum_{i=1}^n \beta_i = \text{mean}(F_\lambda) = n\nu' = O(n)$$

Another way to obtain this inequality is to do more direct computations:

$$\begin{aligned}
\text{var } F_\lambda &= \frac{\varphi_n''(\lambda)}{\varphi_n(\lambda)} - \left( \frac{\varphi_n'(\lambda)}{\varphi_n(\lambda)} \right)^2 \\
&= \sum_{i=1}^n \frac{e^{-\lambda}(1-1/i)}{q_i} + \sum_{i \neq j} \frac{e^{-\lambda}(1-1/i)e^{-\lambda}(1-1/j)}{q_i q_j} \\
&\quad - \left( \sum_{i=1}^n \frac{e^{-\lambda}(1-1/i)}{q_i} \right)^2 \\
&= \sum_{i=1}^n \frac{e^{-\lambda}(1-1/i)}{q_i} - \sum_{i=1}^n \frac{e^{-2\lambda}(1-1/i)^2}{q_i^2} \\
&\leq \sum_{i=1}^n \frac{e^{-\lambda}(1-1/i)}{q_i} = \frac{\varphi_n'(\lambda)}{\varphi_n(\lambda)} = \text{mean } F_\lambda
\end{aligned}$$

Since the variance is  $O(n)$ , by Chebychev's inequality we have that  $F_\lambda(na) - F_\lambda(n\nu) \rightarrow 1$ . Therefore,

$$\liminf_{n \rightarrow \infty} \frac{\log \mathbf{P}[S_n \leq an]}{n \log n} \geq \nu - 1$$

But  $\nu$  is arbitrarily close to  $a$ , so the result is proved.  $\square$

### 3.4 The uniform measure on $\partial B(an)$

Let  $\{\zeta'_i, 1 \leq i \leq n\}$  have the distribution of  $\{\zeta_i, 1 \leq i \leq n\}$  conditional on  $\sum_{i=1}^n \zeta_i = \lfloor an \rfloor$ . Let  $\{\zeta_i^{(\lambda)}, 1 \leq i \leq n\}$  be independent with distribution

$$dF_{\lambda,i}(x) = \frac{1}{\phi_i(\lambda)} e^{-\lambda x} dF_i(x)$$

where  $F_i, \phi_i$  are respectively the distribution function and the Laplace transform of  $\zeta_i$ , and  $\lambda$  is the optimal parameter of the previous section,  $e^{-\lambda} = b/n$ . It is easy to see that  $\zeta_i^{(\lambda)}$  is another Bernoulli random variable with

$$\mathbf{P}[\zeta_i^{(\lambda)} = 1] = \mathbf{P}[\zeta_i = 1] e^{-\lambda} \frac{1}{\phi_i(\lambda)} = \frac{1}{1 + \frac{n}{b(i-1)}} := \beta_i$$

We are now ready to prove

**Theorem 28** *Let  $|\mathcal{C}_1|$  be the length of the cycle that contains 1. Under  $\mu$ , the uniform distribution on  $\partial B(I, an)$ ,*

$$|\mathcal{C}_1| \Rightarrow \mathbf{G}$$

where  $\mathbf{G}$  is a geometric r.v. with  $P(\mathbf{G} > k) = (b/(1+b))^k$  and  $b$  satisfies  $\log(1+b)/b = 1-a$ .

**Sketch of the proof** The first part of demonstrating this is to recall what Arratia, Barbour, and Tavaré (2004) call the Feller coupling. Start with vertex 1 and choose  $\sigma(1)$  uniformly from the  $n$  possible choices. If this is 1, then take vertex 2 and choose  $\sigma(2)$  uniformly from the  $n-1$  remaining possible choices. If  $\sigma(1) \neq 1$  then choose  $\sigma(\sigma(1))$  uniformly from the  $n-1$  remaining choices, and so on, until the final vertex where there is only one possible choice. Although the construction is much different from the Chinese Restaurant Process, the reader should note that the variables  $\xi_i = 1$  if a cycle is not completed at the  $i^{\text{th}}$  stage then  $\{\xi_i : 1 \leq i \leq n\}$  and  $\{\zeta_i : 1 \leq i \leq n\}$  have the same distribution.

From the last observation it follows that  $N = \inf\{i : \xi_i = 1\}$  has the same distribution as the length of the cycle containing 1. We can now conclude the proof of the Theorem, using the large deviations calculation of the volume, and an argument called the Gibbs conditioning principle (see Dembo-Zeitouni (1996)). This principle asserts that the distribution of the  $\zeta_i$  conditional on  $\sum_{i=1}^n \zeta_i = an$  should be asymptotically independent and their law given by that which minimizes the entropy, i.e., the random variables  $\zeta_i^{(\lambda)}$  with distribution:

$$\frac{1}{\phi_i(\lambda)} e^{-\lambda x} dF_i(x) \tag{3.8}$$

where  $F_i, \phi_i$  are respectively the d.f. and the Laplace transform of  $\zeta_i$ , and  $\lambda$  is the parameter that optimizes (3.4), i.e.,  $e^{-\lambda} = b/n$ .

### Proof of Theorem 3

We will first need a lemma.

**Lemma 16.** *For any  $n \geq 1$  and for every  $\lambda > 0$ , then*

$$(\zeta'_1, \dots, \zeta'_n) \stackrel{d}{=} (\zeta_1^{(\lambda)}, \dots, \zeta_n^{(\lambda)}) \text{ given } \sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor$$

*Proof.* Let  $f_1, \dots, f_n$  be bounded nonnegative Borel functions.

$$\begin{aligned} E(f_1(\zeta'_1) \cdot \dots \cdot f_n(\zeta'_n)) &= E\left(f_1(\zeta_1) \cdot \dots \cdot f_n(\zeta_n) \left| \sum_{i=1}^n \zeta_i = \lfloor an \rfloor \right.\right) \\ &= E\left(f_1(\zeta_1) \cdot \dots \cdot f_n(\zeta_n); \sum_{i=1}^n \zeta_i = \lfloor an \rfloor\right) \\ &\quad \times P\left(\sum_{i=1}^n \zeta_i = \lfloor an \rfloor\right)^{-1} \end{aligned}$$

On the other hand,

$$\begin{aligned} E\left(f_1(\zeta_1^{(\lambda)}) \cdot \dots \cdot f_n(\zeta_n^{(\lambda)}); \sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor\right) &= \int_{R^n} f_1(x_1) \cdot \dots \cdot f_n(x_n) \mathbf{1}_{\{\sum x_i = \lfloor an \rfloor\}} \prod_{i=1}^n dF_{i,\lambda}(x) \\ &= \int_{R^n} f_1(x_1) \cdot \dots \cdot f_n(x_n) \mathbf{1}_{\{\sum_i x_i = \lfloor an \rfloor\}} \prod_{i=1}^n \frac{e^{-\lambda x}}{\phi_i(\lambda)} dF_i(x) \\ &= \frac{e^{-\lambda \lfloor an \rfloor}}{\prod_{i=1}^n \phi_i(\lambda)} \int_{R^n} f_1(x_1) \cdot \dots \cdot f_n(x_n) \mathbf{1}_{\{\sum_i x_i = \lfloor an \rfloor\}} \prod_{i=1}^n dF_i(x) \\ &= \frac{e^{-\lambda \lfloor an \rfloor}}{\prod_{i=1}^n \phi_i(\lambda)} E[f_1(\zeta_1) \cdot \dots \cdot f_n(\zeta_n); \sum_{i=1}^n \zeta_i = \lfloor an \rfloor] \end{aligned}$$

We can now divide and multiply by the probability of the events in the two sides of this equation to obtain that for some constant  $C > 0$

$$E\left(f_1(\zeta_1^{(\lambda)}) \cdot \dots \cdot f_n(\zeta_n^{(\lambda)}); \sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor\right) = CE\left(f_1(\zeta_1) \cdot \dots \cdot f_n(\zeta_n) \left| \sum_{i=1}^n \zeta_i = \lfloor an \rfloor \right.\right)$$

By taking  $f_1 = \dots = f_n = 1$  we see that  $C = 1$  and the lemma is proved.  $\square$

We will need another lemma :

**Lemma 17.** *The  $\zeta_i^{(\lambda)}$  satisfy a local central limit theorem:*

$$\mathbf{P}\left(\sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor\right) \sim Cn^{-1/2}$$

*Proof.* The proof of this local limit theorem follows very closely that of the usual i.i.d. case, which can be found in Theorem 5.2 of Durrett (2004). Let  $\beta_m = P(\zeta_m^{(\lambda)} = 1)$ , (i.e.  $\beta_m = (1 + n/b(m-1))^{-1}$ ), and let  $X_{m,n} = n^{-1/2}(\zeta_m^{(\lambda)} - \beta_m)$  be the rescaled Bernoulli variable. We start by noticing that  $X_{m,n}$  satisfy the hypotheses of the Lindeberg-Feller theorem (Theorem 4.5 in Durrett (2004)). Indeed, they are independent by definition; for all  $\varepsilon > 0$ ,  $P(|X_{m,n}| > \varepsilon) = 0$  as soon as  $n^{-1/2} \leq \varepsilon$ , since  $\zeta_m^{(\lambda)} \leq 1$  and  $\beta_m \leq 1$  as well. Moreover,

$$\begin{aligned} \sum_{m=1}^n E(X_{n,m}^2) &= \frac{1}{n} \sum_{m=1}^n \beta_m(1 - \beta_m) \\ &\rightarrow \int_0^1 \frac{x/b}{(1 + (x/b))^2} dx := \sigma^2 \end{aligned}$$

Therefore  $\sum_{m=1}^n X_{m,n} \Rightarrow \mathcal{N}(0, \sigma)$ . At this point, the proof of the local limit theorem from Durrett (2004) can be reproduced exactly. Therefore:

$$\sup_{x \in \mathbf{R}} \left| n^{1/2} P \left( \sum_{m=1}^n X_{m,n} = x \right) - n(x) \right| \rightarrow 0$$

where  $n(x) := (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$ . Since  $\sum_{m=1}^n \beta_m \rightarrow an$ , and since  $n(\cdot)$  is a continuous function, we can conclude the proof of the lemma by the above uniform convergence.  $\square$

Now, by the Feller coupling,  $|\mathcal{C}_1| \stackrel{d}{=} \inf\{k \geq 1 : \zeta'_{n-k} = 0\}$ , i.e. we must reverse the time of the Chinese restaurant process. Hence by lemma 3 and 4:

$$\begin{aligned}
\mathbf{P}[\mathcal{C}_1 > k] &= \mathbf{P}[\zeta'_n = 1, \dots, \zeta'_{n-k+1} = 1] \\
&= \mathbf{P}[\zeta_n^{(\lambda)} = 1, \dots, \zeta_{n-k+1}^{(\lambda)} = 1 \mid \sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor] \\
&= \mathbf{P}[\zeta_n^{(\lambda)} = 1, \dots, \zeta_{n-k+1}^{(\lambda)} = 1; \sum_{i=1}^{n-k-1} \zeta_i^{(\lambda)} \\
&\quad \times \mathbf{P}[\sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor] = \lfloor an \rfloor - k]^{-1} \\
&= \mathbf{P}[\zeta_n^{(\lambda)} = 1] \dots \mathbf{P}[\zeta_{n-k}^{(\lambda)} = 1] \mathbf{P}[\sum_{i=1}^{n-k-1} \zeta_i^{(\lambda)} = \lfloor an \rfloor - k] \\
&\quad \times \mathbf{P}[\sum_{i=1}^n \zeta_i^{(\lambda)} = \lfloor an \rfloor]^{-1} \\
&\sim \frac{1}{1 + \frac{n}{b(n-1)}} \cdots \frac{1}{1 + \frac{n}{b(n-k)}} \\
&\rightarrow \frac{1}{(1 + 1/b)^k}
\end{aligned}$$

Hence Theorem 2 is proved.  $\square$

### 3.5 Asymptotic singularity between $\mu$ and $\nu$

In this section we give a proof to Theorem 30 that follows in an almost straightforward way from Theorems 2 and 3:  $\nu$  and  $\mu$  concentrate on permutations that have a different number of fixed points. First recall the statement of the Theorem:

**Theorem 30** *As  $n \rightarrow \infty$ , the hitting distribution  $\nu$  and the uniform distribution  $\mu$  on a sphere of radius  $an$  are asymptotically singular:*

$$d_{TV}(\mu, \nu) \rightarrow 1$$

**Lemma 18.** *The random partition of  $\{1, \dots, n\}$  derived from  $\nu$  is exchangeable.*

*Proof.* The probability to obtain a certain partition of  $\{1, \dots, n\}$  under  $\nu$  only depends on the size of its blocs, which stays the same under the action of a given permutation. Hence  $\nu$  yields an exchangeable partition of  $\{1, \dots, n\}$ .  $\square$

An immediate consequence is that the expected number of fixed points is  $n\nu(\mathcal{C}_1 = 1) = n/(1+b)$ . Next we show that under  $\nu$  the number of fixed points  $N$  is close to its expected value.

**Lemma 19.**

$$\text{var } N = o(n^2)$$

under  $\nu$ .

*Proof.* Let  $x_i = \mathbf{1}_{\{\zeta'_i=0; \zeta'_{i+1}=0\}}$  be the indicator of the event that in the conditioned Chinese restaurant process, client number  $i$  sits by himself. Then  $N = \sum_i x_i$  and

$$\begin{aligned} \text{var } N &= \sum_{i=1}^n \text{var } x_i + \frac{1}{2} \sum_{i < j} \text{cov}(x_i, x_j) \\ &\leq n + \frac{1}{2} \sum_{i < j} \text{cov}(x_i, x_j) \end{aligned}$$

But when  $j - i > 1$ , by the Gibbs asymptotic independence proved in Theorem 28,  $\text{cov}(x_i, x_j) \rightarrow 0$ . Also, there are only  $O(n)$  terms such that  $j = i + 1$  and in this case  $\text{cov}(x_i, x_{i+1}) \leq 1$  hence the sum  $\sum_{i < j} \text{cov}(x_i, x_j) = o(n^2)$ .  $\square$

To end the proof of Theorem 30 by Chebychev's inequality there remains only to notice that:

**Lemma 20.** For  $0 < a < 1$  and large enough  $n$

$$\nu(|\mathcal{C}_1| = 1) \neq \mu(|\mathcal{C}_1| = 1)$$

*Proof* Recall that  $b$  is defined by  $\log(1+b)/b = 1-a$ . For  $x \in (0, 1)$ , let  $f(x) = 1 - \log(1+x)/x$ , so that  $b = f^{-1}(a)$ .

On the other hand an easy consequence of Berestycki Durrett (2004) or Theorem 0 is  $\mu(|\mathcal{C}_1| = 1) = e^{-u^{-1}(a)}$ . (Indeed, under  $\mu$ ,  $|\mathcal{C}_1|$  is asymptotically the total progeny of a  $PGW$  process with parameter  $u^{-1}(a)$ ).

Hence the lemma is proved if we show that

$$1/(1+b) \neq e^{u^{-1}(a)} \text{ or } u(x) \neq 1 - x/(e^x - 1)$$

for all  $x > 0$ .

We start by noticing that as  $x \rightarrow 0$ ,  $u(x) \sim x$  but  $1 - x/(e^x - 1) \sim x/2$ . Hence  $u(x) > 1 - x/(e^x - 1)$  as  $x \rightarrow 0$ . The same is true as  $x \rightarrow \infty$  (an easy argument shows indeed that  $u(x) = 1 - e^{-x} + o(e^{-x})$ ). Now those functions are both concave as we will see in a moment, hence this has to stay true on the whole open half-line  $x > 0$ . (Notice that we have thus proved that the hitting distribution has always less fixed points than the uniform distribution).

**Lemma 21.** *The function  $u$  appearing in Theorem 0 is concave.*

*Proof.* For  $c \leq 1$  this is obvious. When  $c > 1$ , rather than carrying explicit calculations on the second derivative of  $u$ , we use a theoretic argument that exploits the recent result of Schramm (2004), which says that the size of the pieces of the giant component in the random graph have approximately a Poisson-Dirichlet distribution. Since each fragmentation decreases the distance by 1 and each coalescence increases it by 1, it is easy to see that

$$\frac{d}{dc} \mathbf{E}[d(\sigma_{cn/2}, I) | \mathcal{F}_{cn/2}] = 1 - 2\mathbf{P}[\text{fragm.} | \mathcal{F}_{cn/2}]$$

where  $\mathcal{F}$  is the canonical filtration generated by the random walk. So we need to show that  $\mathbf{P}[\text{fragm.}]$  is an asymptotically increasing function of  $c$ . However the probability of fragmenting a small cycle is asymptotically 0 (by duality and the fact that  $u$  is linear in the subcritical regime), and the probability of fragmenting one of the giant cycles can be computed explicitly using the Poisson-Dirichlet structure:

$$\mathbf{P}[\text{fragm.}] \rightarrow \mathbf{E} \sum_{i=1}^{\infty} (\theta(c) X_i)^2 = \theta(c)^2 \mathbf{E} \sum_{i=1}^{\infty} X_i^2 = \frac{1}{2} \theta(c)^2$$

where  $\theta(c)$  is the survival probability of a  $PGW(c)$  branching process and  $(X_i, i \geq 1)$  follows the  $PD(1)$  distribution. ( $\mathbf{E} \sum X_i^2 = 1/2$  follows from Pitman, formula (128)). Since  $\theta(c)$  is an increasing function of  $c$ , the lemma is proved (and thus, so is Theorem 30).  $\square$

**Remark.** We have thus proved the following formula

$$u(c) = c/2 - \int_0^{c/2} \theta(u)^2 du$$

which is perhaps a little simpler to handle than the expression in Theorem 0.

### 3.6 Number of geodesics and Radon-Nikodym derivative

Here we prove the following theorem, which we will then use to prove a stronger version of the singularity theorem.

**Theorem 31** *Suppose  $d(\sigma) = t$  and  $m_1, \dots, m_j$  are the cycle lengths of  $\sigma$ . The number of paths of length  $t$  from  $I$  to  $\sigma$  is*

$$t! \prod_{i=1}^j \frac{m_i^{m_i-2}}{(m_i-1)!}$$

*From this it follows that if  $t = \lfloor cn/2 \rfloor$  with  $c < 1$  then*

$$r(\sigma) = K_{n,t} \prod_{i=1}^j \frac{m_i^{m_i-2}}{(m_i-1)!}$$

*where  $K_{n,t}$  is a constant that only depends on  $n$  and  $t$ .*

**Sketch of the proof.** To see the first result, note that in order to go from  $\sigma$  to  $I$  in the shortest number of steps we must increase the number of cycles by 1 at each step, and to do this we must fragment a cycle at each step by transposing two of its elements. A cycle of length  $m_i$  will require  $m_i - 1$  fragmentations. The first step in constructing a path is to decide on how to allocate the  $t$  moves between the original cycles which can be done in  $t! / \prod_{i=1}^j (m_i - 1)!$  ways. The next step is to count the number of ways that we can reduce a cycle of length  $m_i$  in  $m_i - 1$  steps, which turns out to be simple:  $m_i^{m_i-1}$ .

### Proof of Theorem 31

Given a partition of  $\{1, 2, \dots, n\}$  into groups  $A_1, \dots, A_j$  of sizes  $m_i$ ,  $1 \leq i \leq j$ , the number of forests that consist of trees with vertex sets  $A_1, \dots, A_j$  is by Cayley's formula for the number of unrooted trees on  $m_i$  vertices

$$\prod_{i=1}^j m_i^{m_i-2}$$

Let  $t = \sum_i (m_i - 1)$ . A given forest can be built up in  $t!$  ways so there are

$$t! \prod_{i=1}^j m_i^{m_i-2}$$

paths for our random graph process that end up producing a given partition. The number of permutations that correspond to a given partition is

$$\prod_{i=1}^j (m_i - 1)!$$

An equal number of paths end at each permutation with cycle sizes  $m_i$ ,  $1 \leq i \leq j$ , so the number of paths to a given permutation is

$$t! \prod_{i=1}^j \frac{m_i^{m_i-2}}{(m_i - 1)!}$$

If  $t = \lfloor cn/2 \rfloor$  with  $c < 1$  then the number of edge choices that end up producing no fragmentations is by Theorem 1 in Berestycki and Durrett (2004)

$$\sim \left( \binom{n}{2} \right)^t e^{-\kappa(c)}$$

Taking the ratio of the last two results gives Theorem 31. □

## 3.7 The size of the support of the hitting distribution

In this section we prove Theorem 32, restated below.

**Theorem 32.** *Suppose  $a < 1/2$ . There exists a set  $S_n \in \partial B(an)$  such that  $\nu(S_n) \rightarrow 1$  as  $n \rightarrow \infty$  and*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{|S_n|}{|\partial B(an)|} = \gamma < 0$$

**Sketch of the proof** Obtaining a decay at least exponential is not very hard, even in the case  $a > 1/2$ . However it is not easy to prove that this is the correct rate for the decay of  $|S|/|\partial B|$ , and we restrict ourselves to the case  $a < 1/2$ .

If  $\sigma \in \partial B(an)$ , then by Theorem 31 we find that

$$\log \nu_0(\sigma) = -an \log n + an + \sum_{k=1}^n a_k \log p_k + o(n)$$

where  $p_k$  is the Borel distribution with parameter  $c$ , and  $a_k$  is the number of cycles of  $\sigma$  of size  $k$ . But by the law of large numbers,  $\nu_0(a_k/n)$  should have a limit as  $n \rightarrow \infty$ . Hence there is a set  $S$  such that  $(\log \nu_0(\sigma) + an \log n)/n$  has a limit  $-c_1$  whenever  $\sigma \in S$ . Because  $\nu_0(S) \approx 1$ ,  $|S| \approx \exp(an \log n + c_1 n)$ . Moreover it is also true that  $\nu(S) \rightarrow 1$ . On the other hand, precise estimates on the size of  $\partial B(an)$  obtained via Kolchin's representation theorem tell us that  $|\partial B(an)| = \exp(an \log n + c_2 n + o(n))$ . (A statement of Kolchin's theorem can be found below.) Thus, the theorem holds with  $\gamma = c_1 - c_2$ . To prove that  $\gamma \neq 0$ , we argue that the decay has to be at least exponential (a consequence of Kolchin's representation theorem).  $\square$

## Proof of Theorem 32

We will need precise estimates on the size of  $\partial B(an)$ . Because we need estimates to order higher than just  $n \log n$ , sticking to the large deviations approach is not good enough. Rather, we will use Kolchin's representation theorem. We would like to thank Jim Pitman for pointing out this reference to us.

Suppose we can partition  $\{1, \dots, n\}$  into a certain number of clusters, which can all have different internal states. To be more specific, suppose that each partition of  $\{1, \dots, n\}$  into  $k$  clusters leads to  $v_k$  possible global states of the system  $\{1, \dots, n\}$ , and that we can further assign each cluster of size  $j$

one of  $w_j$  possible internal states. We call such a combinatorial structure a  $(v, w)$ -partition (of  $\{1, \dots, n\}$ ). Kolchin's representation theorem answers with probabilistic means to the following purely combinatorial question : how many different  $(v, w)$ -partitions are there ? Also, what does a random, uniform,  $(v, w)$ -partition look like ?

Before going into the details of this theorem, let us see its relevance to our problem. The number of permutations at distance  $an$  from the identity is a special instance of the above Kolchin problem, where  $v_k = \mathbf{1}_{\{k=(1-a)n\}}$  and  $w_j = (j-1)!$ . Indeed a permutation at distance  $an$  is exactly a permutation having  $(1-a)n$  cycles and each cluster of size  $j$  can be in one of the  $(j-1)!$  possible orderings of the cycle.

Here is the content of Kolchin's theorem. (See Pitman (2002a)). Let  $v(\theta) = \sum_{k=1}^{\infty} v_k \theta^k / k!$  and let  $w(\xi) = \sum_{j=1}^{\infty} w_j \xi^j / j!$  be the so-called exponential generating function of the sequences  $v$  and  $w$ . Let  $K$  be an integer-valued random variable with distribution

$$P(K = k) = v_k \frac{w(\xi)^k}{k! v(w(\xi))}$$

and let  $X$  be random variable distributed according to

$$P(X = j) = \frac{w_j \xi^j}{j! w(\xi)}$$

Here  $\xi$  is any parameter. In our setting,  $K = (1-a)n, a.s.$  and  $X$  has the so-called logarithmic distribution,  $P(X = j) = b^j / j \cdot \frac{1}{-\log(1-b)}$ , for some parameter  $b = w(\xi)$ .

**Theorem 34.** (Kolchin) *The number of  $(v, w)$ -partitions is given by*

$$\frac{n! v(w(\xi))}{\xi^n} P\left(\sum_{i=1}^K X_i = n\right)$$

where  $X_i$  are *i.i.d* samples of the variable  $X$ . Moreover, the sizes of the clusters in exchangeable random order have the same law as

$$(X_1, \dots, X_K) \text{ given } X_1 + \dots + X_K = n$$

For a precise definition of exchangeable random order, and further discussion of this theorem, see Pitman (2002a). It is to be noted that here  $\xi$  is any parameter. By playing on this parameter so as to make the event  $S_K = n$  not unlikely (e.g., of probability  $\propto n^{-1/2}$  rather than exponentially small), we get that the size of the clusters are approximately drawn from the r.v.  $X$ . Note that as a consequence we get here another proof of Theorem 28. Indeed, we see that the sizes of the cycles of a uniform permutation on  $\partial B$  in exchangeable random order have a logarithmic distribution (asymptotically when the parameter is chosen suitably). Hence, a size-biased pick  $|\mathcal{C}_1|$  should have distribution  $P(X' = j) = \text{const.} \cdot j \cdot \frac{b^j}{j} \propto b^j$ , a geometric random variable. The similarity between the large deviations - statistical mechanics approach and Kolchin's theorem is striking.

Another straightforward consequence of this Theorem is the precise asymptotics for the size of a ball of radius  $an$ . Indeed, in our setting,  $v(\theta) = \theta^{(1-a)n}/[(1-a)n]!$  and  $w(\xi) = -\log(1-\xi)$  hence:

$$|\partial B(an)| = \frac{n!(-\log(1-\xi))^{(1-a)n}}{\xi^n((1-a)n)!} P\left(\sum_{i=1}^{(1-a)n} X_i = n\right)$$

where  $\xi$  is still any parameter. However, when  $\xi$  is chosen such that  $(1-a)E(X) = 1$ , the local central limit theorem shows that  $P(\sum_{i=1}^{(1-a)n} X_i = n) \sim Cn^{-1/2}$ . By Stirling's formula, it is now straightforward to see that

$$|\partial B(an)| = \exp(an \log n + c_2 n + o(n))$$

Let us now turn our attention to the hitting distribution. We will get the corresponding estimate by analyzing the Radon-Nikodym derivative  $r(\sigma)$  and the Law of large numbers for  $\nu$ , as mentioned in the sketch of the proof.

More precisely, it follows from the proof of Theorem 31 that if  $\sigma \in \partial B(an)$ , with cycle decomposition of size  $m_1, \dots, m_{(1-a)n}$ , and  $t = an$  then

$$\nu(\sigma) = \frac{1}{\binom{n}{2}^t e^{-\kappa(c)}} t! \prod_{i=1}^{n(1-a)} \frac{m_i^{m_i-2}}{(m_i-1)!}$$

Let us write  $a_k$  for the number of cycles of  $\sigma$  of size  $k$ , so that  $\sum_{k=1}^n a_k = n(1-a)$  and  $\sum_{k=1}^n k a_k = n$ . We can rewrite the above as

$$\nu(\sigma) = \frac{1}{\binom{n}{2}^t e^{-\kappa(c)}} t! \prod_{k=1}^n \left( \frac{1}{c} \frac{k^{k-2}}{(k-1)!} (ce^{-c})^k \right)^{a_k} c^{(1-a)n} \frac{1}{(ce^{-c})^n}$$

When we take the logarithm, calling  $q_k = \frac{1}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k$  and  $p_k = kq_k$

$$\log \nu(\sigma) = \kappa(c) - t \log \binom{n}{2} + \log t! + \sum_{k=1}^n a_k \log p_k + n(1-a) \log c - n \log c + cn$$

Recalling that  $t = an$ ,  $c = 2a$ , and using Stirling's formula, we find that

$$\log \nu(\sigma) = -an \log n + an + \sum_{k=1}^n a_k \log p_k + o(n)$$

We would now like to use the law of large numbers for  $\nu$  since

$$\nu(\{\sigma \in \partial B : |\frac{a_k}{n} - q_k| \leq \varepsilon; \forall 1 \leq k \leq n\}) \rightarrow 1$$

for given  $\varepsilon > 0$ , and where  $p_k$  is the Borel distribution, but this is not possible directly since we would obtain a bound  $\varepsilon \sum_{k=1}^{\infty} \log p_k = \infty$ . So we need to modify our choice : let

$$S_n = \{\sigma \in \partial B : |\frac{a_k}{n} - q_k| \leq (\log n)^{-5}; \forall 1 \leq k \leq (\log n)^2 \text{ and } a_k = 0 \text{ otherwise}\}$$

First note that  $\nu_0(S_n) \rightarrow 1$ . Indeed, we know that the order of deviation of  $a_k$  from its mean  $nq_k$  should be of order  $n^{1/2}$ , which is much smaller than the  $n(\log n)^{-5}$  in the definition of  $S_n$ . Furthermore, no cycle can be greater than  $\beta \log n$  (for some  $\beta > 0$ ) when  $a < 1/2$ , because of the coupling with the Erdős-Renyi random graph. Thus  $\nu(\partial B - S_n) \rightarrow 0$ . On the other hand, for all  $\sigma \in S_n$ ,

$$\begin{aligned} \frac{\log \nu(\sigma) + an \log n}{n} &\geq a + \sum_{k=1}^{(\log n)^2} (q_k + (\log n)^{-5}) \log p_k + o(1) \\ &\geq a + \sum_{k=1}^{(\log n)^2} q_k \log p_k + o(1) \end{aligned}$$

from which we deduce that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} (\log \nu(\sigma) + an \log n) = a + \sum_{k=1}^{\infty} q_k \log p_k := -c_1$$

After similar treatment for the limsup, we get

$$\frac{1}{n}(\log \nu(\sigma) + an \log n) \rightarrow -c_1$$

Since

$$\nu_0(S_n) = \sum_{\sigma \in S_n} \nu_0(\sigma) \rightarrow 1$$

it must be that  $|S_n| = \exp(an \log n + c_1 + o(n))$ . Therefore

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{|S_n|}{|\partial B|} = c_1 - c_2 := \gamma$$

It now remains to show that  $\gamma \neq 0$ . Observe that by Kolchin's Theorem, we could pursue the asymptotic expansion of  $|\partial B(an)|$  and the next term would be polynomial in  $n$ . From the exact formula of  $\nu(\sigma)$ , we could also find the next term for  $|S|$  and find that it is polynomial. Hence if  $\gamma = 0$ , then we would  $|S|/|\partial B| \sim n^{-\alpha}$  for some  $\alpha \geq 0$ . But, another consequence of Kolchin's Theorem is that the decay has to be at least exponential: for instance, permutations in  $S$  have a number of fixed points characteristic of  $\nu$  and not of  $\mu$ . As we have seen earlier the number of fixed points under  $\mu$ ,  $n/(1+b)$ , is smaller than under the hitting distribution. But since the number of fixed points under  $\mu$  is given by a sum of almost independent random variables:

$$\sum_{i=1}^{n-\lfloor an \rfloor} \mathbf{1}_{\{X_i=1\}} \text{ given } \sum_{i=1}^{n-\lfloor an \rfloor} X_i = n$$

we have that

$$\begin{aligned} \mu(S) &\leq P \left( \frac{1}{n} \sum_{i=1}^{n-\lfloor an \rfloor} \mathbf{1}_{\{X_i=1\}} < \frac{1}{1+b} \middle| \sum_{i=1}^{n-\lfloor an \rfloor} X_i = n \right) \\ &\leq Cn^{1/2} e^{-n\rho} \leq C' e^{-n\rho'} \end{aligned}$$

by standard large deviations (here, simply Markov's inequality), and because the event on which we condition is of probability  $Cn^{-1/2}$ . Hence the decay has to be at least exponential and  $\gamma$  cannot be 0.  $\square$

**Remark.** The same argument shows that the hitting distribution of  $\sigma$  is supported on a set at least exponentially small even in the case  $a > 1/2$ , but of course we do not know whether this is a precise asymptotics. If the decay is still exponential after  $a = 1/2$  it seems likely that the exponential coefficient will not be smooth at  $a = 1/2$ . In figure 3.2 we have plotted the value of this coefficient against a time-change of  $a$ . It would be interesting to compute exact asymptotics in the case  $a > 1/2$  and make this picture complete.

**Remark.** Kolchin's representation theorem could have been used already earlier for the proofs of Theorems 2 and 3. This would actually simplify the proof of both results. However, we have chosen to keep the proofs as they were, because they do not rely on a technical result such as Kolchin's theorem, which is not so well-known as standard large deviations theory.

### 3.8 Asymptotic hyperbolicity under the uniform measure

Here we present a proof of Theorem 33. The sketch of the proof below contains some ideas that will be used and not re-explained in the actual proof that follows.

**Theorem 33** *Let  $0 < a < 1$  and let  $\sigma, \pi$  be two random independent points chosen uniformly from  $\partial B(an)$ . Then*

1. *If  $a < 1 - \log 2$*

$$E(\sigma|\pi)_p \leq \delta(\log n)^2$$

*for some  $0 < \delta = \delta(a) < \infty$ . Moreover, with probability asymptotically 1, there is a geodesic between  $\sigma$  and  $\pi$  that comes within distance at most  $\delta(\log n)^2$  of  $p$ .*

2. *If  $a > 1 - \log 2$*

$$E(\sigma|\pi)_p \sim \delta n$$

*for some  $\delta = \delta(a) > 0$ . Moreover, no geodesic can approach  $p$  closer than  $\delta'n$  for some  $0 < \delta' < \infty$ .*

**Sketch of the proof.** To guess what the answer is, we exploit once again the connection with the theory of random graphs. The first thing to do is to realize that because of the symmetries of the Cayley graph  $G_n$  it is enough

to look at  $d(I, \sigma \cdot \pi)$  and see whether it is approximately  $2an$  or much smaller than  $2an$ .

To construct our graph, we will need some notations. Let

$$\pi = \tau_1 \tau_2 \cdots \tau_{an} \tag{3.9}$$

be a minimal decomposition of  $\pi$  as a product of transpositions, with the following convention. If we list all cycles of  $\pi$  in the order of their left element, then the transpositions  $\tau_i$  are those  $(x, y)$  such that  $y$  comes just after  $x$  in the cyclic decomposition of  $\pi$  and  $x$  and  $y$  are in the same cycle, and we order the  $an$  transpositions according to their position in this canonical decomposition. To clarify the ideas, suppose

$$\pi = (1\ 4\ 3\ 7)(2)(5\ 8)(6\ 10\ 9)$$

then we write

$$\pi = (1\ 4)(4\ 3)(3\ 7)(5\ 8)(6\ 10)(10\ 9)$$

We define the graph  $\Gamma = (V, E)$  on  $n(1 - a)$  vertices as follows. Let  $V = \{\text{cycles of } \sigma\}$ , and there is an edge between  $\mathcal{C}$  and  $\mathcal{C}'$  if there is  $x \in \mathcal{C}$  and  $y \in \mathcal{C}'$  such that  $(x, y)$  is one of the  $an$  transpositions in the minimal decomposition described above. Note that this graph could have self-loops and multi-edges.

A notion that we will use on several occasions is that of being a terminal point. We say  $x \in \{1, \dots, n\}$  is *terminal* if  $x$  doesn't appear more than once in the transpositions of the above minimal decomposition. This means that, with those conventions,  $x$  is situated at the "end" of the cycle of  $\pi$  in which it is contained.

Here is why we are interested in the properties of the graph  $\Gamma$ . If we define  $\sigma_0 = \sigma$  and, for  $1 \leq r \leq an$ ,  $\sigma_r = \sigma \cdot \tau_1 \cdots \tau_r$ , and consider the process  $(\sigma_r, 0 \leq r \leq an)$ , this is a walk on  $G_n$  starting at  $\sigma$  and ending at  $\sigma \cdot \pi$ . Moreover, since at each step we are multiplying by a transposition, the cycles of  $\sigma_r$  evolve according to a discrete coagulation-fragmentation chain, with cycles merging when the transposition involves elements from different cycles, and cycles splitting otherwise (as it is the case for simple random walk on  $G_n$ ). Therefore,  $\Gamma$  is the graph that results from drawing an edge between two cycles of  $\sigma$  as we encounter a transposition joining those two cycles. In particular, the same argument that shows that the Erdős-Renyi random

graph is an upper-bound for the sizes the cycles of simple random walk on  $G_n$ , will show that the cycles of  $\sigma \cdot \pi$  are subcomponents of the connected components of  $\Gamma$ , with possibility of fragmentation whenever there is a cycle in  $\Gamma$ , or a self-loop or a multiple-edge. All other edges represent coalescence of cycles in the walk  $(\sigma_r, 0 \leq r \leq an)$ .

In particular, the property of  $\Gamma$  that we will be mostly interested in, will be to decide whether  $\Gamma$  has a giant component, meaning a component containing a positive fraction of all  $n(1 - a)$  vertices. Indeed when all cycles of  $\Gamma$  are small, we should expect very few cycles in  $\Gamma$  and hence little fragmentation. Hence most steps of the walk  $\sigma_r$  are coalescence events and the number of cycles decreases linearly: in other words, in the case that all cycles are small,  $d(\sigma, \pi) \approx 2an$ . On the other hand, if  $\Gamma$  contains a giant cycle, then we can expect many cycles in the graph and hence many fragmentation events in the walk  $(\sigma_r, 0 \leq r \leq an)$ , which means that  $d(\sigma, \pi) \ll 2an$ .

Here is our strategy to see whether there is a giant component in  $\Gamma$ . Rather than counting the number of cycles of  $\sigma$  that a component of  $\Gamma$  contains, we prefer to compute the exact number of integers in  $\{1, \dots, n\}$  that it actually encloses prior to shrinking all cycles of  $\sigma$  into points. Formally, this means, give weight  $W(\mathcal{C}) = |\mathcal{C}|$  to any vertex  $\mathcal{C}$  of  $\Gamma$ , and ask what is the total weight of a connected component of  $\Gamma$ . Let  $\mathcal{C}_1(\Gamma)$  denote a size-biased pick from the connected components of  $\Gamma$ , i.e. the total weight of the connected component of  $\Gamma$  containing "1" (or, more precisely,  $\mathcal{C}_1(\sigma)$ ).

Lemma 22 shows that  $W(\mathcal{C}_1(\Gamma))$  converges in distribution to the total progeny of a branching process with offspring distribution a shifted geometric random variable. The idea is that by Theorem 28, the various cycles of  $\sigma$  are asymptotically i.i.d., so that each edge in  $\Gamma$  adds to the weight of  $\mathcal{C}_1(\Gamma)$  a contribution which is, by Theorem 28, asymptotically a geometric random variable  $\mathbf{G}$  with parameter  $1/(1 + b)$  where  $b$  satisfies  $\log(1 + b)/b = 1 - a$ . This seems to give an infinite progeny almost surely (since  $\mathbf{G} \geq 1$  a.s.). However, to every point that we examine there is a positive probability that it is a *terminal* point. In this case, that integer does not connect to a new independent cycle of  $\sigma$ , and hence its offspring is 0. This kind of modified branching process is defined more precisely and analyzed in section 3.8.3. The key fact is that because of the special properties of the asymptotic law of  $\pi$ , which involves geometric random variables, this modified branching process is in fact equal in distribution to another branching process where the offspring distribution has been shifted from  $\mathbf{G}$  to  $\mathbf{G} - 1$ . In all that

follows, we call  $X$  a random variable such that

$$X \stackrel{d}{=} \mathbf{G} - 1$$

Hence  $\Gamma$  has a giant component if, and only if,  $E(\mathbf{G}) > 2$ . Since  $p = 1/(1+b)$  and  $\log(1+b)/b = 1-a$ ,

$$P(\Gamma \text{ has a giant component}) > 0 \iff a > 1 - \log 2$$

## Proof of Theorem 33

### 3.8.1 Structure of the proof

As this proof is rather long, we feel that it is appropriate to explain how the various arguments are used. In section 3.8.2, we prove that  $W(\mathcal{C}_1(\Gamma)) \Rightarrow \sum_{t \geq 0} Z_t$  the total progeny of a branching process with offspring distribution  $X$ . Then in section 3.8.3, we define a modified branching process, and prove that in the case of geometric random variables this becomes another branching process with shifted offspring distribution. We then use this to prove by hand that in the subcritical case,  $|\mathcal{C}_1(\sigma \cdot \pi)|$  is dominated by such a modified branching process. Since this is a subcritical branching process, we prove the exponential decay of the tail of  $|\mathcal{C}_1(\sigma \cdot \pi)|$ , uniformly in  $n$ . This enables us to show that as long as  $a < 1 - \log 2$  there are very few fragmentation events in the walk  $(\sigma_r, r = 0, \dots, an)$ . The supercritical case is treated in section 3.8.5. Since we have established branching process asymptotics, we can use the duality principle of a branching process between the subcritical phase and supercritical phase. This shows that the number of clusters of  $\Gamma$  in the supercritical regime can be computed by looking at the number of clusters of  $\Gamma$  for some specific subcritical time. Since we have proved that the distance is linear in this regime, we now know how many clusters does  $\Gamma$  have at any subcritical time, and it follows that the number of clusters of  $\Gamma$  in the supercritical regime is strictly less than what it would be if the distance was still linear. It only remains to prove that at any given time the number of extra cycles that were generated by some fragmentation (and have not been reabsorbed by other large cycles) is  $O(n^{1/2})$ , which is done in section 3.8.6.

### 3.8.2 Branching process asymptotics

To start proving things, we need some more notations. Let  $A_0^n = \{1\}$  and define recursively the  $A_k^n$  by:

$$A_{k+1}^n = \bigcup_{x \in A_k^n} \{\mathcal{C}_{\pi(x)}(\sigma)\} - \bigcup_{1 \leq j \leq k} A_j^n$$

The  $A_k^n$  correspond to growing the branching process generation after generation, rather than cycle after cycle. Let  $(Z_t, t = 0, 1, \dots)$  be a branching process with offspring distributed as  $X$ . Note that by the construction of  $\Gamma$ , we also have that

$$\sum_{k=0}^{\infty} |A_k^n| = W(\mathcal{C}_1(\Gamma))$$

**Lemma 22.** *As  $n \rightarrow \infty$ ,*

$$(|A_0^n|, |A_1^n|, \dots) \Rightarrow (Z_0, Z_1, \dots)$$

*Proof.* Let us start by the convergence of  $(|A_0^n|, |A_1^n|)$ . If  $j = 0$ ,  $P(|A_0^n| = 1, |A_1^n| = 0) = P(\pi(1) = 1) \rightarrow 1/(1+p) := p = P(X = 0)$ . If  $j \geq 1$ , then

$$\begin{aligned} P(|A_0^n| = 1, |A_1^n| = j) &= P(\pi(1) \neq 1; |\mathcal{C}_{\pi(1)}(\sigma)| = j) \\ &= P(\pi(1) \neq 1) \cdot P(|\mathcal{C}_{\pi(1)}(\sigma)| = j | \pi(1) \neq 1) \\ &\rightarrow (1-p) \cdot (1-p)^{j-1} p = P(X = j) \end{aligned}$$

Indeed, conditionally on  $\{\pi(1) \neq 1\}$ ,  $\pi(1)$  is uniform on  $\{2, \dots, n\}$ , so that  $\mathcal{C}_{\pi(1)}(\sigma)$  is as good a size-biased pick as  $\mathcal{C}_1(\sigma)$ . Theorem 4 this observation.

Now let us consider the general case of finite-dimensional distributions. Let  $n_1 > 0, n_2 > 0, \dots, n_k \geq 0$  with  $\sum_i n_i \leq n$ . We are trying to compute the asymptotics of

$$P(|A_0^n| = 1, |A_1^n| = n_1, \dots, |A_k^n| = n_k)$$

To do this, we need to evaluate the probability of a collision occurring in the first  $k$  stages, that is,

$$P\left(\pi(x) \in \bigcup_{1 \leq i \leq k} A_i^n - \mathcal{C}_x(\pi) \text{ for some } x \in A_j^n \text{ with } j \leq k\right)$$

We will say of an  $x$  such as in the event above, that it makes a *backward* connection. Hence an  $x$  makes a backward connection if  $\pi(x)$  maps it to some lower level in the branching process, but  $x$  is not a terminal point. Therefore backward connections (or collisions) are exactly those that may lead to a fragmentation, as explained in the sketch of the proof.

It is easy to see that  $P(\text{collision in first } k \text{ stages}) = O(1/n)$ . In fact, it follows from the uniformity of Lemma 23 that:

$$\begin{aligned} P(|A_0^n| = n_0, \dots, |A_k^n| = n_k; \text{b.w. collisions in } k \text{ first stages}) &\leq \sum_{j=1}^k n_j \frac{\sum_{i=0}^{j-1} n_i}{n} \\ &\leq \frac{(\sum_{i=0}^k n_i)^2}{n} \end{aligned}$$

(see also Lemma 24 where similar estimates are derived.)

Therefore it is enough to consider

$$P(|A_0^n| = 1, |A_1^n| = n_1, \dots, |A_k^n| = n_k | \text{given no backw. connection})$$

Suppose  $A_{k-1}^n = \{x_1, \dots, x_{n_{k-1}}\}$ . Conditionally on the event that there is no collision in the  $k$  first stages,  $\pi(x_1), \dots, \pi(x_{n_{k-1}})$  belong to yet unexplored cycles, (as long as they are not terminal). After decomposition on the number of such  $x$ , the last probability is equal to

$$\begin{aligned} &= \sum_{j=1}^n P\left(\sum_{i=1}^{n_{k-1}-j} |\mathcal{C}_{\pi(x_i)}(\sigma)| = n_k\right) P(\#\{x \in A_{k-1}^n, x \text{ terminal}\} = j | \text{no bw conn.}) \\ &\rightarrow \sum_{j \geq 0} P\left(\sum_{i=1}^{n_{k-1}-j} X_i = n_k\right) p^j \\ &= P\left(\sum_{i=1}^{n_{k-1}} X_i = n_k\right) \end{aligned}$$

by the asymptotic independence property of a finite number of size-biased cycles, and the fact that given there was no backward connection, the  $x$  in level  $A_{k-1}^n$  belong to different cycles of  $\pi$ , so that the events that they are terminal are independent asymptotically.

These are the transition probabilities of a branching process with offspring distribution  $X$ , so the Lemma is proved.  $\square$

### 3.8.3 A modified branching process

One way to formalize the idea that a vertex has a geometric number of children only during finitely many generations, is to use a modified branching process where each individual  $x$  is endowed with a nonnegative, integer-valued random variable  $T(x)$ , that represents the "life-time" of its family. As long as  $T(x) > 0$ ,  $x$  will keep having children according to the original offspring progeny  $L$ . But when  $T(x) = 0$ , the individual will be declared "terminal" and will not be allowed to have any children.

Here is a rigorous description of this modified branching process. Let  $X_{t,i}$  be a collection of i.i.d random variables with distribution  $L$ , a fixed distribution on the nonnegative integers (the original progeny). Let  $T_{t,i}$  be i.i.d nonnegative integer-valued random variables, distributed according to another distribution  $L'$ , the life-time. Let  $Z_t$  be the size of the process at time  $t$  (with discrete time). Define  $Z_0 = 1$ , and give the root life-time  $T_{0,0}$ . Then define recursively  $Z_t$  by

$$Z_{t+1} = \sum_{i=0}^{Z_t} X_{t,i} \mathbf{1}_{\{T(x_i) > 0\}} \quad (3.10)$$

where  $x_1, \dots, x_{Z_t}$  are the  $Z_t$  individuals of generation  $t$ . If  $y_1, \dots, y_{Z_{t+1}}$  are the  $Z_{t+1}$  individuals of generation  $t + 1$ , the rule that we adopt for the value of  $T(y_1), \dots, T(y_{Z_{t+1}})$ , is the following. If  $T(x_i) > 0$  give all  $X_{t,i}$  children of  $x_i$  independent life-times from  $T_{t,i}$ , except for one of its children, say  $y_j$ , for which  $T(y_j) := T(x_i) - 1$ . Rigorously, let  $N_t = \#\{i : T(x_i) > 0\}$ , rewrite the  $x_i$ 's removing the terminal ones, and call them  $x'_1, \dots, x'_{N_t}$ . Let  $T(y_1) = T(x'_1) - 1, \dots, T(y_{N_t}) = T(x'_{N_t}) - 1$ , and let  $T(y_{N_t+1}) = T_{t+1, N_t+1}, \dots, T(y_{Z_{t+1}}) = T_{t+1, Z_{t+1}}$ .

Of course we make this definition because asymptotically,  $W(\mathcal{C}_1(\Gamma))$  will be well-approximated by such a system, where the offspring  $L$  is the size of cycle of  $\sigma$ , and where  $L'$  is the size of a cycle of  $\pi$ . Indeed, suppose we are exploring the cluster containing 1 in the graph of the superposition of  $\sigma$  and  $\pi$ .  $T(1)$  is  $|\mathcal{C}_1(\pi)|$ , which corresponds to the fact that after that many iterations of  $\pi$  we are back to where we started and no longer add anything new to the population of the cluster. However, after one iteration say, all vertices in the first generation, other than  $\pi(1)$  itself, belong to different cycles of  $\pi$  with high probability. Therefore their lifetime should be an independent random variable, distributed as  $L'$ .

In general, the ageing branching process  $(Z_t, t = 0, 1, \dots)$ , where each individual has a "life-time" that it transmits to one of its children, is not a Markov process with respect to its own filtration  $\sigma(Z_0, Z_1, \dots)$ . Indeed the size of the generation  $t + 1$  not only depends on the size of generation  $t$ , but also on the random variables  $T_x$  where  $x$  is an individual of generation  $t$ , so one would need to add in the filtration the values of  $T(x)$  for each generation.

However, a miracle happens due to the fact that the cycles of  $\pi$  have (asymptotically) a geometric distribution  $\mathbf{G}$ . Let  $p'$  be the parameter of  $\mathbf{G} : P(\mathbf{G} = j) = (1 - p')^{j-1}p'$ . Then the distribution  $L'$  of the random variables  $T_{t,i}$  is again  $\mathbf{G}$ . For  $k \geq 1$ , conditionally on  $T > k$ ,  $T - k$  is distributed as  $\mathbf{G}$ . This fact, called "lack of memory", has the following amazing consequence:

**Proposition 2.** *When the lifetime  $L'$  is a geometric random variable,  $(Z_t, t = 0, \dots)$  is a Markovian branching process with offspring distribution  $L\mathbf{1}_{\{L' > 0\}}$ . When  $L \stackrel{d}{=} L' \stackrel{d}{=} \mathbf{G}$ , this distribution is  $X \stackrel{d}{=} \mathbf{G} - 1$ .*

*Proof.* Let  $B_{t,i}$  be Bernoulli random variables with success parameter  $P(B_{t,i} = 1) = p'$ . Because  $\mathbf{G} =_d \inf\{t \geq 0; B_{t,i} = 1\}$ , the event  $\{T(x_i) > 0\}$  is the same as  $\{B_{t,i} = 0\}$ , so (3.10) becomes:

$$Z_{t+1} = \sum_{i=0}^{Z_t} X_{t,i} \mathbf{1}_{\{B_{t,i}=0\}} \quad (3.11)$$

This expresses the fact that for each new vertex visited, we can take the decision of closing the cycle, independently of the past. When the cycle still has some length to be explored, then the vertex has  $X_{t,i}$  children. This decision affects the law of progeny at a given vertex. The new distribution of the progeny is now, by (3.11):

$$P(X = j) = \begin{cases} p' & \text{if } j = 0 \\ (1 - p')P(X_{t,i} = j) & \text{if } j \geq 1 \end{cases} \quad (3.12)$$

Of course, for our problem,  $\sigma =_d \pi$ , so both  $L$  and  $L'$  are distributed as  $\mathbf{G}$ . As can be readily checked from (3.12), the distribution of  $X$  is thus a shift of  $\mathbf{G}$ :

$$X \stackrel{d}{=} \mathbf{G} - 1 \quad (3.13)$$

□

### 3.8.4 Fragmentations in the subcritical case

First note that if  $\sigma$  is a uniform permutation on  $\partial B(an)$ , if we visit all points in  $\{1, \dots, n\}$  according to their order of appearance in the canonical cyclic decomposition of  $\sigma$ , and call this process  $V_t (0 \leq t \leq n)$ , then the successive points are in some sense uniformly chosen from what remains to be found, at least as long as we don't have to start a new cycle. More precisely:

**Lemma 23.** *Given  $(V_0, \dots, V_t)$  and given that  $V_t$  is not a terminal point,*

$$\sigma(V_t) \text{ is uniform on } \{1, \dots, n\} - \{V_0, \dots, V_t\}$$

The proof of this lemma follows directly from the Feller coupling presentation of a uniform permutation on  $G_n$ , which also has (obviously) this property. Conditioning on the number of cycles does not change how the cycles are filled in.

**Lemma 24.** *Suppose the branching process is subcritical, i.e.  $p > 1/2$  or (equivalently)  $a < 1 - \log 2$ . Then the number of fragmentations in the walk  $(\sigma_r, r = 0 \dots n)$  is  $o(n)$ .*

Basically, all cycles are fairly small, so by improving our estimates on the number of collisions, we should get an  $O(1)$  bound, just like in the Erdős-Renyi case. Technicality arises due to the fact that the cycles are conditioned independent random variables, and not just independent. Here is a rigorous proof.

*Proof.* We prove things in two steps.

First, we prove a uniform bound for the size of a cluster: we show that if we write  $\pi = \prod_{i=1}^{an} \tau_i$ , and denote by  $\pi_r := \prod_{i=1}^r \tau_i$

$$P(|\mathcal{C}_1(\sigma \cdot \pi_r)| > u \text{ for some } r \leq an) \leq Cn \exp(-\alpha u) \quad (3.14)$$

where  $C$  and  $\alpha$  are constants independent of  $n$ , and  $u$  is any number.

Once this exponential control is proved, we can bound the number of times that one of the  $\tau_r$ 's will yield a fragmentation. Indeed, recall that to obtain  $\sigma \cdot \pi$  we can perform successively the  $\tau_r$ 's on  $\sigma$ , and each one yields a coagulation or a fragmentation. We hence view this a process indexed by  $1 \leq r \leq an$ . In the course of this process, at all times, by (3.14) applied to  $u = (\log n)^2$ ,

*no cluster* is larger than  $(\log n)^2$  with overwhelming probability, so that by Lemma 23:

$$P(\tau_{r+1} \text{ yields a fragmentation}) \leq 2(\log n)^2/n$$

There are (exactly)  $an$  transpositions to perform, hence:

$$E(\#\text{frag.}) \leq 2a(\log n)^2$$

This is already largely enough to prove Lemma 24.

We will now prove that (3.14) holds, since this is the only thing that remains to be proved. Although we have seen that in the limit each cluster is a subcritical branching process, (for which such an exponential tail of the total progeny holds), when  $n$  is finite there is no real branching process available to dominate  $\mathcal{C}_1(\sigma \cdot \pi_r)$ , essentially because the size of the cycles are not i.i.d. random variables. However, they are conditionally independent, (cf. Theorem 34, or Theorem 28), and we will use this fact to construct a real branching process that dominates  $\mathcal{C}_1(\sigma \cdot \pi_r)$ , when conditioned on some mild event. This conditioning accounts for the extra factor  $n$  in (3.14).

Here is how we proceed. By Kolchin's representation theorem (Theorem 34), there are random variables  $(X_1, \dots, X_{n(1-a)})$  such that the joint law of the size of the cycles of  $\sigma$  is  $(X_1, \dots, X_{n(1-a)})$  given  $\sum_i X_i = n$  (we will call  $A_n$  the event that  $\sum_i X_i = n$ ). The  $X_i$ 's constitute a "pool" of possible cycle sizes. Similarly, there are random variables  $(Y_1, \dots, Y_{n(1-a)})$  such that the joint law of the sizes of the cycles of  $\pi$  is  $(Y_1, \dots, Y_{n(1-a)})$  given  $\sum_i Y_i = n$  (let  $B_n$  be the event that  $\sum_i Y_i = n$ ).

We give an upper bound of  $\mathcal{C}_1(\sigma \cdot \pi)$  in terms of the modified branching processes of section 3.8.3, that uses only the  $X_i$ 's and the  $Y_i$ 's. Start with vertex 1 and chose a size-biased pick  $X'_1$  of the  $X_i$ 's (the cycle containing 1). Put  $T(1) = Y'_1$ , a size-biased pick of the  $Y_i$ 's. Next, given  $X'_1 = k$ , put  $T(2) = Y'_2, \dots, T(k) = Y'_k$ . All vertices with positive lifetime  $T$  have a number of children given by a size-biased pick of the remaining  $X_i$ 's. They transmit their lifetime -1 to one of their children and the rest have lifetimes given by size-biased picks from the remaining  $Y_i$ 's. Then repeat the procedure until we cannot go any further (i.e., until all vertices at a given generation have lifetime  $T = 0$ , or until all  $X_i$ 's  $Y_i$ 's have been picked). Call  $Z'$  the total population obtained at the end of this construction.

We claim that  $Z'$  dominates all stages of  $\mathcal{C}_1(\sigma \cdot \pi_r)$ , because  $Z'$  gives the cycles of  $\sigma$  coagulated by those of  $\pi$ , without taking any account of eventual fragmentations. In particular, in  $Z'$ , as long as a vertex  $x$  is not *terminal* ( $T(x) > 0$ ), the children of  $x$  will be part of the population of  $Z'$ . Of course in the event of a collision or a backward connection,  $Z$  does not contain any additional children, so that  $Z < Z'$ . Therefore

$$\begin{aligned} P(Z > u) &\leq P(Z' > u | A_n \text{ and } B_n) \\ &\leq P(A_n)^{-1} P(B_n)^{-1} P(Z' > u; A_n; B_n) \\ &\leq Cn P(Z' > u) \end{aligned}$$

Indeed, by the local central limit theorem (see Durrett (2004)),

$$P(A_n) = P(B_n) = P\left(\sum_{i=1}^{n(1-a)} X_i = n\right) \sim Cn^{-1/2}$$

To complete the proof, it remains to notice that size-biasing the logarithmic distribution of Kolchin's Theorem gives a geometric random variable. Therefore, by arguments already developed in the sketch of the proof,  $Z'$  is the total population of a branching process with offspring distribution  $X$  of (3.13), and starting with a geometric number of individuals  $\mathbf{G}$ . Because  $p > 1/2$ , this branching process is subcritical. In this case, classical estimates (Athreya and Ney (1972), Durrett (2004)), show the exponential tail

$$P(Z' > u) \leq C \exp(-\alpha u)$$

This concludes the proof of (3.14), and also that of Lemma 24.  $\square$

**Remark.** It is possible to avoid the use of Kolchin's representation Theorem in the above proof. Indeed, by Theorem 28, a size-biased pick of the cycles has, after unconditioning on some event of probability  $\propto n^{-1/2}$ , a distribution which is given by the lengths of sequences of 1 in the Bernoulli trials  $\zeta_i^{(\lambda)}$ . However, since  $\beta_i = P(\zeta_i^{(\lambda)} = 1) \leq b/(1+b)$ , it follows that the distribution of a size-biased cycle is thus (after unconditioning) stochastically dominated by the geometric random variable  $\mathbf{G}$ .

### 3.8.5 Mean in the supercritical regime and duality

Although this may seem surprising a little surprising at first, we use the result from the subcritical case to get that for the supercritical case. The idea is to use the duality of branching processes.

A crucial remark is that the number of cycles of  $\sigma \cdot \pi$  is given by  $\sum_{x=1}^n 1/|\mathcal{C}_x(\sigma \cdot \pi)|$ , hence by exchangeability

$$\frac{1}{n}E(\#\text{clusters of } \Gamma) = E\left(\frac{1}{|\mathcal{C}_1(\Gamma)|}\right) \rightarrow E\left(\frac{1}{T-1} | T > 1\right)$$

where  $T$  is the total progeny of a branching process with offspring distributed as  $X$  and started with 1 individual. Indeed, let us not forget that the first generation  $A_0$  of the branching process is itself a geometric random variable, so we can add an imaginary root and then subtract it (thus  $T-1$ ). Introducing an extra vertex for the root allows us to make use of the duality principle of branching processes. (Athreya-Ney (1972), Durrett (2005)).

The duality principle states that a supercritical branching process, conditioned on extinction, is another branching process, subcritical, whose offspring distribution is given through its generating function. If  $\phi(s) = E(s^X)$  is the generating function of  $X$  and  $\alpha$  is the extinction probability  $\alpha = P(T < \infty)$  then the conditioned process has offspring distribution characterized by

$$\phi'(s) = \phi(s\alpha)/\alpha$$

Here,  $P(X = j) = (1-p)^j p$ , so

$$\phi(s) = \frac{p}{1-s(1-p)}$$

Therefore

$$\phi'(s) = \frac{p/\alpha}{1-s\alpha(1-p)}$$

The fixed point equation for  $\alpha$  yields that

$$\frac{p}{\alpha} + \alpha(1-p) = 1 \tag{3.15}$$

so that  $\phi'$  is the Laplace transform of another shifted geometric random variable  $X'$ , with parameter  $p' = p/\alpha$ . Let  $T'$  be the total progeny of a branching process with offspring  $X'$  and started with 1 individual.

Let us now relate the supercritical and subcritical regimes. By duality,

$$\begin{aligned}
E\left(\frac{1}{T-1} \mid T > 1\right) &= \frac{P(T < \infty)}{P(T > 1)} E\left(\frac{1}{T-1} \mathbf{1}_{\{T > 1\}} \mid T < \infty\right) \\
&= \frac{P(T < \infty)}{P(T > 1)} E\left(\frac{1}{T'-1} \mathbf{1}_{\{T' > 1\}}\right) \\
&= \frac{P(T < \infty)}{P(T > 1)} P(T' > 1) E\left(\frac{1}{T'-1} \mid T' > 1\right)
\end{aligned}$$

However for the subcritical regime, we know by Lemma 24 that there are only  $o(n)$  fragmentations, so the distance between  $\sigma$  and  $\pi$  is  $2an + o(n)$  and the number of clusters is  $(1 - 2a)n + o(n)$ . Hence  $E\left(\frac{1}{T'-1} \mid T' > 1\right) = 1 - 2a'$ , where  $a'$  is the radius corresponding to the conditioned parameter  $p'$ . Since  $p = 1/(1 + b)$  and  $\log(1 + b)/b = 1 - a$ , we have that

$$a = 1 + \frac{p \log p}{1 - p}$$

On the other hand, due to the fixed point equation (3.15), the constant  $P(T' > 1)/P(T > 1) = (1 - p')/(1 - p)$  simplifies into  $\alpha$ .

Therefore, Theorem 33 is proved when we show that

$$\alpha^2(1 - 2a') > 1 - 2a$$

Using the fixed point equation (3.15), we find that  $a' = 1 + p \log p' / ((\alpha^2)(1 - p))$ , so that the above reduces to

$$-\alpha^2 - 2\frac{p \log p'}{1 - p} > -1 - 2\frac{p \log p}{1 - p}$$

or

$$\frac{2p \log \alpha}{1 - p} > \alpha^2 - 1$$

Using one more time the fixed point equation, one gets

$$2p \log \alpha > \alpha - 1$$

Since  $\log(1 - x) > -x$ , it is therefore enough to show

$$-2p(1 - \alpha) > \alpha - 1 \text{ or } 2p < 1$$

which is precisely the condition that the branching process is supercritical.

### 3.8.6 Fragmentations in the supercritical range

In the previous section we have computed asymptotics for the expected number of clusters in the graph resulting from the superposition of the cycle structures of  $\sigma$  and  $\pi$ . We now need to show that at the end of the walk  $(\sigma_r, r = 0, \dots, an)$ , there are no more than  $o(n)$  additional cycles that been generated by fragmentation, compared to the number of clusters of  $\Gamma$ .

To do this, we use once again the dynamic point of view adopted to deal with the subcritical regime. Let  $\tau_1, \dots, \tau_{an}$  be the decomposition of  $\pi$  in product of  $an$  transpositions as evoked earlier, and let  $\sigma_r = \sigma \cdot \tau_1 \dots \tau_r$ .

**Lemma 25.** *For each  $1 \leq r \leq an$ , the expected number of cycles in  $\sigma_r$  generated by fragmentation is  $O(n^{1/2})$ .*

This is similar to the Erdős-Renyi case of Berestycki and Durrett (2004), Theorem 3. Lemma 25 does not claim that the number of fragmentations itself is  $O(n^{1/2})$ , but that the number of extra cycles generated by fragmentation is  $O(n^{1/2})$ . Just like in the Erdős-Renyi case, many of the cycles that are fragmented get reabsorbed by large components fairly quickly.

*Proof.* There can never be more than  $n^{1/2}$  cycles of size larger than  $n^{1/2}$ . On the other hand, by Lemma 23, the probability that  $\tau_t$  will create a fragment of size smaller than  $n^{1/2}$  is at most  $n^{1/2}/n = n^{-1/2}$ . Therefore the expected number of such fragmentations is at most  $an \cdot n^{-1/2} = O(n^{1/2})$ .  $\square$

At this point, theorem 33 is proved.  $\square$

### Acknowledgements

I am very grateful to Laurent Saloff-Coste for introducing me to Gromov hyperbolic spaces and to Jim Pitman for pointing out the reference to Kolchin's representation theorem. I wish to thank Rick Durrett particularly for his help at several important stages of this work and for his support during my stay at Cornell University.

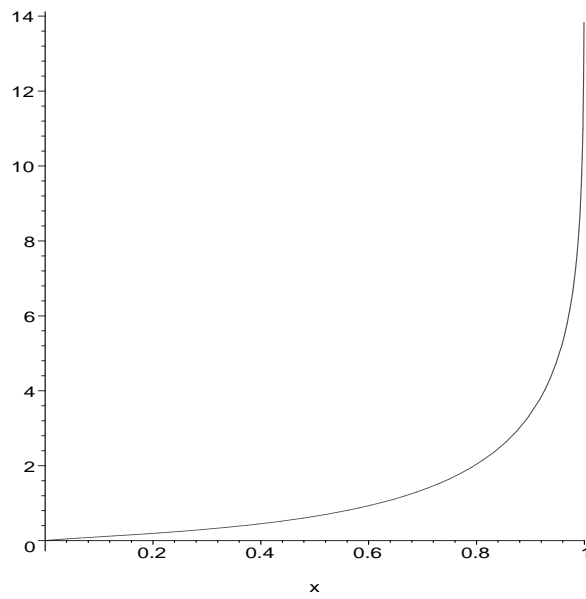


Figure 3.2: Numerical evaluation by Maple of the limiting behavior of the quantity  $-n^{-1} \log \frac{|S_n|}{|\partial B|}$ . The result is plotted as a function of  $\xi = f^{-1}(a)$ , where  $f(\xi) = 1 + \log(1 - \xi)(1 - \xi)/\xi$ .  $f^{-1}$  is an increasing function of  $a$ . The plot is only relevant for  $a < 1/2$  i.e.  $\xi < f^{-1}(1/2) \approx 0.715331863\dots$

# Chapter 4

## Limiting behavior of the distance of a random walk

by Nathanaël Berestycki<sup>1,2</sup> and Rick Durrett<sup>2</sup>

1. Ecole Normale Supérieure, D.M.A.
2. Cornell University, Department of Mathematics.

### 4.1 Introduction

Consider the symmetric group on  $n$  markers, say  $\{1, \dots, n\}$ , and suppose we want to study how this initial order is perturbed by randomly shuffling those elements according to some specific rules. For instance, one can think of repeatedly choosing two elements at random, and exchanging them (this is the so-called random transposition random walk). This is of course the model of card shuffling studied by Diaconis and many others, but we have another application in mind, namely the rearrangement of genome under the action of specific large-scale mutations, which involve changes within a chromosome at the scale of several genes.

When we think of card shuffling it is important to understand how many moves does it take to bring the deck of card close to random. However in the context of computational biology, a quantity of interest is the following:

given the current state of a genome, can we guess how many moves have been performed that have resulted in this specific configuration?

A first approach to this question consists of using what computer scientists refer to as a *parsimony* method, i.e., what is the *minimal* number of moves that have necessarily been performed to obtain this specific configuration? Often (give references), this provides an accurate answer. In some cases, it is even possible to find a polynomial algorithm which computes this *parsimony distance* between two given genome configurations.

However there are some cases where the parsimony approach does not give the right answer. This was first observed on simulations of Bourque and Pevzner (2002) and later confirmed theoretically by Berestycki and Durrett (2004). More precisely,  $X_t$  be the random transposition random walk on  $n$  markers. This means, in discrete time,  $X_k$  is the composition of  $k$  independent uniform transpositions  $(i, j)$ , or, in continuous time, that we perform a random uniform transposition at the jump times of some Poisson clock with intensity 1, and this model corresponds to exchanging two markers chosen uniformly at random as a rearrangement rule. Let  $D_t$  be the distance of  $X_t$  from its starting point, i.e. the *minimal* number of transpositions necessary to build  $X_t$  from transpositions (so,  $D_t$  is the *parsimony* distance of  $X_t$ ). The main result of Berestycki and Durrett (2004) is that  $D_t$  has a phase transition at time  $n/2$  as  $n \rightarrow \infty$ :

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

$$n^{-1}D_{tn} \rightarrow_p f(t)$$

where  $f(t)$  is defined by:

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

In particular, the parsimony method gives an accurate answer if (and only if) fewer than  $n/2$  transpositions have been performed, or equivalently, the observed parsimony distance is smaller than one half the number of markers. The authors then show that this result remains valid for random reversals, a much more realistic model for the inversions within a chromosome.

However the methods used to prove this result are very specific to this random walk, and it is an obvious question to ask whether this striking phase

transition is shared by other random walks. More generally, the purpose of the present paper is to see through a rather large set of examples what can be said about the distance  $D_{t\theta_n}$  of a random walk to its starting point at some time  $t\theta_n$  where  $\theta_n$  is a time-scale that must be chosen appropriately. We are particularly interested in situations where a phase transition similar to Theorem 0 may or may not occur.

**Organization of the paper.** In the next section we present our main results and try to explain where the results come from. Each of the following sections is devoted to a particular example (3-cycle random walk, random adjacent transpositions, random walk on a random 3-regular graph).

## 4.2 Summary of main results

We start by describing the results that we obtain for the random walks under consideration.

### 4.2.1 Random $p$ -cycles

A first walk that we consider is a simple generalization of the result for random transpositions. Suppose that the step distribution is now uniform on cycles of lengths  $p$ , where  $p \geq 3$  (so  $p = 2$  is exactly the case of random transpositions). To avoid useless complications we explain our results for  $p = 3$ . As Theorem 35 shows, this random walk displays a phase transition very similar to that of random transpositions.

For  $c > 0$ , let

$$u(c) = 1 - \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

For  $t > 0$  denote by  $D_t$  the distance to the identity of  $\sigma_t$ , where  $\sigma_t$  denotes the random walk which is a product of a Poisson number of 3-cycles.

**Theorem 35.** *Let  $\omega(n)$  be any sequence with  $\omega(n) \rightarrow \infty$  as slowly as we want. Then*

$$\frac{D_{cn} - u(c)n}{\omega(n)\sqrt{n}} \rightarrow 0$$

*in probability.*

Of course the function has similar characteristics as in the case of random transpositions for  $c < 1/6$ ,  $u(c) = c$ , it has no-second derivative at  $c = 1/6$  and  $u(c) < c$  for  $c > 1/6$ . The value  $c = 1/6$  for the phase transition is perhaps surprising. Indeed the naive reasoning that we are about to present suggests the value  $c = 1/4$  as a first guess. Indeed, any 3-cycle  $(a b c)$  can be decomposed as the product of 2 transpositions  $(a b c) = (a b)(b c)$ , so it seems at first that this random walk goes "twice as fast" as the random transposition, since there are "two steps at once". Of course, in this decomposition the two transpositions share a common point ( $b$ ), but intuitively this should actually slow down the walk - whereas in fact we observe the phase transition *before*  $c = 1/4$ , at  $c = 1/6$ .

Our proof sheds light on this rather surprising fact. We use random hypergraphs as a substitute for Erdős-Renyi random graphs. Random hypergraphs are graphs where edges ("hyperedges") may connect several vertices at the same time. The natural thing to do is then to couple the random walk with a hypergraph process, adding a hyperedge connecting  $a, b$  and  $c$  whenever the 3-cycle  $(a b c)$  is applied. The result of doing this is, morally, adding the edges  $(a, b)$ ,  $(b, c)$  *but also the edge*  $(c, a)$ : therefore this walk goes 3 times as fast as random transpositions, and this is where the value  $c = 1/6$  comes in. More generally, for a random  $p$ -cycle, the phase transition is located at time  $cn$  with  $c = 1/p(p-1)$ . Again, loosely speaking, this walk has been sped up by a factor of  $\binom{p}{2}$  compared to random transpositions.

## 4.2.2 Random adjacent transpositions

As we will see, the random adjacent transposition provides a first example of a walk which has a very different behavior, with different regimes but no sharp phase transition. The walk  $(X_t, t \geq 0)$  is once again a  $\mathcal{S}_n$ -valued Markov process where the step distribution of the discrete-time walk is uniform on transpositions  $(i, i+1)$  for  $1 \leq i \leq n-1$ . To evaluate the (adjacent) distance of a permutation  $\sigma$ , i.e., the minimum number of adjacent transpositions needed to build  $\sigma$ , or the graph distance between  $I$  and  $\sigma$  on the Cayley graph of  $\mathcal{S}_n$  generated by the set of adjacent transpositions, there is a very convenient formula

$$d_{\text{adj}}(\sigma) = \text{Inv}(\sigma) := \#\{1 \leq i < j \leq n : \sigma(i) > \sigma(j)\} \quad (4.1)$$

$\text{Inv}(\sigma)$  is called the number of inversions of  $\sigma$ .

Having in mind roughly the same applications to computational biology as those exposed earlier in the introduction, Eriksson et al. (2000) considered the problem of evaluating the distance after  $k$  random adjacent transpositions were applied. Relying heavily on formula (4.1) they were able to carry some explicit combinatorial analysis, to obtain various exact formulae for this expected distance, such as this one:

$$Ed_{\text{adj}}(X_{t(k)}) = \sum_{r=0}^k \frac{(-1)^r}{n^r} \left[ \binom{k}{r+1} 2^r C_r + 4d_r \binom{k}{r} \right] \quad (4.2)$$

where  $t(k) = \inf\{s > 0 : N_s = k\}$  is the time at which the  $k^{\text{th}}$  random adjacent transposition is applied,  $C_r$  are the Catalan numbers and  $d_r$  is some non-negative integer sequence. This formula is obtained by Eriksson et al. (2000), and then improved by Eriksen (2005) (who calculates explicitly the sequence  $d_r$  and also proves that the formula holds for a wider range of  $k$ ), based on the observation that the matrix with entries  $p_{i,j}(t) = P(X_t(i) > X_t(j))$  performs a discrete heat-flow process, a fact that will become obvious from our approach.

However getting useful asymptotics from exact formulae such as (4.2) seems an almost untractable task. On the other hand, we give a much more probabilistic approach, which has the advantage of yielding intuitive results about the asymptotic behavior of this walk. Also, in certain regimes our results go beyond beyond the simple expected value analysis.

Let us explain very briefly the starting point of our analysis. By (4.1), we get

$$d_{\text{adj}}(X_t) = \sum_{i < j} \mathbf{1}_{\{X_t(i) > X_t(j)\}}$$

On the other hand, if  $1 \leq i \leq n$  is fixed, the trajectory  $(X_t(i), t \geq 0)$  of the  $i^{\text{th}}$  particle under the action  $X_t$  is nothing but a continuous time simple random walk, on  $\{1, \dots, n\}$  with reflecting boundaries, starting at  $i$ , and with rate  $2/(n-1)$  except at the boundaries where the rate is  $1/(n-1)$ .

Taking two such trajectories, say those of particles  $i$  and  $j$  with  $i < j$ , they now both perform independent simple random walk, except when they are adjacent, because it is never the case that  $X_t(i) = X_t(j)$ . When they are adjacent, the only thing that can happen is an exchange of the two particles, or one of them has to leave away from the other. In other words, this couple of

particles performs simple exclusion on  $\{1, \dots, n\}$  (with reflecting boundaries and modified rates near the boundaries).

Taking expected values for instance, this simply becomes

$$Ed_{\text{adj}}(X_t) = \sum_{i < j} P^{i,j}(Y_t > Y'_t)$$

where the law of  $(Y_t, Y'_t)$  under  $P^{i,j}$  is just simple excursion of two particles started at  $i$  and  $j$ .

From this representation we can deduce the following result after some analysis.

**Theorem 36.** *Let  $t > 0$ . Then*

$$\frac{1}{n} Ed_{\text{adj}}(X_{nt}) \rightarrow f(t)$$

as  $n \rightarrow \infty$  for some explicit function  $f(t)$ .  $f(t)$  is infinitely differentiable, and moreover it has the asymptotic behavior

$$\lim_{t \rightarrow \infty} \frac{f(t)}{\sqrt{t}} = \frac{1}{2} E[\max_{0 \leq s \leq 1} B_{2s}] = \sqrt{2}/2$$

where  $B_t$  is a standard Brownian motion.

This should be regarded as a "diffusive behavior" type of result. The next result looks at the distance of the random walk on time-scales of the order  $n^3 t$ . On this time-scale, any two particles may have had a chance to be exchanged. Indeed, it takes about  $n^2$  unit of time for particles  $n$  units of space apart to meet, and recall that a given pair of particles evolves with rate  $O(1/n)$ , which accounts for the extra factor  $n$  in the scaling. Let  $p_t(u, v)$  denotes the transition function of 1-dimensional Brownian motion reflecting at 0 and 1.

**Theorem 37.** *Let  $t > 0$ .*

$$\frac{1}{n^2} d_{\text{adj}}(X_{n^3 t}) \rightarrow_p \int_0^1 du \int_u^1 dv \int_0^1 p_t(u, x) dx \int_0^y p_t(v, y) dy = \mathbf{P}[B_1(t) > B_2(t)]$$

where  $B_1$  and  $B_2$  are two reflecting Brownian motions started uniformly on  $0 \leq B_1(0) < B_2(0) \leq 1$  evolving independently.

### 4.2.3 Random walk on a random 3-regular graph

Here is an example of a random walk which has a phase transition in a context much different from the case considered so far of random walks on  $\mathcal{S}_n$ .

A *3-regular* graph is a graph where all vertices have degree equal to 3. If  $n$  is even, it is always possible to draw such a graph on  $n$  vertices, and a *random 3-regular graph* is a uniform member of such graphs. Bollobàs and de la Vega (1982) have given a representation of a random 3-regular graph in terms of random matchings. Rapidly, expand each vertex  $i$  into 3 vertices  $3i$ ,  $3i + 1$  and  $3i + 2$ . If  $n$  is even, we may consider a random matching of those  $3n$  vertices. A random 3-regular graph  $G_n$  is then obtained by collapsing back the groups of 3 vertices together and keeping the edges of the random matching. (In fact this only holds on the event that there are no self-loops or multiple edges in  $G_n$  but this can be shown to have positive probability asymptotically). Alternatively, a random 3-regular graph is also a special case of the Molloy and Reed (1995) graph model where vertices have a prescribed degree distribution.

We make use of these two (equivalent) representations of a random 3-regular graph. The result that we prove for the random walk  $(X_k, k \geq 0)$  on this graph is the following, assuming that  $X_0 = 1$ .

**Theorem 38.** *For fixed  $t > 0$*

$$\frac{d(X_{\lfloor t \log_2 n \rfloor})}{\log_2 n} \xrightarrow{p} f(t) := \sup \left( \frac{1}{3}t, 1 \right)$$

An intuitive description of a random 3-regular graph as seen from vertex 1 can be given as follows. Let us grow the cluster containing 1 by exposing successive vertices with increasing distance away from 1. As long as we have not exposed more than  $o(n)$  vertices this should look very much like a regular tree (i.e., each vertex has 2 edges going away from the root and 1 leading back towards the root). Therefore this phase roughly lasts until distance  $\log_2 n$ . After distance  $\log_2 n$  however, this tree approximation is no longer valid and there are many edges inter connecting the vertices.

Therefore, as long as we are below critical distance  $\log_2 n$ , we expect that  $d(X_k)$  evolves like a biased random walk on  $\mathbf{Z}_+$ , with transition probabilities  $p(x, x + 1) = 2/3$  and  $p(x, x - 1) = 1/3$ . After  $k$  moves we expect this walk to be at distance  $k(2/3 - 1/3) = k/3$ .

On the other hand, once the walk reaches a distance corresponding to the average diameter of the graph ( $\log_2 n$  by Bollobàs - de la Vega (1982), or Theorem 2.13 in Wormwald (2005)) it should remain at this level. Indeed it cannot go any further since this is the diameter, and on the other hand the tree structure below makes it hard for it to come down back toward the root.

As we already mentioned, this gives a lower bound on the time to reach equilibrium for a random walk on a random 3-regular graph. Indeed, the walk cannot have reached equilibrium before it has reached a distance corresponding to the average diameter of the graph. In fact, we conjecture that this lower bound gives the right answer.

**Conjecture 2.** *This lower bound is sharp: random walk on a random 3-regular graph mixes in time  $3 \log_2 n$ .*

Right now it is known that  $O(\log n)$  is the correct order of magnitude for the mixing time. More precisely, if we let  $\Delta(t) = \max_{i,j} \frac{|p_{i,j}^t - \pi_j|}{\pi_j}$  be the relative pairwise distance between the chain and the invariant (here uniform) measure  $\pi$ , then  $\Delta(b \log n) \rightarrow 0$  for  $b$  large enough. See Durrett (2005) for a proof.

#### 4.2.4 Informal speculation

Loosely speaking, the study of different examples coming from different contexts suggests that the behavior of the distance of a *reversible* random walk on a large but finite graph is characterized by the existence or the absence of a phase transition for some suitable time-scale. In all the examples that we study, transition phases for the distance are preceded by a phase where the distance evolves in a purely linear fashion (random transpositions, random 3-cycles, random 3-regular graph), and then followed by a sublinear phase, where the distance grows more slowly towards a limit which is the average distance in the graph.

On the other hand when such a phase transition does not happen the walk may exhibit several regimes but the behavior of the walk throughout each phase is then always (in all the examples we study) smooth and sublinear - actually, close in some sense to the classical diffusive behavior of random walks on  $\mathbf{Z}^d$ . In this category we find random walk on a large-dimensional hypercube and the composition of random adjacent transpositions.

This informal claim should be tempered by the fact that even in the simplest examples it is always possible to come up with a time-scale that is so rough that the distance becomes actually discontinuous. Requiring that the distance be continuous (for instance) on the time-scale considered allows one to take care of trivialities like that.

#### 4.2.5 Other random walks and some open problems

We hope to initiate a programme of identifying which random walks have a phase transition and which remain smooth. More generally, we view the analysis carried on the examples above as the start of a more complete investigation of what are the possible behaviors for the distance of a random walk. In particular, does there exist random walks with different behavior from one of the examples above, e.g. a walk with a smooth transition from a linear regime to a sublinear regime, or one with a non-smooth point but sublinear at all times.

Some more work is already going in this direction. For instance, Fulman (2004) studies the evolution of the distance when the random walk is the so-called Gilbert-Shannon-Reeds riffle shuffle. For this shuffling method (which, we emphasize, is *nonreversible*), Bayer and Diaconis (1992) proved an explicit formula for the distribution of  $\sigma_t$  after  $r$  shuffles. In particular from this formula it follows that the riffle-shuffle distance of a permutation is given by

$$d(\sigma) = \lceil \log_2(\text{Des}(\sigma) + 1) \rceil \text{ where } \text{Des}(\sigma) = \#\{1 \leq i \leq n-1 : \sigma(i) > \sigma(i+1)\}$$

$\text{Des}(\sigma)$  is called the number of descents of  $\sigma$ . The main result of Fulman (2004) is

**Theorem 39.** Fulman (2004). *After  $r = \log_2(\alpha n)$  shuffles*

$$\frac{1}{n} E(\text{Des}(\sigma_r)) \rightarrow \alpha - \frac{1}{e^{1/\alpha} - 1}$$

*at least if  $\alpha > 1/(2\pi)$ .*

This says that for this range of  $r$  the walk is already in a sublinear regime. In particular, as  $\alpha \rightarrow \infty$  we get that  $d(\sigma_r) \sim \log_2 n$  since the expression  $\alpha - \frac{1}{e^{1/\alpha} - 1}$  is asymptotic to  $1/2$ . This is not surprising since this is the diameter of the graph. It is not clear at this point whether this formula also

holds for smaller values of  $\alpha$ , although it is tempting to let  $\alpha \rightarrow 0$  and get that for small values of  $\alpha$  the walk is "almost" linear (the fraction term with the exponential is much smaller than the other term).

Also, the techniques developed for the random walk on a 3-regular graph should be very useful when dealing with graph which have a strong "tree"-like geometry. For instance, the giant cluster of a supercritical Erdős-Renyi random graph should be another example where the random walk exhibits a phase transition. Essentially, the arguments will be the same, except that there is no available good upper bound in the supercritical regime of the walk. Indeed for this we would need good estimates on the (topological) diameter of the giant component. This was easy in the random 3-regular graph (it is a 20-year old result of Bollobàs and de la Vega), but in the supercritical random graph this is an important open problem (see Chung and Lu (2001) for partial results). Indeed it is not even known if the diameter of the graph is realized on the giant component or on some "small" cluster!

## 4.3 Random $p$ -cycles

### 4.3.1 Random hypergraphs, definitions and notations

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

**Definition 6.** *A hypergraph is given by a set  $V$  of vertices and a subset  $E$  of  $\mathcal{P}(V)$ , where  $\mathcal{P}(V)$  denotes the set of all subsets of  $V$ . The elements of  $E$  are called hyperedges. A  $d$ -regular hypergraph is a hypergraph where all edges connect  $d$  vertices, i.e. for all  $e \in E$ ,  $|e| = d$ .*

We call  $\mathbf{G}_d(n, M)$  the probability distribution on  $d$ -regular hypergraphs on  $V = \{1, \dots, n\}$  which is the uniform distribution over all hypergraphs on  $V = \{1, \dots, n\}$  with exactly  $M$  hyperedges.

So when  $d = 2$  this is just the usual Erdős-Renyi random graph case, since a hyperedge connecting two vertices is nothing else than a usual edge.

In the study of Erdős-Renyi random graphs, and in coupling those graphs with the random transposition random walk, the notion of a complexity of a component played an important role. In particular, trees were the simplest

possible connected components and outnumbered all other type of components by a large amount, a fact that was crucial for counting the number of connected components of a random graph. The correct way to bring this concept in this context is to define the *excess* of a subgraph  $H$  of a hypergraph.

**Definition 7.** *The excess  $ex(H)$  of a given subgraph  $H$  of a  $d$ -regular hypergraph  $G$  is the quantity*

$$ex(H) = (d - 1)s - r$$

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

This definition finds its justification in the fact that if  $H$  is connected then  $ex(H) \geq -1$ . Indeed, each hyperedge adds at most  $(d - 1)$  points to the component, except for the first edge considered which brings in  $d$  vertices. Hence if any two hyperedges don't intersect at more than one vertex at a time the quantity  $(d - 1)s - r$  is exactly equal to  $-1$ . In this situation we will say that  $H$  is a *tree* (or a *hypertree* if we want to emphasize the difference with usual random graphs). Otherwise,  $ex(H) \geq 0$ . If  $ex(H) = 0$  and  $H$  is connected we will say that  $H$  is *unicyclic* and if the excess is positive we will say that the component is *complex*.

**Remark.** We write this with some care because there is at least one other natural way to define what a tree is in this context, which is to say that  $H$  is a tree if it is a *minimally* connected set, i.e. removing any hyperedge would make  $H$  disconnected. As can be easily seen from the discussion above, the two notions do not coincide when  $d \geq 3$ . Karoński and Łuczak in their paper of (2002) also use Definition 7 for a hypertree, but in their older paper of (1993) they use the minimal connectedness notion. For the particular application we have in mind, it is preferable to use Definition 7.

### 4.3.2 Critical point for random hypergraphs

In classical Erdős-Renyi random graphs, it is well known that for  $M = cn/2$  and  $c > 0$  the behavior is much different according to the position of  $c$  with respect to 1. Namely, when  $c < 1$  the graph consists *a.a.s.* (i.e., with probability 1 asymptotically) of trees and unicyclic components, all smaller than  $O(\log n)$ . When  $c > 1$  there is a single *giant component*, that is a component whose size is approximately  $\theta n$  for some  $\theta > 0$ . The giant component

has a complexity of order  $n$ , while all other components are not larger than  $O(\log n)$  and are either trees or unicyclic (all these properties are of course only valid *a.a.s.*). Here a similar phenomenon has been proved by Karoński and Luczak (2002).

**Theorem 40.** *Let  $M = cn$ , and consider a realization of  $\mathbf{G}_d(n, M)$ .*

- *When  $c < c_d = 1/(d(d-1))$  then *a.a.s* there are only trees and unicyclic components. The largest component is  $O(\log n)$ .*
- *When  $c > c_d$  then there is *a.a.s* a unique complex component, of size  $\theta_d(c)n$  asymptotically. all other component are not larger than  $O(\log n)$ .*

This result contains the Erdős-Renyi theorem as a particular case ( $d = 2$ ). We complement this result with a result on the number  $N_{\text{clusters}}$  of connected components of  $G_d(n, M)$  when  $M = cn$ . We believe this result to be new. To avoid useless complications we specialize from now on to the case  $d = 3$ .

**Theorem 41.** *The number of clusters  $N_{\text{clusters}}$  in a random 3-regular hypergraph  $\mathbf{G}_3(n, M = cn)$  satisfies*

$$\frac{1}{n} N_{\text{clusters}} \rightarrow_p \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

**Remark.** Differentiating twice and using Stirling's formula we see that the function of the variable  $c$ ,  $c \mapsto \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$  is differentiable but has no second derivative at  $c = 1/6$ .

*Proof.* By Theorem 40 it is enough to count the number of trees in  $\mathbf{G}_3(n, M)$ . We will first compute the expected value and then prove a law of large numbers using a second moment method.

Let  $s \geq 0$ , we first compute the number of trees with  $s$  hyperedges ( $s = 0$  corresponds to isolated vertices). These have  $r = 2s + 1$  vertices. By Karoński-Luczak (1997), there are

$$\frac{(2s)!(2s+1)^{s-1}}{s!2^s}$$

trees on  $r = 2s + 1$  vertices (this is the equivalent of Cayley's (1889) formula that there are  $k^{k-2}$  ways to draw a tree on  $k$  labelled vertices).

If  $T$  is a given hypertree with  $s$  edges, there are a certain number of conditions that must be fulfilled in order for  $T$  to be one of the components of  $G$ : (i)  $s$  hyperedges must be open, (ii)  $\binom{r}{d} - s$  hyperedges must be closed on the inside of the tree, (iii)  $T$  must be disconnected from the rest of the graph, which requires closing  $r\binom{n-r}{d-1}$  hyperedges.

Now, just like in Erdős-Renyi random graphs, we first assume that each given hyperedge is open independently of the others, and the probability of its being open is

$$p = 1 - \exp\left(-\frac{cn}{\binom{n}{3}}\right) \sim 6c/n^2$$

Indeed if show this for the *Binomial* model (where edges are independent) the result will follow for the uniform model  $\mathbf{G}_d(n, M)$  by considering that the edges of  $\mathbf{G}_d(n, p)$  arrive according to a Poisson point process. Conditioning on the time of the  $M^{\text{th}}$  arrival and using the central limit theorem for Poisson variables will conclude the proof in the case of the uniform model. (This last complication is useless for the application of the Theorem that we have in mind since on our case hyperedges *do* arrive with a Poisson point process).

Therefore the probability that  $T$  is a component of  $G$  is

$$p^s(1-p)^{\binom{r}{d}-s+r\binom{n-r}{d-1}}$$

Hence the expected number of trees in  $G$  with  $s$  edges is

$$\binom{n}{r} \frac{(2s)!(2s+1)^{s-1}}{s!2^s} p^s(1-p)^{\binom{r}{d}-s+r\binom{n-r}{d-1}} \quad (4.3)$$

On the other hand

$$(1-p)^{\binom{r}{d}-s+r\binom{n-r}{d-1}} \sim \exp\left(-r\binom{n-r}{2}\frac{6c}{n^2}\right) \sim \exp(-3rc) = \exp(-6c(s+1/2))$$

So, after simplification, summing over all possible values of  $s$  gives by Fatou's Lemma

$$\liminf_{n \rightarrow \infty} \frac{1}{n} E(N_{\text{clusters}}) \geq \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

In the other direction we note that for every  $\varepsilon > 0$  we can choose  $s_0$  such that the infinite series  $\sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$  can be approximated by the sum of its first  $s_0$  terms within  $\varepsilon$ . Then by (4.3) it suffices to show that

$$\frac{1}{n} \sum_{s=s_0+1}^{C \log n} \binom{n}{r} \frac{(2s)!(2s+1)^{s-1}}{s!2^s} p^s (1-p)^{\binom{r}{3}-s+r} \binom{n-r}{2}$$

can be made arbitrarily small (indeed by Theorem 40 there is at most one cluster larger than  $C \log n$ , namely the giant component). This is done by using Stirling's formula and proving an exponential decay for the terms in the series when  $c$  is not at the critical value  $c = 1/6$ . When  $c = 1/6$  we note that the number of clusters is increasing so we easily bound  $N_{\text{clusters}}$  from both sides by its subcritical and its supercritical one. Since the function  $c \mapsto \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$  is continuous, this proves the asymptotics

$$\frac{1}{n} E(N_{\text{clusters}}) \rightarrow \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

To get the convergence in probability and not just in expectation, we use a second moment estimate and remark that

$$N_{\text{clusters}} = \sum_{i=1}^n \frac{1}{|\mathcal{C}_i|}$$

where  $\mathcal{C}_i$  is the cluster containing  $i$ . In fact it is enough to prove that for every  $s_0 < \infty$ ,  $\text{var}(N_{\text{clusters}}^{\leq s_0}) = o(n^2)$  where

$$N_{\text{clusters}}^{\leq s_0} = \#\{\text{clusters with no more than } 2s_0 + 1 \text{ vertices}\}$$

(We adopt this definition since those clusters have usually no more than  $s_0$  hyperedges and correspond to the terms with  $s \leq s_0$  in the sum).

Indeed, if this we assume  $\text{var}(N_{\text{clusters}}^{\leq s_0}) = o(n^2)$ , and if we denote by  $m = \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$  is the mean number of clusters, then we have by Markov's inequality for any  $\varepsilon > 0$

$$\begin{aligned} P(N_{\text{cluster}} > nm(1 + \varepsilon)) &\leq P\left(N_{\text{clusters}}^{> s_0} > \frac{nm\varepsilon}{2}\right) + P\left(N_{\text{clusters}}^{\leq s_0} > nm\left(1 + \frac{\varepsilon}{2}\right)\right) \\ &\leq \frac{E(N_{\text{clusters}}^{> s_0})}{nm\varepsilon/2} + o(1) \end{aligned}$$

since  $N_{\text{clusters}}^{\leq s_0}$  has mean smaller than  $nm$  and we have assumed a variance condition which implies by Chebyshev's inequality that it satisfies a weak law of large numbers. On the other hand it was proved above that

$$E(N_{\text{clusters}}^{> s_0}) \sim n \sum_{s=s_0+1}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

In particular,

$$\limsup_{n \rightarrow \infty} P(N_{\text{cluster}} > nm(1 + \varepsilon)) \leq (m\varepsilon/2)^{-1} \sum_{s=s_0+1}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$$

which can be made as small as desired when  $s_0 \rightarrow \infty$ . With similar methods one can show that under the assumption that  $\text{var}(N_{\text{clusters}}^{\leq s_0}) = o(n^2)$

$$\limsup_{n \rightarrow \infty} P(N_{\text{cluster}} < nm(1 - \varepsilon)) = 0$$

which proves the convergence of probability of Theorem 41. Hence let us know prove that for any  $s_0$  we have  $\text{var}(N_{\text{clusters}}^{\leq s_0}) = o(n^2)$ .

If we let  $f_i = \frac{1}{|C_i|} \mathbf{1}_{\{|C_i| \leq 2s_0+1\}}$  to ease notations

$$\text{var}(N_{\text{clusters}}^{\leq s_0}) = n \text{var}(f_1) + \binom{n}{2} \text{cov}(f_1, f_2)$$

Hence it suffices to show that  $\text{cov}(f_1, f_2) \rightarrow 0$ . On the other hand, we can separate the covariance on the events that 1 and 2 are in the same cluster or not: if  $m = E(f_i)$

$$\begin{aligned} \text{cov}(f_1, f_2) &\leq \text{var}(f_1 \mathbf{1}_{\{1 \text{ and } 2 \text{ are in the same cluster}\}}) \\ &\quad + E((f_1 - m)(f_2 - m) \mathbf{1}_{\{1 \text{ and } 2 \text{ are in different clusters}\}}) \end{aligned}$$

The first term is easy to deal with since this is smaller than

$$P(|C_1| \leq s_0 \text{ and } 2 \in C_1) \leq s_0/n \rightarrow 0$$

The second term is *a priori* more delicate. To see how we control it we rewrite it as (with  $C = C_1$  and  $C' = C_2$ , and  $1 \leftrightarrow 2$  denotes the event that 1

and 2 are not in the same cluster)

$$\begin{aligned}
E((f_1 - m)(f_2 - m); 1 \leftrightarrow 2) &= E(f_1 f_2; 1 \leftrightarrow 2) - 2mE(f_1; 1 \leftrightarrow 2) + m^2 P(1 \leftrightarrow 2) \\
&= \sum_{k, j \leq 2s_0+1} \frac{1}{kj} P(C = k; C' = j; 1 \leftrightarrow 2) \\
&\quad - \sum_{k, j \leq 2s_0+1} \frac{1}{kj} P(C = k) P(C' = j) \\
&\quad + O(P(1 \leftrightarrow 2 \text{ and } C \leq s_0))
\end{aligned}$$

On the other hand,  $P(C = r; C' = r'; 1 \leftrightarrow 2)$  is given by a formula similar to (4.3), which is

$$\begin{aligned}
&\binom{n}{r-2} \frac{(2s)!(2s+1)^{s-1}}{s!2^s} p^s (1-p)^{\binom{r}{d}-s+r\binom{n-r}{d-1}} \\
&\times \binom{n-r-1}{r'} \frac{(2s')!(2s'+1)^{s'-1}}{s'!2^{s'}} p^{s'} (1-p)^{\binom{r'}{d}-s'+r'\binom{n-r'}{d-1}}
\end{aligned}$$

from which it is obvious that this is asymptotic to  $P(C = r)P(C' = r')$ , since  $r$  and  $r'$  are fixed. Therefore  $\text{cov}(f_1, f_2) \rightarrow 0$  and  $\text{var}(N_{\text{clusters}}^{\leq s_0}) = o(n^2)$ .  $\square$

### 4.3.3 Bounds for the Cayley distance on the symmetric group

Let  $(\sigma_t, t \geq 0)$  be the random 3-cycle walk. This is the walk on  $\mathcal{S}_n$  whose discrete-time walk has transition probabilities

$$p(\sigma, \pi) = \begin{cases} \frac{3}{\binom{n}{3}} & \text{if } \pi\sigma^{-1} \in S \\ 0 & \text{otherwise} \end{cases}$$

for  $S = \{3\text{-cycles}\}$ . In words, at rate 1 we choose a 3-cycle  $(a b c)$  uniformly among all possible such 3-cycles, which we apply on the current state of the permutation (i.e.,  $\sigma_t = (a b c) \cdot \sigma_{t-}$ ). In order to prove a phase transition for  $\sigma_t$  it will be necessary to have a formula to work with for the distance of a permutation  $\sigma$  to the identity element. Recall that the distance  $d_{3\text{-cycle}}(\sigma)$  of a permutation  $\sigma$  to the identity is the minimal number of 3-cycles needed to construct  $\sigma$  starting at the identity, or, equivalently, the graph distance

between  $\sigma$  and  $I$  where there is an edge between  $\sigma$  and  $\pi$  if and only if  $p(\sigma, \pi) > 0$ . Notice first that since we are only applying 3-cycles, the permutation stays even forever, so  $d_{3\text{-cycle}}(\sigma) = \infty$  if  $\sigma \notin \mathcal{A}_n$ .

In the case of random transpositions we had the convenient formula that if  $\sigma \in \mathcal{S}_n$  then  $d_{\text{RTRW}}(\sigma) = n - \#\text{cycles}$ , a formula originally due to Cayley. In the case of 3-cycles unfortunately there isn't such a nice formula to work with. However we can prove some useful bounds, which also relate  $d_{3\text{-cycle}}(\sigma)$  to its number of cycles when  $\sigma \in \mathcal{A}_n$ .

**Proposition 3.**

$$\frac{1}{2}(n - |\sigma|) \leq d_{3\text{-cycle}}(\sigma) \leq \frac{1}{2}(n - |\sigma|) + \frac{1}{2}\#\text{cycles of even length}$$

where  $|\sigma|$  is the number of cycles of  $\sigma$ .

*Proof.* For each cycle of odd length  $(i_1, \dots, i_{2r+1})$  we can write

$$(i_1, \dots, i_{2r+1}) = (i_1, i_2, i_3)(i_3, i_4, i_5) \dots (i_{2r-1}, i_{2r}, i_{2r+1})$$

which has exactly  $r$  3-cycles factors. Now, because  $\sigma \in \mathcal{A}_n$ , the number of cycles of even length must be even. So let  $(i_1, \dots, i_{2r})(j_1, \dots, j_{2m})$  be a pair of even cycles. Then we start by building

$$(i_1, i_2)(j_1, j_2) = (i_1, i_2, j_1)(i_2, j_1, j_2)$$

in two moves and then completing each of the cycle in the same way as above. The total number of moves to build this pair of cycles is thus  $2 + (r - 1) + (m - 1) = r + m$

Hence each  $\sigma$  can be made up of at most

$$\sum_{c \text{ odd cycle}} \frac{1}{2}(|c| - 1) + \sum_{c \text{ even cycle}} \frac{1}{2}|c| = \frac{1}{2}(n - |\sigma|) + \frac{1}{2}\#\text{ even cycles}$$

So  $d_{3\text{-cycle}}(\sigma) \leq \frac{1}{2}(n - |\sigma|) + \frac{1}{2}\#\text{ even cycles}$ . On the other hand, each multiplication by a 3-cycle can create at most two new cycles. Hence, after  $k$  multiplications the resulting permutation can't have more than  $|\sigma| + 2k$  cycles. Therefore the distance must be at least that  $k_0$  for which  $|\sigma| + 2k_0 \geq n$ . Indeed, if

$$(\gamma_d \dots \gamma_1) \cdot \sigma = I$$

then

$$\sigma = (\gamma_d \cdot \dots \cdot \gamma_1)^{-1} = \gamma_1^{-1} \cdot \dots \cdot \gamma_d^{-1}$$

If  $\gamma_i$  is a 3-cycle then so is  $\gamma_i^{-1}$ . Hence the number of multiplications needed to turn  $\sigma$  into the identity is also the distance  $d_{3\text{-cycle}}(\sigma)$ .  $\square$

**Remark** If considering the case of random  $p$ -cycles, then the above argument shows that:  $d_{p\text{-cycle}}(\sigma) \geq \frac{1}{p-1}(n - |\sigma|)$ . The reverse inequality is more intricate but there is a number  $M(p) \geq 0$  such that

$$d_{p\text{-cycle}} \leq \frac{1}{p-1}[n - |\sigma|] + M(p)\#\{\text{cycles of length } l \neq 1 \pmod{p-1}\}$$

in which case the argument below for the law of large numbers for the distance still holds.

### 4.3.4 Phase transition for the 3-cycle random walk

For  $c > 0$ , let  $u(c) = 1 - \sum_{s=0}^{\infty} \frac{(2s+1)^{s-2}}{s!} (3c)^s e^{-6c(s+1/2)}$ .

**Theorem 42.** *Let  $\omega(n)$  be any sequence with  $\omega(n) \rightarrow \infty$  as slowly as we want. Then*

$$\frac{d_{3\text{-cycle}}(\sigma_{cn}) - u(c)n}{\omega(n)\sqrt{n}} \rightarrow 0$$

*in probability.*

*Proof.* The proof is adapted from Theorem 4 of Berestycki and Durrett (2004). We couple the random walk  $(\sigma_t, t \geq 0)$  with a random hypergraph process  $G(t)$ , where we add the hyperedge  $\{a, b, c\}$  to  $G(t^-)$  whenever the 3-cycle  $(a b c)$  is applied to the random walk. Of course this process has the properties that the cycles of  $\sigma_t$  are subsets of the connected components of  $G(t)$ . This can be seen by remarking that the cycles of  $\sigma_t$  perform a coagulation and fragmentation process. To see this, write any 3-cycle  $(a b c)$  as  $(a b)(b c)$ . It is known (see Berestycki and Durrett (2004), where this is extensively discussed) that multiplication by a transposition may result in a coagulation or a fragmentation of the cycles, depending on whether the two integers transposed are part of the same cycle or not. Here, we do two such operations at the same time, so the result of a 3-cycle can be a coagulation of 3 cycles at the same time, a double-fragmentation, or a coagulation and a followed by a fragmentation and vice-versa.

Let us note that cycles of even length in  $\sigma_t$  can only result from fragments of clusters in the hypergraph that have nonnegative excess. Indeed, for a hypertree, there must be an odd number of vertices (by Definition 7) and there has been no fragmentation on such a component because on a tree no two hyperedge share more than one vertex. (In the random transposition random walk we also had to take care of events where the same transposition is chosen twice, but here this doesn't happen with high probability since the probability of choosing a given hyperedge is of order  $O(1/n^2)$ , so the probability of choosing twice the same hyperedge on a time-span of  $O(n)$  is  $O(1/n)$ ). We will show that at any time  $cn$  with  $0 < c < \infty$  there are no more than  $O(\sqrt{n})$  extra cycles in the permutations compared to the number of clusters in  $G(cn)$ . Hence the result will follow immediately from the analog result for cluster count in  $G(cn)$  (Theorem 41) and from the bounds on the distance (Proposition 3) since the number of non-tree components is  $o(n^{1/2})$  in  $G(cn)$  by Theorem 40.

Now, there can of course never be more than  $n^{1/2}$  cycles larger than  $n^{1/2}$ . Hence we are interested in counting those extra cycles of size  $\leq n^{1/2}$  generated by fragmentation of a hypergraph cluster. Since a 3-cycle  $(i, j, k)$  is nothing but the product of two transpositions  $(i, j)(j, k)$  we see that to generate a fragment of size smaller than  $A$ ,  $j$  (resp.  $k$ ) must be located at most  $A$  points away from  $i$  (resp.  $i$  or  $j$ ) in the cycle containing  $i$  (resp.  $i$  or  $j$ ). Hence the rate at which such fragments are created is at most  $2A/n + 4A/n = 6A/n$ . When  $A = n^{1/2}$  and when we integrate over a time span of order  $O(n)$  we see that the expected numbers of such fragments is  $O(n^{1/2})$ .  $\square$

**Remark.** A similar result holds for general  $p$ -cycle random walks,

$$\frac{1}{n} d_{p\text{-cycle}}(\sigma_{cn}) \rightarrow_p u(c)$$

where  $u(c)$  has a phase transition from linear to sublinear at  $c = 1/(p(p-1))$ .

## 4.4 Random adjacent transpositions

Let  $S$  be the set of adjacent transpositions,  $S = \{(1, 2), (2, 3), \dots, (n-1, n)\}$ . We consider the random walk  $(X_t, t \geq 0)$  in continuous time on  $\mathcal{S}_n$  generated by  $S$ , i.e. the composition of a Poisson number of uniformly chosen adjacent

transpositions (elements of  $S$ ). It is well known that for  $\sigma \in \mathcal{S}_n$ , the distance from  $\sigma$  to the identity is

$$d_{\text{adj}}(\sigma) = \#\{i < j : \sigma(i) > \sigma(j)\}$$

the number of inversions of  $\sigma$ .

#### 4.4.1 Trivial scaling

Here we just explain briefly how choosing any time-scale may lead to results that do not really reflect the real nature of behavior of the random walk. We look at times  $n^\alpha$  for  $0 \leq \alpha \leq 1$ . This time-scale is chosen so that for  $\alpha < 1$  we are still so close to the identity that nothing has happened, while  $\alpha = 1$  presents a discontinuity: the walk is already in a sublinear regime.

We first prove that for  $\alpha < 1$ ,

$$\frac{1}{n^\alpha} \mathbf{E}d_{\text{adj}}(X_{n^\alpha}) \rightarrow 1$$

Let  $N_t$  be the number of transpositions performed by time  $t$  and  $F_t$  be the number of moves that have decreased the distance, so that

$$d_{\text{adj}}(X_t) = N_t - 2F_t$$

Then  $F_t$  is a counting process whose underlying point process has intensity  $\leq d_{\text{adj}}(X_t)/n$ . (Indeed to decrease the distance we have to transpose a pair  $(i, j)$  such that  $X_t(i) = X_t(j) + 1$  and there are at most  $d_{\text{adj}}(X_t)$  such pairs.) Hence

$$\begin{aligned} \mathbf{E}d_{\text{adj}}(X_{n^\alpha}) &\geq n^\alpha - 2\mathbf{E} \int_0^{n^\alpha} d_{\text{adj}}(X_s)/nds \\ &\geq n^\alpha - 2n^{2\alpha-1} \end{aligned}$$

Of course it is always true that  $\mathbf{E}d_{\text{adj}}(X_{n^\alpha}) \leq n^\alpha$  so this part is proved.

Now on the other hand when  $\alpha = 1$  a consequence of Theorem 43 is that

$$\frac{1}{n} \mathbf{E}d_{\text{adj}}(X_n) \rightarrow c \text{ where } c < 1$$

## 4.4.2 Random walk scaling

To compute  $\mathbf{E}d_{\text{adj}}(X_{nt})$  it is enough to compute the contribution of the couple  $(i, j)$  to this sum and then add everything up. I.e,

$$\mathbf{E}d_{\text{adj}}(X_{nt}) = \sum_{i < j} \mathbf{P}[X_{nt}(i) > X_{nt}(j)]$$

Then the crucial observation is that  $\{X_{nt}(i), t \geq 0\}$ , is a rate 1 Simple Random Walk in continuous time on  $\{1 \dots n\}$ , starting at  $i$ , with 1 and  $n$  as reflecting boundaries. Moreover, the only interaction between  $\{X_{nt}(i), t \geq 0\}$  and  $\{X_{nt}(j), t \geq 0\}$  is that they cannot sit on the same location. In other words, they perform simple exclusion on  $\{1, \dots, n\}$  with reflecting boundaries.

For an integer  $x \geq 0$ , let  $T^x$  denote the hitting time of the level  $x$  by a rate 2 random walk on  $\mathbf{Z}$  starting at 0. We will need another definition. Suppose  $\{Y(t), t \geq 0\}$  and  $\{Y'(t), t \geq 0\}$  perform simple exclusion on  $\mathbf{Z}$ , with  $Y(0) = 0$  and  $Y'(0) = 1$ . Then we define  $p(u)$  to be the probability that at time  $u$  the two particles are exchanged, i.e.  $p(u) = \mathbf{P}[Y(u) > Y'(u)]$ .

**Theorem 43.** *Let  $t > 0$ . Then*

$$\frac{1}{n} \mathbf{E}d_{\text{adj}}(X_{nt}) \rightarrow f(t) := \sum_{x=0}^{\infty} \int_0^t \mathbf{P}[T^x \in ds] p(t-s)$$

as  $n \rightarrow \infty$ .

The fact that this series converges is an (easy) part of the result. To see that  $f(t) < t$  note that  $p(u) < 1$  for all  $u > 0$ , so  $f(t) < \sum_{x=0}^{\infty} \int_0^t P(T^x \in ds) = E(\#x \text{ discovered by the random walk during } (0, t)) \leq t$

**Proof** Theorem 4 is easy to understand. Suppose that we neglect boundary effects in the sum  $\sum_{i < j} \mathbf{P}[X_t(i) > X_t(j)]$ . Then theorem 4 just expresses the Markov property at the first time that the two random walks sit next to each other, after which we are indeed left with processes like  $\{Y(t), t \geq 0\}$  and  $\{Y'(t), t \geq 0\}$  as long as we ignore any boundary effect.

In all that follows, let  $H_{i,j}(t)$  denote the event that the two random walks starting at  $i$  and  $j$  sit next to each other at some time  $s \leq t$  :

$$H_{i,j}(t) := \{\exists s \leq t : X_t(j) = X_t(i) + 1\}$$

Let  $I_x = \{i \geq \log n, i + x = j \leq n - \log n\}$ .  $I_x$  is the set of points  $i$  such that for  $i$  and  $j = i + x$  the boundary contributions will be small.

$$\sum_{i < j} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] = \sum_{x=1}^n \sum_{i \in I_x} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] + \sum_{x=1}^n \sum_{i \notin I_x} \mathbf{P}[X_{nt}(i) > X_{nt}(j)]$$

**Lemma 26.**

$$\frac{1}{n} \sum_{x=1}^n \sum_{i \notin I_x} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] \rightarrow 0$$

**Proof** There are first those  $i, j$  who are both "outside" (i.e., neither  $i$  nor  $j$  are in  $\{\log n, \dots, n - \log n\}$ ). There cannot be more than  $O(\log n)^2$  of those. More interesting is the case where  $i < \log n$  and  $\log n \leq j < n - \log n$  :

$$\begin{aligned} & \sum_{x=1}^n \sum_{i \notin I_x} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] \\ & \leq O(\log n)^2 + 2 \sum_{x=1}^t \sum_{\substack{i < \log n, \\ j = i + x \geq \log n}} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] \\ & \quad + 2 \sum_{x=t+1}^n \sum_{\substack{i < \log n, \\ j = i + x \geq \log n}} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] \\ & \leq O(\log n)^2 + 2t^2 + 2 \sum_{x=t+1}^n \sum_{\substack{i < \log n, \\ j = i + x \geq \log n}} \mathbf{P}[H_{i,j}(nt)] \\ & \leq O(\log n)^2 + 2t^2 + 2 \sum_{x=t+1}^n \sum_{\substack{i < \log n, \\ j = i + x \geq \log n}} \mathbf{P}[\text{Poisson}(t) > x] \\ & \leq O(\log n)^2 + 2t^2 + 2 \log nt = o(n) \end{aligned}$$

since for  $X_{nt}(i)$  and  $X_{nt}(j)$  to hit there must be at least  $x$  moves performed in duration  $t$ .  $\square$

We next turn our attention to the main contribution to the double sum, which is given by "inside" terms  $i \in I_x$ . For those we need first to show the boundary reflections can be safely neglected. Let  $Y_i$  and  $Y_j$  perform simple exclusion on  $\mathbf{Z}$  starting at  $i$  and  $j$  respectively ; let  $A_{i,j}$  denote the event that neither random walk hits one of the boundaries, and let  $A'_{i,j}$  be the event

that simple exclusion started at  $i$  and  $j$  doesn't hit one of the boundaries. For  $x \geq 1$  and  $i \in I_x$ ,

$$\begin{aligned} \mathbf{P}[X_{nt}(i) > X_{nt}(j)] &= \mathbf{P}[X_{nt}(i) > X_{nt}(j); A_{i,j}] + \mathbf{P}[X_{nt}(i) > X_{nt}(j); A_{i,j}^c] \\ &= \mathbf{P}[Y_i(t) > Y_j(t); A'_{i,j}] + \mathbf{P}[X_{nt}(i) > X_{nt}(j); A_{i,j}^c] \\ &= \mathbf{P}[Y_i(t) > Y_j(t)] + O(\mathbf{P}[X_{nt}(i) > X_{nt}(j); A_{i,j}^c]) \end{aligned}$$

**Lemma 27.**

$$\frac{1}{n} \sum_{x=1}^n \sum_{i \in I_x} \mathbf{P}[X_n(i) \text{ or } X_n(j) \text{ have hit the boundary before time } t] \rightarrow 0$$

**Proof** Let  $\mathbf{P}_i$  denote the law of a rate 1 simple random walk starting at  $i$ .

$$\begin{aligned} \frac{1}{n} \sum_{x=1}^n \sum_{i \in I_x} \mathbf{P}[A_{i,j}^c] &= \frac{1}{n} O\left(\sum_{x=1}^n \sum_{I_x} \mathbf{P}[Y_i \text{ hits } 1 \text{ or } n]\right) \\ &= O\left(\sum_{i \in I_x} \mathbf{P}[Y_i \text{ hits } 1 \text{ or } n]\right) \\ &= O\left(\sum_{i \in I_x} \mathbf{P}_i[T^0 < t]\right) = O\left(\sum_{i \in I_x} \mathbf{P}_0[T^i < t]\right) \\ &= O\left(\sum_{i > \log n} \mathbf{P}_0[M_t > i]\right) = o(1) \end{aligned}$$

where  $M_t := \sup_{s \leq t} W_s$  is the supremum of a rate 1 random walk on  $\mathbf{Z}$ , and where the last equality follows from the fact that  $\mathbf{E}M_t < \infty$ . (To see this, just note that  $M_t$  is at most a Poisson random variable with mean  $t$  in the case where the random walk is increasing at each step).  $\square$

Combining lemma 1 and 2 gives the following :

$$\begin{aligned} \frac{1}{n} \mathbf{E}d_{\text{adj}}(X_{nt}) &= o(1) + \frac{1}{n} \sum_{x=1}^n \sum_{i \in I_x} \mathbf{P}[Y_i(t) > Y_j(t)] \\ &= o(1) + \frac{1}{n} \sum_{x=1}^n \sum_{i \in I_x} \int_0^t \mathbf{P}[T^{x-1} \in ds] p(t-s) \\ &= \sum_{x=0}^{n-1} \frac{|I_x|}{n} \int_0^t \mathbf{P}[T^x \in ds] p(t-s) + o(1) \end{aligned}$$

where we have used the strong Markov property at the first time  $T^{i,j}$  that  $Y_i$  and  $Y_j$  are adjacent and where we have used the fact that  $T^{i,j} \stackrel{D}{=} T^{x-1}$ . Indeed the distance between  $Y_i$  and  $Y_j$  is a rate 2 random walk on  $\{1, \dots, \}$  reflecting at 1.

Now, given that  $|I_x| = n - 2 \log n - x$ , it is a simple game to show that the last sum converges to the series in the statement of the theorem,  $f(t)$  : let  $u_x(t) = \int_0^t \mathbf{P}[T^x \in ds] p(t-s)$ . Then  $\sum_{x \geq 0} u_x(t) < \infty$ , by the crude bound  $p(t-s) \leq 1$  and using the fact already noted that  $\mathbf{E}M_t < \infty$ .

Let  $\varepsilon > 0$ . Let  $N \geq 0$  be such that  $\sum_{x=N+1}^{\infty} u_x(t) < \varepsilon$ . Then

$$\sum_{x=0}^{n-1} \frac{|I_x|}{n} \int_0^t \mathbf{P}[T^x \in ds] p(t-s) \geq \frac{n - 2 \log n - N}{n} \sum_{x=0}^N u_x(t)$$

Hence

$$\liminf \sum_{x=0}^{n-1} \frac{|I_x|}{n} \int_0^t \mathbf{P}[T^x \in ds] p(t-s) \geq \sum_{x=0}^N u_x(t) \geq f(t) - \varepsilon$$

We can let  $\varepsilon \rightarrow 0$ . On the other hand, the limsup is dealt with even more easily by  $|I_x| \leq n$ . Hence the theorem.  $\square$

**Theorem 44.**

$$\lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n\sqrt{t}} \mathbf{E}d_{\text{adj}}(X_{nt}) = \frac{1}{2} \mathbf{E}[\max_{0 \leq s \leq 1} B_{2s}] = \sqrt{2}/2$$

where  $B$  is a standard Brownian Motion.

We start the proof by a lemma :

**Lemma 28.**  $p(t) \rightarrow 1/2$  as  $t \rightarrow \infty$ .

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

**Lemma 29.**

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

*Proof.* We write this as

$$\begin{aligned} t^{-1/2} E\left(\sum_{x=1}^{\infty} \mathbf{1}_{\{T^x \in (t - \log t, t)\}}\right) &\leq t^{-1/2} E(\#\text{jumps in } (t - \log t, t)) \\ &\leq t^{-1/2} \log t \rightarrow 0 \end{aligned}$$

Indeed we can only uncover a site  $x$  once at a time, so the number of sites  $x$  such that  $T^x \in (u, v)$  is smaller than the number of jumps of the random walk in the interval  $(u, v)$ . On the other hand, this number is a Poisson random variable with mean  $(v - u)$ , which is exactly what we apply here.  $\square$

**Proof** of Theorem 5. Let  $\varepsilon > 0$ . Fix  $T$  large enough so that  $|p(t) - 1/2| \leq \varepsilon$  as soon as  $t \geq T$ . Then by the above lemma, for  $t \geq T' := e^T$ ,

$$\begin{aligned} t^{-1/2} f(t) &= t^{-1/2} \sum_{x=0}^{\infty} \int_0^{t - \log t} \mathbf{P}[T^x \in ds] p(t - s) + o(1) \\ &\leq \left(\frac{1}{2} + \varepsilon\right) t^{-1/2} \sum_{x=0}^{\infty} \mathbf{P}[T^x < t - \log t] + o(1) \\ &\leq \left(\frac{1}{2} + \varepsilon\right) t^{-1/2} \sum_{x=0}^{\infty} \mathbf{P}\left[\max_{s \leq t - \log t} W_{2s} > x\right] \\ &\rightarrow \left(\frac{1}{2} + \varepsilon\right) \mathbf{E} \max_{s \leq 1} B_{2s}. \end{aligned}$$

by Donsker's theorem. On the other hand

$$\liminf_{t \rightarrow \infty} t^{-1/2} f(t) \geq \left(\frac{1}{2} - \varepsilon\right) \mathbf{E} \max_{s \leq 1} B_{2s}$$

by similar methods, so Theorem 5 is proved.  $\square$

### 4.4.3 Brownian scaling

Let  $p_t(x, y)dy$  be the transition kernel of Brownian Motion reflected at 0 and 1. Then the following result holds.

**Theorem 45.** *Let  $t > 0$ .*

$$\frac{1}{n^2} d_{\text{adj}}(X_{n^3t}) \rightarrow \int_0^1 du \int_u^1 dv \int_0^1 p_t(u, x) dx \int_0^y p_t(v, y) dy = \mathbf{P}[B_1(t) > B_2(t)]$$

*in probability and where  $B_1$  and  $B_2$  are two reflecting Brownian motions started uniformly on  $0 \leq B_1(0) < B_2(0) \leq 1$  evolving independently.*

**Proof** We first show that the convergence holds in expectation, i.e.

$$\frac{1}{n^2} \mathbf{E} d_{\text{adj}}(X_{n^3t}) = \frac{1}{n^2} \sum_{i < j} \mathbf{P}[X_{n^3t}(i) > X_{n^3t}(j)] \rightarrow \mathbf{P}[B_1(t) > B_2(t)]$$

Of course, with this time-scale, the rescaled random walks  $\{\frac{1}{n}X_{n^3t}(i), t \geq 0\}$  converge to reflecting Brownian Motion on  $[0, 1]$ . In fact, Durrett and Neuhauser (1994) showed that the following stronger statement holds: for fixed  $i < j$ , the rescaled couple of random walks converge to two independent Brownian Motions. In other words, the only interaction, which only comes from when the two particles are adjacent, can be neglected at this time-scale. (This is well understood with the fact that Brownian particles will hit many times and so the actual result of their interaction is "as random" as if there was no interaction.) This shows

$$\frac{1}{n^2} \mathbf{E} d_{\text{adj}}(X_{n^3t}) = \frac{1}{n^2} \sum_{i < j} \mathbf{P}[X_{n^3t}(i) > X_{n^3t}(j)] \rightarrow \mathbf{P}[B_1(t) > B_2(t)]$$

To get the convergence in probability, let us compute second moments estimates:

$$\mathbf{E} \left( \frac{1}{n^2} d_{\text{adj}}(X_{n^3t}) \right)^2 = \frac{1}{n^4} \sum_{i < j} \sum_{k < l} \mathbf{P}[X_{n^3t}(i) > X_{n^3t}(j); X_{n^3t}(k) > X_{n^3t}(l)]$$

We can again apply Durrett and Neuhauser's (1994) results to the 4-uplet of random walks  $(X(i), X(j), X(k), X(l))$ , except when  $i = k$  or  $j = l$ . In this case the contributions are asymptotically 0. Indeed by the above,

$$\frac{1}{n^4} \sum_{i < j} \mathbf{P}[X_{n^3t}(i) > X_{n^3t}(j)] \rightarrow 0$$

and

$$\frac{1}{n^4} \sum_{i < j} \sum_{l \neq j} \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(j); X_{n^{3t}}(i) > X_{n^{3t}}(l)] \rightarrow 0$$

because the same quantity, when divided by  $n^3$  instead of  $n^4$ , converges to  $\mathbf{P}[B_1(t) > B_2(t); B_1(t) > B_3(t)]$  by the same type of arguments, where  $(B_1, B_2, B_3)$  are three independent reflecting Brownian motions.

On the other hand

$$\begin{aligned} \frac{1}{n^4} \sum_{i < j} \sum_{l \neq j} \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(j)] \cdot \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(l)] \\ \leq \frac{1}{n^3} \sum_{i < j} \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(j)] \rightarrow 0 \end{aligned}$$

and

$$\frac{1}{n^4} \sum_{i < j} \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(j)]^2 \leq \frac{1}{n^4} \sum_{i < j} \mathbf{P}[X_{n^{3t}}(i) > X_{n^{3t}}(j)] \rightarrow 0$$

as well.

Hence

$$\mathbf{E} \left( \frac{1}{n^2} d_{\text{adj}}(X_{n^{3t}}) \right)^2 \rightarrow \mathbf{P}[B_1(t) > B_2(t)]^2$$

In other words, the variance of  $\frac{1}{n^2} d_{\text{adj}}(X_{n^{3t}})$  is asymptotically 0 : by applying Chebyshev's inequality, we get the convergence in probability.  $\square$

## 4.5 Random walk on a random 3-regular graph

Let  $(X_k, k \geq 0)$  be a random walk on a random 3-regular graph,  $G_n$ . This is a graph on  $n$  vertices such that each vertex has degree 3, and what we mean here by "random 3-regular graph" is of course a uniform member of such graphs.

Although it is not obvious how to construct such graphs, Bollobàs and de la Vega (1982) have given the following very useful representation. Consider the random graph  $G_n^*$  obtained from a random matching on  $3n$  vertices. Then, collapse vertices  $3i, 3i+1$  and  $3i+2$  to form a single vertex. On the event (of

asymptotically positive probability) that the resulting graph has no self-loops or multi-edges, then this is a realization of a random 3-regular graph.

This representation will be very useful for computing the probability of various events describing the configuration of the vertex on which the random walk steps. It was originally designed to compute the isoperimetric constant of a random regular graph. We can also use it to prove that with high probability this graph is connected (meaning that there is a single connected component, as opposed to a giant component and possibly many other small components).

We will assume that  $X_k$  starts at a distinguished vertex 1. We will now state the theorem.

**Theorem 46.** *For fixed  $t > 0$*

$$\frac{d(X_{\lfloor t \log_2 n \rfloor})}{\log_2 n} \xrightarrow{p} f(t) := \sup \left( \frac{1}{3}t, 1 \right)$$

An intuitive description of a random 3-regular graph as seen from vertex 1 can be given as follows. Let us grow the cluster containing 1 by exposing successive vertices with increasing distance away from 1. As long as we have not exposed more than  $o(n)$  vertices this should look very much like a regular tree (i.e., each vertex has 2 edges going away from the root and 1 leading back towards the root). Therefore this phase roughly lasts until distance  $\log_2 n$ . After distance  $\log_2 n$  however, this tree approximation is no longer valid and there are many edges inter connecting the vertices.

Therefore we expect that  $d(X_k)$  evolves like  $(1/3)k$  as long as we are below critical distance  $\log_2 n$ . After that time, it should stay at this distance, since this is the diameter of the graph (by Bollobàs and de la Vega (1982)) and thus cannot go any further, and on the other hand the tree structure makes it hard for it to come down back to the root.

### 4.5.1 Proof of subcritical regime

Let  $v$  be a vertex at distance  $l$  from the root. We say that  $v$  is a "good" vertex if it has two edges leading away from the root (at distance  $l + 1$ ) and one leading back to distance  $l - 1$ , and if moreover its two children have this same property (two edges leading out). Otherwise we say that  $v$  is a "bad" vertex. Let  $B(l)$  be the set of all bad vertices at distance  $l$ .

**Lemma 30.** *Let  $2 \leq v \leq n$  be a vertex distinct from the root. Given that  $v$  is at distance  $l$  from the root,*

$$P(v \in B(l)) \leq 2i/n$$

where  $i = 2^l$ .

*Proof.* This is a consequence of the Bollobàs representation of a random regular graph. First consider the event that  $v$  has an edge leading at some other vertex at distance  $l$ . Since it is at distance  $l$ , it must have at least one edge leading backwards, so there are only two other edges left. (In particular there are at most  $2^l = i$  vertices at distance  $l$ ). In  $G_n^*$  those  $i$  vertices at distance  $l$  correspond to  $3i$  vertices, so this probability is smaller than  $2i/3n$ .

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

A simple heuristic now allows us to understand that the random walk will not encounter more than  $O(1)$  bad vertices (as long as we are in the subcritical regime). Because of the push forward we will not spend more than  $O(1)$  time on each level, hence the expected number of bad vertices encountered until time distance  $(1 - \varepsilon) \log_2 n$  is smaller than

$$\sum_{l=1}^{(1-\varepsilon) \log_2 n} O(1) \frac{2^l}{n} = O(n^{-\varepsilon}) \rightarrow 0$$

More rigorously, let  $A_k$  denote the event that by time  $k$  the random walk has never stepped on a bad vertex, and let  $A_k^c$  denote its complement.

**Lemma 31.** *Suppose  $k \leq 3(1 - \varepsilon) \log_2 n$  so that  $k$  belongs to the subcritical regime. For any  $\rho > (1/3)$  there is a  $C_\rho$  such that*

$$P(A_{k+1}^c | A_k) \leq C_\rho \frac{2^{\rho k}}{n}$$

*Proof.* Let us first remark that for  $\rho > 1/3$  with overwhelming probability the distance of the random walk is not more than  $\rho k$  since for each vertex

there are at most two edges leading out and one leading back. More precisely, for all  $k > 0$

$$P(d(X_k) > \rho k) \leq C e^{-\alpha k} \quad (4.4)$$

by standard large deviations arguments ( $C$  and  $\alpha$  depend on  $\rho$ ). Remark that for the same reason the same estimate holds under the conditional probability given  $A_k$ .

Since (4.4) is smaller than  $C_\rho \frac{2^{\rho k}}{n}$ , we may assume that  $d(X_k) \leq \rho k$ . Given  $A_k$ , the only way that we can step on a bad vertex at time  $k + 1$  is if the random walk is visiting a vertex never visited before which happens to be bad. Since we can imagine that the environment is generated only as the random walk steps on a vertex never visited before, this is a situation where we can apply Lemma 30: this event that we step on a bad vertex never visited previously has probability smaller than

$$\sum_{l \leq \rho(k+1)} P(d(X_{k+1}) = l; X_{k+1} \notin \{X_0, \dots, X_k\} | A_k) 2 \frac{2^l}{n}$$

which, in turn, is smaller than  $2 \sum_{l \leq \rho(k+1)} 2^l / n = O(1) 2^{\rho k} / n$  for some universal constant  $O(1)$  (but the constant  $C_\rho$  in the statement of Lemma 31 is not universal because (4.4) depends on  $\rho$ ).  $\square$

*End of proof for the subcritical regime:* We apply Lemma 31 recursively. Let  $k = \lfloor t \log_2 n \rfloor$  with  $t = 3(1 - \varepsilon)$  for some  $\varepsilon > 0$ .

$$\begin{aligned} P(A_{k+1}) &\geq P(A_k) P(A_{k+1} | A_k) \\ &\geq P(A_k) \left(1 - \frac{2^k C_\rho}{n}\right) \\ &\geq \prod_{j=1}^k \left(1 - \frac{2^j C_\rho}{n}\right) \end{aligned}$$

Taking the logarithm:

$$\begin{aligned} \log P(A_{k+1}) &\geq \sum_{j=1}^k \log \left(1 - \frac{2^j C_\rho}{n}\right) \\ &\sim - \sum_{j=1}^k \frac{2^j C_\rho}{n} \sim -n^{\alpha(\varepsilon, \rho)} C_\rho \end{aligned}$$

for  $\alpha(\varepsilon, \rho) = -1 + 3\rho(1 - \varepsilon)$ . By choosing  $\rho > 1/3$  close enough to  $1/3$  so that

$$\rho(1 - \varepsilon) < 1/3$$

we get a negative power  $\alpha(\varepsilon, \rho) < 0$  in the above calculation, which proves that  $P(A_k) \rightarrow 1$  for  $k = t \log_2 n$  and  $t < 3$ . In other words, with overwhelming probability  $(X_j; 1 \leq j \leq t \log_2 n)$  never steps on a bad vertex. On this event, for each  $1 \leq j \leq t \log_2 n$   $X_j$  has probability  $2/3$  to move away from the root and  $1/3$  to move back towards the root, which results in a linear speed of  $1/3$  for the random walk: we conclude easily to the first part of Theorem 46.

## 4.5.2 Proof of the supercritical regime

Here we wish to prove that if  $k = t \log_2 n$ , with  $t > 3$ , then  $d(X_k) \approx \log_2 n$ . As already noted, this is the diameter of  $G_n$  so all we have to prove is that once it reaches this distance the random walk cannot go back downstairs - it is trapped upstairs where there are many interconnections.

It turns out that the estimates proved in the subcritical regime are again very useful. In particular, Lemma 30 will again play a key role.

Let  $\varepsilon > 0$ . Theorem 46 is proved if we show that  $P(d(X_k) \leq (1 - \varepsilon) \log_2 n) \rightarrow 0$ .

The idea is the following. Suppose  $t = 3$  first. If  $d(X_k) \leq (1 - \varepsilon) \log_2 n$  we can find  $t' < 3$  such that  $d(X_{t' \log_2 n}) \approx (1 - \delta) \log_2 n$ , where  $\delta < \varepsilon$  (indeed, it suffices to take  $t' = 3(1 - \delta)$  by the previous section). It suffices to show that the expected number of bad vertices encountered in the strip  $L(1 - \varepsilon, 1 - \delta)$  where  $L(a, b) := \{2 \leq v \leq n : d(v) \in [a \log_2 n, b \log_2 n]\}$  is small. When we do so it becomes almost impossible for the random walk to resist the  $1/3$  drift forward.

The proof for  $t > 3$  exploits the exact same idea, and that is the case we will concentrate on so as to avoid repeating the same arguments.

In what follows,  $\varepsilon > 0$ , is a fixed number  $k = t \log_2 n$  and  $t > 3$ ,  $\delta$  is some fixed number with  $\delta < \varepsilon$ , and we let  $k' = t' \log_2 n$  for  $t' = 3(1 - \delta)$ , so that  $d(X_{k'}) = (1 - \delta + o(1)) \log_2 n$ .

Here we have to be a little careful, there are many bad vertices in the strip  $L(1 - \varepsilon, 1 - \delta)$ , (of order  $n^{1-\delta}$ ), but in fact we will see that most of them

cannot affect the behavior of the random walk with high probability on the environment. To do so, we divide the vertices in the strip  $L(1 - \varepsilon, 1 - \delta)$  into roughly  $m \approx n^{1-\varepsilon}$  categories. Let  $v_1, \dots, v_m$  be the  $m$  vertices at level  $(1 - \varepsilon) \log_2 n$ . For  $j = 1, \dots, m$  if  $v \in L(1 - \varepsilon, 1 - \delta)$ , we say that  $v \in T_j$  if  $v_j$  is the closest vertex to  $v$  among  $v_1, \dots, v_m$ . (We think of  $T_j$  as subtrees below  $v_j$ , this definition allowing to eliminate any ambiguity from this vague idea).

The crucial fact we shall prove is that, even though some vertex in  $T_j$  might also connect to  $T_{j'}$ , the 'subtrees'  $T_j$  are very loosely interconnected. More precisely, Lemma 33 below shows that with high probability no  $T_j$  is connected to more than  $O(1)$  other subtrees if we only allow to use vertices in  $L(1 - \varepsilon, 1 - \delta)$  for connecting different subtrees.

More precisely, we will prove that the number of bad vertices in trees connected to  $T_j$  is dominated by the total progeny of a branching process where the offspring distribution has mean  $O(n^{\varepsilon-2\delta})$ . Since sideways connections may only exist on bad vertices, this will prove that  $T_j$  is, with high probability, connected to at most  $O(1)$  other trees. We will also be able to deduce from this that while crossing from  $(1 - \delta) \log_2 n$  down to  $(1 - \varepsilon) \log_2 n$  the random walk may only step on at most  $o(\log n)$  bad vertices, from which it is easy to conclude that the probability of reaching level  $(1 - \varepsilon) \log n$  is tiny.

To prove the domination by a branching process, we remark that we can construct the regular graph using Bollobàs' representation, so that it is enough to examine the connections between vertices one level after another. Suppose that for some level  $l$  we determine the connections of each vertex one after another. In a random uniform matching, we may also generate the graph one vertex after another by saying that at a given step, the particular vertex under examination is linked to a vertex uniformly chosen among those that have no neighbor yet.

To fix the ideas, suppose we generate the graph in the following order: inside each level  $l$ , we number the vertices according to their "lexicographical order", i.e., all the vertices at distance  $l$  from 1 in  $T_1$ , then those in  $T_2, \dots$  and then those in  $T_m$ .

**Lemma 32.** *The number of subtrees that  $T_j$  is connected to is dominated by a branching process with offspring distribution  $\text{Binomial}(n^{\varepsilon-\delta}, n^{-\delta})$ .*

*Proof.* Let  $n_l \in \{1, 2, \dots\}$ . We start by remarking that the probability of each given vertex being bad is  $O(2^l/n) = O(n^{-\delta})$  when it is at level  $l$ , and

that there are at most  $n^{\varepsilon-\delta}$  vertices in  $T_j$ . Therefore the expected number of bad vertices in  $T_j$  is  $O(n^{\varepsilon-2\delta}) \rightarrow 0$  if  $\delta$  is chosen  $> \varepsilon/2$ , which we will assume from now on. In particular, there is a good probability that  $T_j$  is not connected to any other tree at all. But we wish to establish more precise bounds on the event that it does connect to other trees. Suppose  $T_j$  is connected to exactly  $m_1$  other subtrees (so, we work conditionally given this event). Let us denote by  $T(j_1), \dots, T(j_{m_1})$  the corresponding indices of the subtrees, and we think of them as the first generation in the branching process, i.e. they are the children of  $T_j$ . Pick one of these children, say  $T(j_1)$ . By Bollobàs' random matching construction of 3-regular graphs, we see that the number of connections that  $T(j_1)$  has with trees different from  $T_j$  and  $T(j_2), \dots, T(j_{m_1})$ , is dominated by a Binomial random variable with parameters  $n^{\varepsilon-\delta}$  and  $n^{-\delta}$ . Indeed if we generate the graph level by level and we count how many connections does  $T(j_1)$  has with other subtrees, except  $T_j$  and  $T(j_2), \dots, T(j_{m_1})$ , we see that the probability for each vertex at level  $(1-\varepsilon)\log_2 n + k$  to connect to one those trees, is smaller than

$$(m - m_1 - 1) \frac{2^{k+1}}{3n - 3n^{1-\varepsilon}2^k} \leq n^{1-\varepsilon} \frac{2^k}{n}$$

when the vertex has one edge leading backward, and is half as much when it has two edges leading backward. The justification for this inequality is the following: there are  $(m-m_1-1)$  other trees, each has at most  $2^k$  individuals at this level, and  $3n^{1-\varepsilon}2^k$  is the number of vertices that have been used so far in the construction of the graph, so  $3n - 3n^{1-\varepsilon}2^k$  is the total number of available vertices at this stage (in Bollobàs' construction). Then  $m - m_1 - 1 \leq m \leq n^{1-\varepsilon}$  accounts for the second inequality. Since  $k \leq (\varepsilon - \delta)\log_2 n$  it follows that each vertex has probability less than  $n^{-\delta}$  of establishing a connection with one of the remaining  $m-m_1-1$  subtrees. Moreover, with this construction we see that it is independent for different vertices and different trees. Therefore each of those subtrees  $T(j_1), \dots, T(j_{m_1})$  is again connected to at most a Binomial( $n^{\varepsilon-\delta}, n^{-\delta}$ ) number of trees. Of course, some of those trees may be identical, but counting them as different only improves our branching process bound. Since we have eliminate all dependence in this bound to the number  $m_1$  of children of  $T_j$ , we can uncondition. Proceeding like this with the children of  $T(j_1), \dots, \dots, T(j_{m_1})$  yields the desired domination by a branching process.  $\square$

**Lemma 33.** *If  $\delta > \varepsilon/2$  then there exists some  $K = K(\varepsilon, \delta) > 0$  such that,*

$$P(\exists j \text{ such that } T_j \text{ is connected to more than } K \text{ subtrees}) \rightarrow 0$$

*Proof.* Let

$$X \stackrel{d}{=} \text{Bin}(n^{\varepsilon-\delta}, 2n^{-\delta/2})$$

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

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

Let  $c = n^{\varepsilon-2\delta}$ , and let  $N = n^{\varepsilon-\delta}$  be the total number of vertices in  $T_j$ , so

$X \stackrel{d}{=} \text{Binomial}(N, c/N)$ . Lemma 33 follows from a simple evaluation of the tail of the total progeny  $Z$  of a branching process with offspring distributed as  $X$ . To do this, we let

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

be the moment generating function of  $X - 1$ . Let  $S_k$  be a random walk that takes steps with this distribution and  $S_0 = 1$ , so that  $S_k$  explores the Galton-Watson tree. Then  $\tau = \inf\{k : S_k = 0\}$  has the same distribution as  $Z$ . Let  $R_k = \exp(\theta S_k) / \phi_N(\theta)^k$ .  $R_k$  is a nonnegative martingale. Stopping at time  $\tau$  we have  $e^{\theta} \geq E(\phi_N(\theta)^{-\tau})$ . If  $\phi_N(\theta) < 1$  it follows that

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

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

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

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

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

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

$$P(Z \geq y) \leq n^\alpha \exp(y(1 - \alpha \log(n)))$$

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

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

so if

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

then the probability than one cluster contains more than  $y$  trees tends to 0.  $\square$

**Lemma 34.** *With probability asymptotically 1, no cluster of trees has more than  $\log \log n$  bad vertices.*

*Proof.* We have to take care of the fact that there may be more than one connection between two trees, so the number we are trying to bound is not just the sum of the bad vertices in each tree plus the number of trees in each cluster (each accounting at least for one bad vertex). Instead, we have to use again arguments coming from the construction of the regular graph. The probability of each vertex being bad is smaller than  $n^{-\delta}$ , and there are at most  $Kn^{\varepsilon-\delta}$  vertices in each cluster by Lemma 33, so the total number of bad vertices in each cluster is smaller than a random variable  $Y$  with Binomial  $(Kn^{\varepsilon-\delta}, n^{-\delta})$  distribution. Taking again  $N = Kn^{\varepsilon-\delta}$  and  $c = Kn^{\varepsilon-2\delta}$  so that  $Y$  is a Binomial  $(N, c/N)$ , we have that since  $c < 1$  for  $n$  large enough,

$$Y \leq \text{Poisson}(c)$$

in the sense of stochastic domination. Hence,

$$\begin{aligned} P(Y > \log \log n) &\leq e^{-c} \sum_{k=\log \log n}^{\infty} \frac{c^k}{k!} \\ &\sim \frac{c^{\log \log n}}{(\log \log n)!} \end{aligned}$$

It is then trivial to see that  $n^{1-\varepsilon} P(Y > \log \log n) \rightarrow 0$ , which concludes the proof of this lemma.  $\square$

**End of the proof of the supercritical regime.** We denote by  $A_n$  the event that there are no more than  $\log \log n$  bad vertices per cluster of trees,  $P(A_n) \rightarrow 1$  by the previous Lemma. We see that there is a subcritical time ( $t' \log_2 n$ ) at which the random walk is strictly above level  $(1 - \delta) \log_2 n$  with almost full probability. Hence the event that the random walk is at distance  $(1 - \varepsilon) \log_2 n$  by time  $t \log_2 n$  with  $t > 3$ , is a subset of the event that between times  $t' \log_2 n$  and  $t \log_2 n$  the random walk crosses down from level  $(1 - \delta) \log_2 n$  to level  $(1 - \varepsilon) \log_2 n$ . But each time it attempts to cross, it has to stay in a single cluster of trees, otherwise it has to go above the end of the strip to change cluster. But on the event  $A_n$ , for each given cluster, only  $\log \log n$  levels may contain bad vertex, which means that apart from those rare levels, the distance of  $X_k$  behaves as an asymmetric  $(2/3, 1/3)$  random walk on  $\mathbf{Z}$ . What is the probability of going, in time  $O(\log n)$ ,  $(\varepsilon - \delta) \log_2 n - \log \log n$  far apart in the opposite directions of the drift, for such a random walk  $S_n$ ? Writing  $S_n = \sum_i X_i$  for i.i.d. random variables with  $P(X_i = 1) = 2/3$  and  $P(X_i = -1) = 1/3$ , standard large deviations arguments give  $P(S_N \leq -aN) \leq e^{-cN}$  where  $c > 0$  is some constant. Since we are interested in  $N = O(\log n)$ , the probability that we are interested in is smaller than  $n^{-c}$  where  $c$  is another constant. There are at most  $O(\log n)$  attempts to cross the strip  $L$ , so the probability to cross it is smaller than  $n^{-c} \log n \rightarrow 0$ . Thus

$$P\left(\frac{1}{\log_2 n} d(X_{t \log_2 n}) \leq (1 - \varepsilon)\right) \leq P\left(\frac{1}{\log_2 n} d(X_{t' \log_2 n}) < (1 - \delta)\right) + P(A_n^c) + n^{-c} \log_2 n$$

All those terms converge to 0, so the Theorem is proved.  $\square$

# Bibliography

Aldous, D. (1997) Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Prob.* 25, 812–854

Aldous, D. (1999) Deterministic and stochastic models for coalescence (aggregation and coagulation) : a review of the mean-field theory for probabilists. *Bernoulli*. 5, 3–48

Angel, O. (2003) Random infinite permutations and the cyclic time random walk. Pages 9–16 in Banderier and Krattenthaler (2003)

Arratia, R. and Barbour, A. and Tavaré, S. (2003) *Logarithmic combinatorial structures : a probabilistic approach*. European Math. Society Monographs, 1.

Athreya, K.B. and Ney, A. (1972) *Branching processes*, Springer-Verlaag, New York.

Bafna, V. and Pevzner, P. (1995) Sorting by reversals: Genome rearrangement in plant organelles and evolutionary history of X chromosome. *Mol. Biol. Evol.* 12, 239–246

Banderier, C., and Krattenthaler, C. (2003) Proceedings of the conference Discrete Random Walks. *Discrete Math and Computer Science*. [dmtcs.loria.fr/proceedings/dmACind.html](http://dmtcs.loria.fr/proceedings/dmACind.html)

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

Berestycki, N. (2005) The hyperbolic geometry of random transpositions. *Ann. Probab.*, to appear.

Berestycki, N. and Durrett, R., (2003) A phase transition in the random transposition random walk. Pages 17-26 in Banderier and Krattenthaler (2003)

- Berestycki, N. and Durrett, R. (2004) A phase transition in the random transposition random walk. *Probab. Theor. Rel. Fields*, to appear.
- Berestycki, N. and Durrett, R. (2005) Limiting behavior for the distance of a random walk. In preparation.
- Bollobás, B. (1984) The evolution of random graphs. *Trans. Amer. Math. Soc.* 286, 257–274
- Bollobás, B. (1985) *Random Graphs*, Cambridge University Press.
- Bollobás, B. (1988) The isoperimetric number of a random graph, *European Journal of Combinatorics*, 9, 241-244.
- Bollobás, B. and de la Vega, F. (1982) The diameter of random regular graphs. *Combinatorica*, 2, 125-134
- Borel, E. (1942) Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au problème de l'attente à un guichet. *C.R. Acad. Sci. Paris.* 214, 452–456
- Bourque, G. and Pevzner, P. A. (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Research.* 12, 26–36
- Cayley, A. (1889). A theorem on trees. *Quarterly Journal of Pure and Applied Mathematics*, 23, 376-378. Also available in *The Collected Mathematical Papers of Arthur Cayley*, Vol XIII, 26-28, Cambridge University Press, 1897.
- Chung, F.K. and Lu, L. (2001) The diameter of sparse random graphs. *Adv. Appl. Math.* 26, 257-279.
- Dembo, A. and Zeitouni, O. (1993) *Large Deviations Techniques and Applications* Jones and Bartlett, Boston, MA.
- Devroye, L. (1992) The branching process method in the Lagrange random variate generation, [cgm.cs.mcgill.ca/~luc/branchingpaper.ps](http://cgm.cs.mcgill.ca/~luc/branchingpaper.ps)
- Diaconis, P. (1988) *Group representation in Probability and Statistics*, Institute of Mathematical Statistics Lecture Notes, Vol. 11.
- Diaconis, P., Mayer-Wolf, E., Zeitouni, O., and Zerner, M. (2004) The Poisson-Dirichlet law is the unique invariant distributions for uniform split-merge transformations. *Ann. Probab.*, 32, 1B, 915-938.
- Diaconis, P. and Shahshahani, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* 57, 159-179.

- Diaconis, P. and Graham, R.L. (1977). Spearman's footrule as a measure of disarray. *J. R. Statist. Soc. B*, 39, 262-268.
- Donnelly, P. and Grimmett, G. (1993) On the asymptotic distribution of large prime factors, *J. London Math. Soc.*, (2) 47, 395-404 for all transpositions
- Durrett, R. (1996) *Probability: Theory and Examples*, Second Edition, Duxbury Press
- Durrett, R. (2002) *Probability Models for DNA Sequence Evolution*. Springer-Verlag, New York
- Durrett, R. (2003) Shuffling Chromosomes. *J. Theor. Prob.* 16, 725–750
- Durrett, R. (2005) *Random Graphs*. In preparation
- Durrett, R. and Neuhauser, C. (1994) Particle systems and Reaction-Diffusion equations. *Ann. Prob.*, Vol. 22, No. 1, 289-333.
- Durrett, R., Nielsen, R., and York, T.L. (2003) Bayesian estimation of genomic distance. *Genetics*, to appear
- Erdős, P. and Renyi, A. (1960). On the evolution of random graphs *Acta Math. Inst. Hungar. Acad. Sci.*, 5, 17-61
- Eriksen, N. (2005) Expected number of inversions after a sequence of random adjacent transpositions - an exact expression. *Discrete Mathematics*
- Eriksson, H., Eriksson, K. and Sjöstrand, J. (2000) Expected number of inversions after  $k$  random adjacent transpositions. In D. Krob, A.A. Mikhalev, A.V. Mikhalev, eds. *Proceedings of Formal Power Series and Algebraic Combinatorics*, Springer-Verlag (2000) 677-685
- Fulman, J. (2004) Stein's method and minimum parsimony distance after shuffles. To appear. Available at <http://www.math.pitt.edu/~fulman/articles.html>
- Gamburd, A. Poisson-Dirichlet distribution for random Belyi surfaces. Preprint available at <http://arXiv:math.PR/0501283>
- Grimmett, G. (1999) *Percolation*. Springer, Grundlehren der mathematischen Wissenschaften, Vol. 321.
- Gromov, M. (1987) Hyperbolic Groups. *Math. Sci. Res. Inst. Publ.* 8, Springer, 75-263.

- Hannehalli, S. and Pevzner, P.A. (1995) Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). *Proceedings of the 27<sup>th</sup> Annual Symposium on the Theory of Computing*, 178–189. Full version in the *Journal of the ACM*. 46, 1–27
- Jacod, J. and Shiryaev, A. (1987) *Limit Theorems for Stochastic Processes* Springer New-York
- Janson, S., Knuth, D. E., Luczak, T. and Pittel, B. (1993) The birth of the giant component. *Rand. Struct. Algor.* 4, 231–358
- Janson, S., Luczak, T., and Rucinski, A. (2000) *Random Graphs*, Wiley - Interscience, New York
- Karoński, M. and Luczak, T. (1993). Random hypergraphs. In: *Combinatorics, Paul Erdős is eighty, Vol. 2*. (Keszthely, 1993), 283–293, Bolyai Soc. Math. Stud., 2, *János Bolyai Math. Soc.*, Budapest, 1996
- Karoński, M. and Luczak, T., (1997) The number of connected sparsely edged uniform hypergraphs. *Discrete Math.* 171, 153–167.
- Karoński, M. and Luczak, T., (2002) The phase transition in a random hypergraph. *J. Comp. Appl. Math.* 142, 125–135
- Kendall (1970) *Rank Correlation Methods*, 4th edn. London: Griffin.
- Knuth, D. (1973) *The Art of Computer Programming*, Vol. 2. reading, Mass.: Addison-Wiley.
- Luczak, T., Pittel, B., and Wierman, J. C. (1994) The structure of a random graph near the point of the phase transition. *Trans. Amer. Math. Soc.* 341, 721–748
- Lyons, R. and Peres, Y. (2005) *Probability on trees and networks*. In preparation. Available at <http://mypage.iu.edu/~rdlyons/#book>
- Mayer-Wolf, E. and Zeitouni, O. and Zerner, M. (2002) Asymptotics of certain coagulation - fragmentation processes and invariant Poisson-Dirichlet measures. *Electr. Journ. Prob.* 7, 1–25
- Molloy, M. and Reed, B. (1995) A critical point for random graphs with a given degree sequence. *Rand. Struct. and Algor.*, 6:161–179, 1995
- Pevzner, P.A. (2000) *Computational Molecular Biology: An Algorithmic Approach*. MIT Press, Cambridge

- Pevzner, P.A. and Tesler, G. (2003) Genome rearrangement in mammalian evolution: lessons from human and mouse genomes. *Genome Research*. 13, 37–45
- Pitman, J. (1998) Enumerations of trees and forests related to branching processes and random walks. *Microsurveys in Discrete Probability*, D. Aldous and J. Propp editors. DIMACS Ser. Discrete Math. Theoret. Comp. Sci no.41 163-180. Amer. Math. Soc. Providence RI.
- Pitman, J. (1999) Coalescent random forests, *J. Comb. Theory A*. 85 165-193.
- Pitman, J. (2002) Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Combin. Prob. Comput.* 11, 501–514
- Pitman, J. (2003) Combinatorial stochastic processes. *Lecture Notes for St. Flour Course*. Available at <http://stat-www.berkeley.edu/users/pitman/>
- Pittel, B. (1990) On tree census and the giant component in sparse random graphs, *Rand. Struct. Algor.*, **1**, 311–342
- Prat, J.J. (1971) Etude asymptotique du mouvement Brownien sur une variété Riemannienne à courbure négative. *C. R. Acad. Sci., Sér A*, 280, 1539-1542.
- Ranz, J.M. and Casals, F. and Ruiz, A. (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Research*. 11, 230–239
- Revuz, D. and Yor, M., (1999) *Continuous martingales and Brownian Motion*, Springer-Verlag, New York
- Saloff-Coste, L. (2003) Random Walks on Finite Groups. In: H. Kesten, ed. *Probability on Discrete Structures*, Encyclopaedia of Mathematical Sciences (110), Springer.
- Schramm, O. (2004) Composition of random transpositions, to appear
- Schmidt-Pruzan, J. and Shamir, E. (1985) Component structure in the evolution of random hypergraphs. *Combinatorica* 5, 81–94
- Sütő, A. (2002) Percolation transition in the Bose gas: II. *J. Phys. A: Math. Gen.*, 35, 6995-7002.

Tanner, J.C. (1961) A derivation of the Borel distribution. *Biometrika* 48, 222–224

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

York, T.L., Durrett, R., and Nielsen, R. (2002) Bayesian estimation of inversions in the history of two chromosomes. *J. Comp. Bio.* 9,808–818

Viràg, B. (2000) On the speed of random walks on graphs, *Ann. Probab.* 28, no. 1, 379–394.

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