THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Mixing times for neighbour transposition shuffles on graphs

Stefan Erikshed

CHALMERS | UNIVERSITY OF GOTHENBURG

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden 2011

Mixing times for neighbour transposition shuffles on graphs
Stefan Erikshed
ISSN 1652-9715
NO 2011:11

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Typeset with LaTeX.
Printed in Gothenburg, Sweden, 2011

Mixing times for neighbour transposition shuffles on graphs

Stefan Erikshed

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis will treat Markov chains on the symmetric group S_n, i.e. the set of permutations of n distinct objects. Markov chains with the state space S_n are often referred to as card shuffling chains. The state space is thus the reorderings of a deck of n cards. A certain type of card shuffling chains is considered here; neighbour transpositions on graphs. This is a generalization of ordinary random transpositions shuffle. Each step of the ordinary random transpositions shuffle consists of randomly selecting any pair of cards in the deck and then switch their places. Neighbour transpositions on a graph means that the n cards is placed on the vertices of a n-vertex graph. At each step a neighbour pair of cards in the graph (i.e. two cards at positions connected with an edge) is selected and transposed.

If the graph is connected, the deck will eventually be well mixed. In other words, the distribution of the chain converges to uniformity on S_n. This thesis deals with the rate of convergence to uniformity for card shuffling on two families of graphs, lollipop graphs and random graphs, $\mathcal{G}(n,p)$. More precisely, bounds on the mixing time of these shuffles is determined. The mixing time is the number of steps of the Markov chain until it is close to its stationary uniform distribution. As usual when dealing with convergence rate problems we let $|S_n| \to \infty$, yielding asymptotic results in n. Lower and upper bounds, both of order $n^4 \log n$, on the mixing time for neighbour transpositions on lollipop graphs is derived. Further, lower bounds of order $n \log n$, on the mixing time for neighbour transpositions on connected random graphs is established. Upper bounds of the same order is proved for random graphs with bounded diameter.

Keywords: card shuffling, mixing time, neighbour transpositions, convergence rate of Markov chains, lollipop graphs, random graphs.

Acknowledgements

I would like to thank my supervisor Johan Jonasson, for his care and help during the work with this thesis. I also want to thank my loving wife Zandra for always encouraging me. I give thanks to my savior Jesus Christ, who is worthy of praise and worship every day.

Stefan Erikshed
Gothenburg, April 2011

Contents

Chapter 1

Introduction

1.1 Background

This thesis will treat Markov chains on the symmetric group S_n, i.e. the set of permutations of n distinct objects. Markov chains with the state space S_n are often referred to as *card shuffling chains*. More specifically we are interested in the *mixing time* for a certain type of such card shuffling chains; *neighbour transpositions* on graphs. Any finite state irreducible aperiodic Markov chain approaches a unique stationary distribution (also called steady state, equilibrium or invariant distribution) as the number of steps increases. The mixing time of a Markov chain is the time until it is "close" to its steady state. There are several variants of definitions of the mixing time. The most common variant, and the one we will use, is the *variation distance mixing time*, where the closeness to stationarity is expressed in terms of the *total variation distance*. The total variation distance measures the distance between the distribution of the Markov chain and its stationary distribution. The formal definition of the total variation distance and the mixing time will be done in chapter 2.

An interesting fact is that for many classes of Markov chains there is a threshold time for the mixing process, where in a short time interval the distance to the stationary distribution drops quickly. Diaconis [9] has written a survey on this cutoff phenomenon. Card shuffling is one of the cases where this sharp cutoff in the convergence to stationarity often can be proved. When dealing with convergence rate problems, one usually lets the size of the state space tend to infinity in some way. In the determination of bounds on the mixing time of a Markov chain on S_n, one lets $n \to \infty$, implying that also $|S_n| = n! \to \infty$, see Definition 4 in chapter 2.

A multitude of different types of card shuffling chains, together with their mixing times, has been studied over the last decades. Important achievement in the searching for mixing times was made already in the 1980's. Worth to mention is the finding of the $\frac{3}{2}\log_2 n$ cutoff for the riffle shuffle (also known as the dovetail shuffle or the Gilbert-Shannon-Reeds shuffle) in 1983, by Aldous [1]. Aldous' result was later sharpened by Bayer and Diaconis, [2]. Since the riffle shuffle is a model for the most frequently used method to shuffle an ordinary deck of cards, an interesting result in Bayer and Diaconis paper is that they point out that one needs about seven shuffles to mix a 52 card deck. Another practically usable shuffle is the overhand shuffle, where you successively take small packets of cards from the top of the deck, piling them together in reverse order. This was first analyzed by Pemantle [18], and the correct order of the mixing time, $n^2 \log n$, was established by Jonasson [16].

However, the type of shuffle we will focus on is so called random transpositions. Each step of the ordinary random transposition shuffle consists of randomly selecting a pair of cards and then switch their places. If we allow for the possibility that the two selected cards are the same, we have a probability of $1/n$ for the identity shuffle, i.e. nothing happens. This also prevents periodicity. The random transpositions shuffle was treated by Diaconis and Shahshahani in [8]. They observed that most of the action in terms of the total variation distance to stationarity occurs around $\frac{1}{2} n \log n$. That $\frac{1}{2} n \log n$ really is a sharp threshold was proved by Matthews, via the method of strong stationary times, in [17].

In the ordinary random transpositions shuffle, any pair of cards can be chosen at each step. But we could also limit ourselves, and only allow for certain transpositions. One example of this is the transposing neighbours shuffle. At each step a pair of cards that sits next to each other in the deck is selected and swapped. An upper bound of correct order for the transposing neighbours shuffle was derived by Aldous [1]. Later Wilson [21] proved that $n^3 \log n$ is the right threshold value. In this paper, and the previous [20], Wilson developed the first systematic technique that gives tight lower bounds on the mixing time. Wilson's method uses the fact that a lower bound on the mixing time of a single card is also a lower bound for the whole chain. If we are able to find an eigenvalue close to 1, and corresponding eigenvector, to the transition matrix of the motion of a single card, we can use this eigenvector to construct a test function that provides a good lower bound on the mixing time of the whole Markov chain.

Shuffling by random transpositions can be thought of as a kind of random walk on a graph. Suppose we have a deck of n distinct cards. Place n cards on the vertices of a graph $G = (V, E)$ with $|V| = n$ vertices, one card on each vertex. Consider the process where we at each step with some positive probability do nothing (this is to avoid periodicity), and otherwise randomly choose one edge (i, j) of the graph and switch places of the objects on the vertex i and j, i.e. do the neighbour transposition $(i \ \ j)$. This process will induce the same Markov chain on the symmetric group S_n as the random transposition shuffle where the allowed transpositions are the ones that correspond to an open edge in the graph. We denote the set of transpositions generating the random neighbour transposition shuffle on the graph $G = (V, E)$ by $\mathfrak{E}(G) = \{(i \ \ j) | (i, j) \in E\}$.

In terms of shuffling on graphs, the ordinary random transposition shuffle thus corresponds to neighbour transpositions on the complete graph $G_n = (V_{G_n}, E_{G_n})$, where $|V_{G_n}| = n$ and $E_{G_n} = \{(i, j) | i, j \in V_{G_n}, i \neq j\}$, so that all vertex pairs are connected by open edges. Aldous' original neighbour transpositions described above will correspond to shuffling on a simple path with n vertices. Note that if the graph is connected, the Markov chain induced by neighbour transpositions will be irreducible. Moreover, shuffling chains on a connected graph will have a uniform stationary distribution, i.e. at stationarity we will have uniform distribution on all the $|S_n| = n!$ permutations in the symmetric group, denoted by π^n. We will consider neighbour transpositions on two families of graphs; *lollipop graphs*, in chapter 3 and *random graphs*, in chapter 4.

A lollipop graph is obtained by appending a complete graph, the "clique", to an end vertex of a simple path. The graph will then, as the name reveal, look like a lollipop. The structure of the lollipop graph, with an edge dense part and an edge sparse part, leads to some interesting properties of the graph. This constitution will make mixing process of the card shuffling chain unusually slow. The slowness properties of lollipop graphs has been observed earlier. Brightwell and Winkler [3] analyzed hitting times for simple random walks on n-vertex graphs, $\max_{i,j} h(i,j)$, where $h(i,j)$ is the expected number of steps in a random walk to reach vertex i starting from vertex j. They proved that the lollipop graph with $\lceil \frac{2n+1}{3} \rceil$ vertices in the clique actually is optimal in terms of hitting times for random walks on n-vertex graphs, i.e. no other graph can have a higher hitting time. Feige [12] made similar conclusions regarding the cover time for random walks, i.e. the expected number of steps it takes to visit all vertices. Furthermore, Jonasson [15] showed that lollipop graphs are extremal also for commute times, $\max_{i,j}[h(i,j)+h(j,i)]$, of simple random walks on n-vertex graphs.

When studying processes on graphs, random graph models are interesting in the sense of being a more appropriate models for "real" graphs and networks than the deterministic models. In general, a random graph is a graph that is generated by some random process. It could either be that the number of graph vertices, graph edges, or the connections between them are determined in some random way. Different models produce different probability distributions on graphs. One of the most common is the Erdős-Rényi model denoted, $\mathcal{G}(n,M)$, proposed by the two name givers in 1959, [10]. In this model a graph is chosen uniformly at random from the collection of all graphs which have n nodes and M edges.

A closely related model is the one suggested by Gilbert [13] in the same year, denoted by $\mathcal{G}(n,p)$. In Gilbert's model we have a fixed number of vertices, n, and between each pair of vertices an open edge occurs independently with probability p. Thus the number of edges is not fixed, like in the Erdős-Rényi model, but instead Bin$(\binom{n}{2},p)$-distributed. Since Gilbert's model is the most commonly used today, we will use this one in this thesis. As long as p is relatively large, the properties of the transposing neighbours chain on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$ is very similar to transposing neighbours on the complete graph. A very small edge probability p makes it more interesting to analyze neighbour transpositions on $G_{n,p}$.

Since we are dealing with shuffling on random graphs, the best we can achieve is results that are correct with probability 1, or in other words, *almost sure* results. And when determining the mixing time of neighbour transpositions on $G_{n,p}$ we are dealing with asymptotic analysis of a random process. Then, one says that a property holds *asymptotically almost surely* (a.a.s.) if, over a sequence of sets, the probability converges to 1. Thus, our goal is to derive asymptotically almost sure bounds on the mixing time.

In a second paper on random graphs by Erdős and Rényi, [11], they establish results equivalent to the following. If $p < \frac{(1-\epsilon)\log n}{n}$, a graph in $\mathcal{G}(n,p)$ will a.a.s. contain isolated vertices, and thus be disconnected. In contrast, if $p > \frac{(1+\epsilon)\log n}{n}$, the random graph is a.a.s. connected. Thus $p = \frac{\log n}{n}$ is a sharp threshold for connectedness of $\mathcal{G}(n,p)$. Now, as mentioned before, a card shuffling chain on a graph will eventually be well mixed, i.e. converge to uniformity, if and only if the graph is connected. We will stick to connected graphs, that is graph with edge probability $p > \frac{\log n}{n}$, in this thesis. For the case $p < \frac{\log n}{n}$, an idea for the future is to analyze card shuffling on the largest connected component of the graph.

Upper bounds on mixing times of card shuffling chains can be obtained in many different ways. The two basic techniques are strong stationary times and coupling times, these are however often hard to determine. There are other, more advanced techniques based on L^2-theory. One example of this is, and the one we will use here, is the so called comparison technique. It was developed by Diaconis and Saloff-Coste in [6], and later generalized to any reversible Markov chain in [7], by the same authors. The idea of the comparison technique is to bound the mixing time of a difficult Markov chain by comparing it with another Markov chain, with a known mixing time. This is done (at least in the theory behind the technique) with comparison of the eigenvalues of the two Markov chains and, and this comparison is in turn made via comparison of the *Dirichlet forms* corresponding to the two processes.

1.2 Summary

Here is a short summary of the contents and results of this thesis. Next chapter contains definitions of the total variation distance between two measures, and the L^2-norm of a measure. Furthermore we will define the mixing time of a Markov chain, and bounds of the mixing time in terms of total variation distance. Finally, the chapter contains the theorem of Wilson that is the key to his technique for lower bounds on the mixing time, together with a proof.

In chapter 3 the neighbour transposition shuffle on lollipop graphs is treated. We will consider the lollipop graph with n vertices, where the number of vertices in the linear part is $\frac{n}{a}$, for some $a > 1$. For simplicity suppose that $\frac{n}{a}$ is an integer. We call this the $\left(\left(1 - \frac{1}{a}\right) n, \frac{n}{a}\right)$-lollipop graph, and denote it by $L_{n,a} = (V_{L_{n,a}}, E_{L_{n,a}})$.

At each step of the process do nothing with probability $\frac{1}{|E_{L_{n,a}}|+1}$, and otherwise do a random neighbour transposition. The following theorem will be proved.

Theorem (Lollipop graphs). *The mixing time τ_{mix} for random neighbour transpositions on the $\left(\left(1 - \frac{1}{a}\right) n, \frac{n}{a}\right)$-lollipop graph, $a > 1$ such that $\frac{n}{a}$ is an integer, has the lower bound*

$$\tau_{mix} \geq (1 + o(1))\frac{(a-1)^2}{4a^4 x_0} n^4 \log n,$$

where x_0 is the smallest positive number that satisfies $(a-1)x_0 + \tan x_0 = 0$.

Moreover τ_{mix} has the upper bound

$$\tau_{mix} \leq (1 + o(1))\frac{(a-1)^2(2a-1)^2}{2a^6} n^4 \log n.$$

Thus, neighbour transpositions on the lollipop graph is an extremely slow shuffle, the mixing time is of order $n^4 \log n$. The lower bound on the mixing time is derived using the technique of Wilson, mentioned in the previous section. In section 3.3, the upper bound will be established via the comparison technique of Diaconis and Saloff-Coste. We will use card shuffling on the complete graph, ordinary random transpositions, as reference shuffle. The theoretical background for this technique will be treated in section 3.2.

Chapter 4 is devoted to neighbour transpositions on random graphs, $\mathcal{G}(n,p)$. Consider a realization of the random graph model, $G_{n,p} = (V_{G_{n,p}}, E_{G_{n,p}})$. The updating procedure for neighbour transpositions is as follows. At each step do nothing with probability $\frac{n}{|E_{G_{n,p}}|+n}$, and otherwise do a random neighbour transposition. The following result will be proved.

Theorem (Random graphs). *For random neighbour transpositions on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$, the mixing time τ_{mix} asymptotically almost surely has the following lower bound for the given ranges of $p < 1$.*

For $p = \omega\left(\frac{\log n}{n}\right)$,

$$\tau_{mix} \geq \frac{1-o(1)}{2} n \log n.$$

For $p = c\frac{\log n}{n}$, $c > 1$,

$$\tau_{mix} \geq \frac{1-o(1)}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n.$$

Moreover, τ_{mix} a.a.s. has the following upper bound for p such that $1 > p > n^{\delta-1}$, for some $\delta > 0$,

$$\tau_{mix} \leq Cn \log n,$$

for some large enough constant C.

To reach the lower bounds on τ_{mix}, we first consider the event

$$A = A^n = \{\text{at least } \log n \text{ cards are in their starting positions}\}.$$

Let $\{X_t^n\}_{t=0}^{\infty}$ be the Markov chain on S_n induced by neighbour transpositions on $G_{n,p}$. Then it is easy to prove that $t = t(n)$ such that $\mathbb{P}(X_{t(n)}^n \in A^n) \to 1$, is an a.a.s. lower bound on the mixing time. In section 4.1, lower bounds on the mixing time is established in this way for the two ranges of the edge probability p.

In section 4.2, the upper bound is derived, again via comparison with the shuffling chain on the complete graph. Note that the upper bound is limited to the case $p > n^{\delta-1}$, for some $\delta > 0$. This is equivalent to random graphs with bounded diameter, see for example [5].

Before we begin our searching for mixing times, we need some mathematical and notational preliminaries.

Chapter 2

Preliminaries

2.1 Definitions

To determine the mixing time of a Markov chain we need a measure of how close the chain to its stationary distribution. To measure the distance between two probability measures μ_1 and μ_2 on a finite state space S we will use the following two measures.

Definition 1 (Total variation distance)**.** *Let μ_1 and μ_2 be two probability measures on a finite state space S. Then the total variation distance between μ_1 and μ_2 is given by*

$$\|\mu_1 - \mu_2\|_{TV} = \frac{1}{2}\sum_{s\in S} |\mu_1(s) - \mu_2(s)| = \max_{A\subseteq S}(\mu_1(A) - \mu_2(A)).$$

Definition 2 (L^2-norm)**.** *Let μ_1 and μ_2 be two finite measures on a finite state space S. Then the L^2-norm of μ_1 with respect to μ_2 is*

$$\|\mu_1\|_2 = \|\mu_1\|_{L^2(\mu_2)} = \left(\sum_{s\in S} \frac{|\mu_1(s)|^2}{\mu_2(s)}\right)^{1/2}.$$

Using the Cauchy-Schwarz inequality we can prove that

$$\|\mu_1 - \mu_2\|_2 \geq 2\|\mu_1 - \mu_2\|_{TV}. \tag{2.1}$$

As a result, convergence in L^2 is stronger than convergence in total variation. Let $\{X_t\}_{t=0}^{\infty}$ be an irreducible aperiodic Markov chain with state space S. As mentioned in the Introduction, a fundamental fact about an irreducible aperiodic Markov chain, is that regardless of the initial state, the time-t distribution of the chain, $\mathbb{P}(X_t \in \cdot)$, converges to a unique steady state distribution π as t tends to infinity. The mixing time of a Markov chain may refer to any of several variant formalizations of the idea: how large must t be until $\mathbb{P}(X_t \in \cdot)$ is approximately π? In this thesis we will use the following definition of the mixing time.

Definition 3 (Mixing time)**.** *The mixing time τ_{mix} of a Markov chain $\{X_t\}$ with stationary distribution π is defined by*

$$\tau_{mix} = \inf\{t : \|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \leq 1/4\}. \tag{2.2}$$

Note that if we have $\|\mathbb{P}(X_\tau \in \cdot) - \pi\|_2 \leq 1/2$, then by (2.1) $\tau \geq \tau_{\text{mix}}$. In particular we have

$$\tau_{\text{mix}} \leq \hat{\tau} := \inf\{t : \|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 \leq 1/4\}. \tag{2.3}$$

We will deal with mixing times of Markov chains induced by card shuffling on graphs, where the number of cards, n, tends to infinity. Our goal is to derive lower and upper bounds on the mixing time of this sequence of Markov chains. The bounds on the mixing time are defined as follows.

Definition 4 (Bounds on mixing time)**.** *Let $\{X_t^n\}_{t=0}^\infty, n = 1, 2, \ldots,$ be a sequence of irreducible aperiodic Markov chains with state spaces S^n and stationary distributions π^n. Suppose $|S^n| \to \infty$ as $n \to \infty$. A sequence $\{t(n)\}_{n=1}^\infty$ is said to be a lower bound on the mixing time of the sequence of Markov chains if*

$$\liminf_{n\to\infty} \|\mathbb{P}(X_{t(n)}^n \in \cdot) - \pi^n\|_{TV} \geq 1/4,$$

and an upper bound if

$$\limsup_{n\to\infty} \|\mathbb{P}(X_{t(n)}^n \in \cdot) - \pi^n\|_{TV} \leq 1/4.$$

2.2 Wilson's theorem

To our aid when deriving lower bounds on mixing time we have a technique introduced by Wilson in [20] and [21]. The first step is to find an eigenvector to the transition matrix of the motion a single card, with eigenvalue close to 1. This eigenvector can then be used to build an eigenvector for the whole Markov chain, which in turn can be used to lower bound the mixing time. Below follows the theorem that is the key to Wilson's technique.

Let $1 - \gamma$ be an eigenvalue of the transition matrix for a Markov chain $\{X_t\}$, and Φ be a right eigenvector, i.e. a function on S s.t. almost surely

$$\mathbb{E}[\Phi(X_{t+1})|X_t] = (1 - \gamma)\Phi(X_t).$$

When $\{X_t\}$ is reversible, it can be proved that the eigenvector Φ and the eigenvalue $1 - \gamma$ are real-valued. Assume that $\gamma \in (0, 1/2)$. Define R by

$$R := \max_{s\in S} \mathbb{E}[(\Phi(X_{t+1}) - \Phi(X_t))^2|X_t = s].$$

Theorem 1 (Wilson). *For a fixed $\varepsilon > 0$ let*

$$T = \frac{\log \Phi(X_0) - \frac{1}{2}\log \frac{4R}{\gamma\varepsilon}}{-\log(1-\gamma)}. \tag{2.4}$$

Then $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \geq 1 - \varepsilon$ for all $t \leq T$.

Proof. We have

$$\mathbb{E}[\Phi(X_t)] = \mathbb{E}[\mathbb{E}[\Phi(X_t)|X_{t-1}]] = (1-\gamma)\mathbb{E}[\Phi(X_{t-1})].$$

Applying this inductively yields

$$\mathbb{E}[\Phi(X_t)] = (1-\gamma)^t \mathbb{E}[\Phi(X_0)]. \tag{2.5}$$

Thus $\mathbb{E}[\Phi(X_t)|X_0] = (1-\gamma)^t\Phi(X_0)$. Let $X = \lim_{t\to\infty} X_t$, so that X has the stationary distribution π. From (2.5) we get $\mathbb{E}[\Phi(X)] = 0$.

With the notation $\triangle\Phi_t := \Phi(X_{t+1}) - \Phi(X_t)$, we have $\Phi(X_{t+1})^2 = \Phi(X_t)^2 + 2\Phi(X_t)\triangle\Phi_t + (\triangle\Phi_t)^2$. Using that $\mathbb{E}[\triangle\Phi_t|X_t] = -\gamma\Phi(X_t)$, and $\mathbb{E}[(\triangle\Phi_t)^2|X_t] \leq R$, we get

$$\mathbb{E}[\Phi(X_{t+1})^2|X_t] \leq (1-2\gamma)\Phi(X_t)^2 + R.$$

Hence

$$\begin{aligned} \mathbb{E}[\Phi(X_t)^2] &= \mathbb{E}[\mathbb{E}[\Phi(X_t)^2|X_{t-1}]] \\ &\leq (1-2\gamma)\mathbb{E}[\Phi(X_{t-1})^2] + R, \end{aligned}$$

and by induction

$$\begin{aligned} \mathbb{E}[\Phi(X_t)^2] &\leq (1-2\gamma)^t\mathbb{E}[\Phi(X_0)^2] + R\sum_{i=0}^{t-1}(1-2\gamma)^i \\ &\leq (1-2\gamma)^t\mathbb{E}[\Phi(X_0)^2] + \frac{R}{2\gamma}. \end{aligned} \tag{2.6}$$

In the last inequality we use the geometric sum and the assumption that $\gamma \in (0, 1/2)$. For a given starting value X_0, (2.5) and (2.6) implies

$$\begin{aligned} \mathbf{Var}(\Phi(X_t)) &= \mathbb{E}[\Phi(X_t)^2] - \mathbb{E}[\Phi(X_t)]^2 \\ &\leq \left((1-2\gamma)^t - (1-\gamma)^{2t}\right)\Phi(X_0)^2 + \frac{R}{2\gamma} \\ &\leq \frac{R}{2\gamma}. \end{aligned}$$

The last inequality follows from $(1-\gamma)^2 = 1-2\gamma+\gamma^2 \geq 1-2\gamma$, and thus, since $1-2\gamma > 0$, $(1-2\gamma)^t - (1-\gamma)^{2t} \leq 0$. By Chebyshev's inequality

$$\mathbb{P}\left(|\Phi(X_t) - \mathbb{E}[\Phi(X_t)]| \geq \sqrt{\frac{R}{\gamma\varepsilon}}\right) \leq \frac{\varepsilon}{2}. \tag{2.7}$$

Let $A = \left\{s \in S | \Phi(s) \geq \sqrt{\frac{R}{\gamma\varepsilon}}\right\}$. Then we have

$$\begin{aligned}
\pi(A) &= \mathbb{P}\left(\Phi(X) \geq \sqrt{\frac{R}{\gamma\varepsilon}}\right) \\
&\leq \mathbb{P}\left(|\Phi(X)| \geq \sqrt{\frac{R}{\gamma\varepsilon}}\right) \\
&\leq \frac{\varepsilon}{2},
\end{aligned}$$

where the last inequality comes from letting t tend to infinity in (2.7), and the fact that at stationarity we have $\mathbb{E}[\Phi(X)] = 0$.

Now, let t be such that $\mathbb{E}[\Phi(X_t)] \geq \sqrt{\frac{4R}{\gamma\varepsilon}}$. Then we have

$$\begin{aligned}
\mathbb{P}(X_t \in A) &= \mathbb{P}\left(\Phi(X_t) \geq \sqrt{\frac{R}{\gamma\varepsilon}}\right) \\
&\geq \mathbb{P}\left(\Phi(X_t) - \mathbb{E}[\Phi(X_t)] \geq \sqrt{\frac{R}{\gamma\varepsilon}} - \sqrt{\frac{4R}{\gamma\varepsilon}}\right) \\
&= 1 - \mathbb{P}\left(\Phi(X_t) - \mathbb{E}[\Phi(X_t)] < -\sqrt{\frac{R}{\gamma\varepsilon}}\right) \\
&\geq 1 - \mathbb{P}\left(|\Phi(X_t) - \mathbb{E}[\Phi(X_t)]| > \sqrt{\frac{R}{\varepsilon}}\right) \\
&\underset{(2.7)}{\geq} 1 - \frac{\varepsilon}{2}.
\end{aligned}$$

Thus, for such t such that $\mathbb{E}[\Phi(X_t)] \geq \sqrt{\frac{4R}{\gamma\varepsilon}}$, we have

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} = \max_{A \subseteq S}(\mathbb{P}(X_t \in A) - \pi(A)) \geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon.$$

Finally, observe that for all $t \leq T = \frac{\log \Phi(X_0) - \frac{1}{2}\log\frac{4R}{\gamma\varepsilon}}{-\log(1-\gamma)}$,

$$
\begin{aligned}
\mathbb{E}[\Phi(X_t)] &= (1-\gamma)^t \Phi(X_0) \geq (1-\gamma)^T \Phi(X_0) \\
&= (1-\gamma)^{-\frac{1}{\log(1-\gamma)} \log\left(\Phi(X_0) \middle/ \sqrt{\frac{4R}{\gamma\varepsilon}}\right)} \Phi(X_0) \\
&= e^{-\log\left(\Phi(X_0) \middle/ \sqrt{\frac{4R}{\gamma\varepsilon}}\right)} \Phi(X_0) = \sqrt{\frac{4R}{\gamma\varepsilon}},
\end{aligned}
$$

and we have proved that $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \geq 1 - \varepsilon$ for all $t \leq T$.

□

Now, taking for example $\varepsilon = \frac{1}{2}$ in Theorem 1 we get $\|\mathbb{P}(X_T \in \cdot) - \pi\|_{TV} \geq \frac{1}{2}$. From Definition 4 we see that

$$
\tau_{\text{mix}} \geq T = \frac{\log \Phi(X_0) - \frac{1}{2}\log\frac{8R}{\gamma}}{-\log(1-\gamma)}, \tag{2.8}
$$

thus we have a lower bound on the mixing time for the Markov chain $\{X_t\}$.

Chapter 3

Neighbour transpositions on lollipop graphs

In this chapter we treat random neighbour transpositions on the lollipop graph. The lollipop graph consists of a ”clique” of vertices where all vertex pairs are connected with edges, i.e. a complete subgraph, and to this clique attached a simple path. We will consider the lollipop graph with n vertices, where the number of vertices in the linear part is $\frac{n}{a}$, including the vertex that connects the path and the complete graph, for some $a > 1$ s.t. $\frac{n}{a}$ is an integer. We call this the $\left(\left(1-\frac{1}{a}\right)n, \frac{n}{a}\right)$-lollipop graph, and denote it by $L_{n,a} = (V_{L_{n,a}}, E_{L_{n,a}})$.

We will se that the partition of the graph into an edge dense part and an edge sparse part, makes the convergence of the card shuffling chain on the graph extremely slow, the mixing time is of order $n^4 \log n$. In the schematic picture below, note that in the clique on the left, all vertices are connected to each other. Number the vertices so that vertex 1 to $\frac{n}{a}$ is in the linear part, with $\frac{n}{a}$ as connection to the complete part, which in turn consists of vertices $\frac{n}{a}+1$ to n.

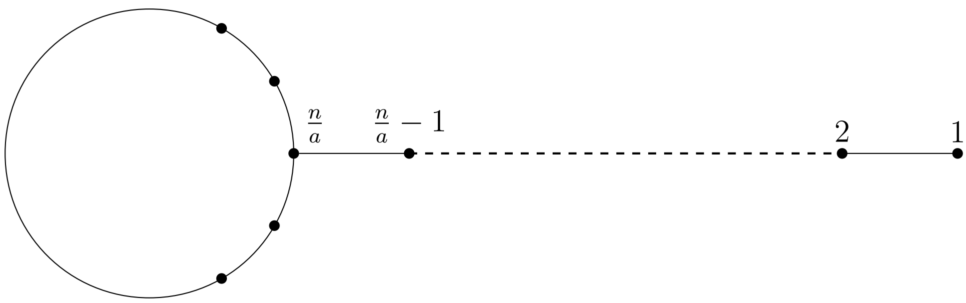

Figure 3.1: The $\left(\left(1-\frac{1}{a}\right)n, \frac{n}{a}\right)$-lollipop graph.

Let $m = |E_{L_{n,a}}| + 1$ (the $+1$ is for convenient computations later on). It is not hard to derive the following expression for $m = m(a)$.

$$m = \frac{1}{2}\left(1-\frac{1}{a}\right)^2 n^2 + \frac{a+1}{2a}n \tag{3.1}$$

At each step of the card shuffling process, do nothing with probability $1/m$, and otherwise do a random neighbour transposition. Let $\{X_t^n\}_{t=0}^{\infty}$ be the Markov chain on the symmetric group S_n, induced by this random process. In section 3.1 to 3.3 we will prove the following theorem.

Theorem 2. *Let $a > 1$ be a constant such that $\frac{n}{a}$ is an integer. The mixing time τ_{mix} for random neighbour transpositions on the $\left(\left(1-\frac{1}{a}\right)n, \frac{n}{a}\right)$-lollipop graph has the following bounds*

$$(1+o(1))\frac{(a-1)^2}{4a^4x_0}n^4\log n \leq \tau_{\mathit{mix}} \leq (1+o(1))\frac{(a-1)^2(2a-1)^2}{2a^6}n^4\log n,$$

where x_0 is the smallest positive number that satisfies $(a-1)x_0 + \tan x_0 = 0$.

3.1 Lower bound on the mixing time on lollipop graphs

Lemma 1. *For random neighbour transpositions on the $\left(\left(1-\frac{1}{a}\right)n, \frac{n}{a}\right)$-lollipop graph, $a > 1$, we have*

$$\tau_{\mathit{mix}} \geq (1+o(1))\frac{(a-1)^2}{4a^4x_0}n^4\log n,$$

where x_0 is the smallest positive number that satisfies $(a-1)x_0 + \tan x_0 = 0$.

Proof. First, consider the motion of a single card. This is in itself a Markov chain. Denote the transition matrix for this chain by $\boldsymbol{P} = [p_{ij}]_{n\times n}$. Put $1-\gamma$ for an eigenvalue of this chain, and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ for corresponding eigenvector. We have $(1-\gamma)v_k = \sum_{i=1}^{n} p_{ki}v_i$, yielding the following system of equations for $(1-\gamma, \boldsymbol{v})$.

$$\begin{cases} (1-\gamma)v_1 = \left(1-\frac{1}{m}\right)v_1 + \frac{1}{m}v_2 & \\ (1-\gamma)v_k = \left(1-\frac{2}{m}\right)v_k + \frac{1}{m}v_{k-1} + \frac{1}{m}v_{k+1}, & 2 \leq k \leq \frac{n}{a}-1 \\ (1-\gamma)v_{\frac{n}{a}} = \left(1 - \frac{1+\left(1-\frac{1}{a}\right)n}{m}\right)v_{\frac{n}{a}} + \frac{1}{m}v_{\frac{n}{a}-1} + \frac{1}{m}\displaystyle\sum_{i=\frac{n}{a}+1}^{n} v_i & \\ (1-\gamma)v_k = \left(1 - \frac{\left(1-\frac{1}{a}\right)n}{m}\right)v_k + \frac{1}{m}\displaystyle\sum_{\substack{i=\frac{n}{a}\\ i\neq k}}^{n} v_i, & \frac{n}{a}+1 \leq k \leq n \end{cases} \tag{3.2}$$

Since by symmetry $v_n = v_{n-1} = \ldots = v_{\frac{n}{a}+1}$ the two last equations reduce to

$$\begin{cases} (1-\gamma)v_{\frac{n}{a}} = \left(1 - \frac{1+\left(1-\frac{1}{a}\right)n}{m}\right)v_{\frac{n}{a}} + \frac{1}{m}v_{\frac{n}{a}-1} + \frac{\left(1-\frac{1}{a}\right)n}{m}v_n & \\ (1-\gamma)v_k = \left(1 - \frac{\left(1-\frac{1}{a}\right)n}{m}\right)v_k + \frac{1}{m}v_{\frac{n}{a}} + \frac{\left(1-\frac{1}{a}\right)n-1}{m}v_n, & \frac{n}{a}+1 \leq k \leq n. \end{cases} \tag{3.3}$$

First we focus on the equations for the linear part of the graph, the first two equations of system (3.2). The second equation is a recurrence relation with characteristic equation

$$(1-\gamma)r = \left(1 - \frac{2}{m}\right) r + \frac{1}{m} + \frac{1}{m} r^2. \tag{3.4}$$

This yields the solutions

$$r_{1,2} = 1 - \frac{m\gamma}{2} \pm \sqrt{m\gamma\left(\frac{m\gamma}{4} - 1\right)}.$$

Assuming that $m\gamma \to 0$ as $n \to \infty$ we get complex solutions $r_{1,2} = \varrho e^{\pm\omega i}$, where $\varrho := |r_1| = 1$ and $\omega := \arctan\left(\frac{\sqrt{m\gamma\left(1-\frac{m\gamma}{4}\right)}}{1-\frac{m\gamma}{2}}\right)$.

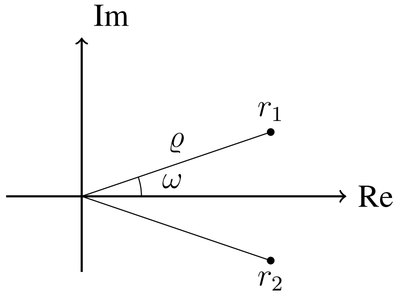

Figure 3.2: Solutions r_1 and r_2 to the characteristic equation (3.4), in the complex plane.

Note that $\cos\omega = 1 - \frac{m\gamma}{2}$ and $\sin\omega = \sqrt{m\gamma\left(1 - \frac{m\gamma}{4}\right)}$. The recurrence relation for v_k in the linear part has the following general solution.

$$v_k = \varrho^k \left(C_1 \cos(k\omega) + C_2 \sin(k\omega)\right) = C_1 \cos(k\omega) + C_2 \sin(k\omega), \quad 1 \le k \le \frac{n}{a}$$

Scale the eigenvector so that $v_1 = 1$. From the first equality in (3.2) we then get $v_2 = 1 - m\gamma$. Thus

$$v_1 = C_1 \cos(\omega) + C_2 \sin(\omega) = 1, \tag{3.5}$$
$$v_2 = C_1 \cos(2\omega) + C_2 \sin(2\omega) = 1 - m\gamma. \tag{3.6}$$

With these two boundary conditions we can compute the constants C_1 and C_2. Inserting $C_2 = \frac{1 - C_1 \cos\omega}{\sin\omega}$ into (3.6) yields

$$\begin{aligned}
&C_1 \cos(2\omega) + \frac{1 - C_1\cos\omega}{\sin\omega} \sin(2\omega) = 1 - m\gamma \quad \Leftrightarrow \\
&C_1(2\cos^2\omega - 1) + (1 - C_1\cos\omega) 2\cos\omega = 1 - m\gamma \quad \Leftrightarrow \\
&-C_1 + 2\cos\omega = 1 - m\gamma \quad \Leftrightarrow \\
&C_1 = 1,
\end{aligned}$$

where in the last step we use that $\cos\omega = 1 - \frac{m\gamma}{2}$. Furthermore, inserting $C_1 = 1$ into (3.5) gives

$$C_2 = \frac{1-\cos\omega}{\sin\omega} = \frac{m\gamma/2}{\sqrt{m\gamma\left(1-\frac{m\gamma}{4}\right)}} = \sqrt{\frac{m\gamma}{4-m\gamma}}.$$

Thus, we have

$$v_k = \cos(k\omega) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin(k\omega), \quad 1 \le k \le \frac{n}{a}. \tag{3.7}$$

We will obtain an eigenvalue $1-\gamma$ via two different expressions for v_n, derived from (3.3). Rewriting the first equation, and taking $k = n$ in the second equation we get

$$\begin{cases} v_n = \left(1 + \frac{1-m\gamma}{\left(1-\frac{1}{a}\right)n}\right) v_{\frac{n}{a}} - \frac{1}{\left(1-\frac{1}{a}\right)n} v_{\frac{n}{a}-1} \\ v_n = \frac{1}{1-m\gamma} v_{\frac{n}{a}} \end{cases}$$

Putting equality between the two expressions for v_n gives

$$\left(1 + \frac{1-m\gamma}{\left(1-\frac{1}{a}\right)n} - \frac{1}{1-m\gamma}\right) v_{\frac{n}{a}} - \frac{1}{\left(1-\frac{1}{a}\right)n} v_{\frac{n}{a}-1} = 0.$$

Multiply the above equality by $1-m\gamma$. We want to find the root of the following function of γ.

$$f(\gamma) = \left(\frac{(1-m\gamma)^2}{\left(1-\frac{1}{a}\right)n} - m\gamma\right) v_{\frac{n}{a}} - \frac{1-m\gamma}{\left(1-\frac{1}{a}\right)n} v_{\frac{n}{a}-1}$$

With the expressions for $v_{\frac{n}{a}-1}$ and $v_{\frac{n}{a}}$ from (3.7) we get

$$\begin{aligned} f(\gamma) \;=\; & \left(\frac{(1-m\gamma)^2}{\left(1-\frac{1}{a}\right)n} - m\gamma\right)\left(\cos\left(\frac{n}{a}\omega\right) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin\left(\frac{n}{a}\omega\right)\right) - \\ & - \frac{1-m\gamma}{\left(1-\frac{1}{a}\right)n}\left(\cos\left(\left(\frac{n}{a}-1\right)\omega\right) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin\left(\left(\frac{n}{a}-1\right)\omega\right)\right). \end{aligned}$$

Assume that $f(\gamma)$ has a root $\gamma = \frac{a^2x^2}{mn^2}$, for some constant $x > 0$, leaving us with the task to derive the value of x. With this γ we get $\omega = \arctan\left(\frac{\sqrt{m\gamma\left(1-\frac{m\gamma}{4}\right)}}{1-\frac{m\gamma}{2}}\right) = \frac{ax}{n} + O(n^{-3})$, from the power series expansion of $\arctan(\cdot)$. Thus

$$
\begin{aligned}
f(\gamma) &= \left(\frac{1}{\left(1-\frac{1}{a}\right)n} - \frac{a^2x^2}{n^2} + O(n^{-3})\right)\Bigg(\cos\left(x+O(n^{-2})\right) + \\
&\quad + \sqrt{\frac{a^2x^2/n^2}{4-a^2x^2/n^2}}\sin\left(x+O(n^{-2})\right)\Bigg) - \left(\frac{1}{\left(1-\frac{1}{a}\right)n} + O(n^{-3})\right)\cdot \\
&\quad \cdot\left(\cos\left(x-\frac{ax}{n}+O(n^{-2})\right) + \sqrt{\frac{a^2x^2/n^2}{4-a^2x^2/n^2}}\sin\left(x-\frac{ax}{n}+O(n^{-2})\right)\right) = \\
&= \left(\frac{1}{\left(1-\frac{1}{a}\right)n} - \frac{a^2x^2}{n^2} + O(n^{-3})\right)\left(\cos x + \frac{ax}{2n}\sin x + O(n^{-2})\right) - \\
&\quad - \left(\frac{1}{\left(1-\frac{1}{a}\right)n} + O(n^{-3})\right)\left(\cos x + \frac{ax}{n}\sin x + \frac{ax}{2n}\sin x + O(n^{-2})\right) = \\
&= \left(-\frac{ax}{1-\frac{1}{a}}\sin x - a^2x^2\cos x\right)n^{-2} + O(n^{-3}).
\end{aligned}
$$

Some algebra shows that $f(\gamma) = 0$ is equivalent to

$$(a-1)x + \tan x + O(n^{-1}) = 0. \tag{3.8}$$

Suppose x_0 is a solution to $(a-1)x + \tan x = 0$. Then it is not hard to prove that equation (3.8) must have a solution between $x_0\left(1 - \frac{1}{\log n}\right)$ and $x_0\left(1 + \frac{1}{\log n}\right)$. Thus we have a root to $f(\gamma)$

$$\gamma = (1+o(1))\frac{a^2x_0^2}{mn^2}. \tag{3.9}$$

Since we are looking for γ as small as possible we take the smallest positive solution to $(a-1)x + \tan x = 0$. The table below gives this solution for some different values of a.

a	Smallest root to $(a-1)x + \tan x$
$\downarrow 1$	$\uparrow \pi$
1.001	3.13845...
1.01	3.11049...
2	2.02875...
3	1.83659...
4	1.75816...
5	1.71550...
10	1.63850...
100	1.57720...
1000	1.57143...
$\uparrow \infty$	$\downarrow \frac{\pi}{2}$

This means that the Markov chain induced by the motion of a single card has an eigenvalue $1-\gamma = 1 - \frac{a^2x_0^2}{mn^2}(1+o(1))$, where x_0 is the smallest positive solution to $(a-1)x + \tan x = 0$, and an eigenvector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ where

$$v_k = \begin{cases} \cos(k\omega) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin(k\omega) & 1 \le k \le \frac{n}{a} \\ \frac{1}{1-m\gamma}\left(\cos(\frac{n}{a}\omega) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin(\frac{n}{a}\omega)\right) & \frac{n}{a}+1 \le k \le n, \end{cases} \tag{3.10}$$

with

$$\omega = \arctan\left(\frac{\sqrt{m\gamma\left(1-\frac{m\gamma}{4}\right)}}{1-\frac{m\gamma}{4}}\right).$$

Now let Z_t^j denote the position of card j at time t, and put $\Phi^j(X_t) = v_{Z_t^j}$. Then

$$\mathbb{E}[\Phi^j(X_{t+1})|X_t] = (1-\gamma)\Phi^j(X_t). \tag{3.11}$$

Put

$$\Phi(X_t) = \sum_{j=1}^{\lfloor n/2 \rfloor} \Phi^j(X_t).$$

By linearity of expectation,

$$\mathbb{E}[\Phi(X_{t+1})|X_t] = (1-\gamma)\Phi(X_t).$$

Hence, we can use $\Phi(X_t)$ as test function in Theorem 1. To get a good lower bound $T = \frac{\log\Phi(X_0) - \frac{1}{2}\log\frac{4R}{\gamma\varepsilon}}{-\log(1-\gamma)}$, we want to start with the cards in an order that maximizes $\Phi(X_0)$. Choosing positions for the $\lfloor n/2 \rfloor$ first cards carefully we can certainly reach $\Phi(X_0) \ge C_a n$ for some constant $C_a > 0$. Just place the $\lfloor n/2 \rfloor$ first cards on the positions with the largest entries in the eigenvector $\boldsymbol{v}$. Next we want to bound $R = \max_{s\in S}\mathbb{E}[(\Phi(X_{t+1}) - \Phi(X_t))^2|X_t = s]$. Note that $\Phi(X_{t+1})$ and $\Phi(X_t)$ differs only if step t consists of a transposition of one of the cards $1, 2, \ldots, \lfloor n/2 \rfloor$ and one of the other cards. The transposition also have to occur in the linear part of the graph, or involve vertex $\frac{n}{a}$. The probability for this condition is $\frac{n-1}{m}$.

Consider all such transpositions, and look the expressions in (3.10). Since the change of rate for the dominating cosine function is greatest for argument around $\frac{\pi}{2}$, we see that $(\Phi(X_{t+1}) - \Phi(X_t))^2$ is maximized if the transposition involves the card at position k such that $k\omega \approx \frac{\pi}{2}$. Thus

$$
\begin{aligned}
R &\leq \frac{n-1}{m}\left(v_{\lfloor \frac{\pi}{2\omega} \rfloor} - v_{\lfloor \frac{\pi}{2\omega} \rfloor + 1}\right)^2 \\
&= \frac{(1+o(1))n}{m}\Bigg(\left(\cos\frac{\pi}{2} + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin\frac{\pi}{2}\right) - \\
&\qquad - \left(\cos\left(\frac{\pi}{2}+\omega\right) + \sqrt{\frac{m\gamma}{4-m\gamma}}\sin\left(\frac{\pi}{2}+\omega\right)\right)\Bigg)^2 \\
&= \frac{(1+o(1))n}{m}\left(\sqrt{\frac{m\gamma}{4-m\gamma}} + \sin\omega - \sqrt{\frac{m\gamma}{4-m\gamma}}\cos\omega\right)^2 \\
&= \frac{(1+o(1))n}{m}\left(\sqrt{\frac{m\gamma}{4-m\gamma}} + \sqrt{m\gamma\left(1-\frac{m\gamma}{4}\right)} - \sqrt{\frac{m\gamma}{4-m\gamma}}\left(1-\frac{m\gamma}{2}\right)\right)^2 \\
&= (1+o(1))\frac{n}{m}m\gamma \\
&= (1+o(1))\frac{a^2x_0^2}{mn}. \qquad (3.12)
\end{aligned}
$$

From (3.1) we have that $m = \frac{1+o(1)}{2}\left(1-\frac{1}{a}\right)^2 n^2$. Inserting this into (3.9) gives

$$\gamma = (1+o(1))\frac{a^2x_0^2}{mn^2} = (1+o(1))\frac{2a^4x_0}{(a-1)^2n^4},$$

which in turn, together with (3.12), entails

$$\frac{R}{\gamma} \leq (1+o(1))\frac{a^2x_0^2/mn}{a^2x_0^2/mn^2} = (1+o(1))n.$$

We are ready to sum up our calculations of a lower bound on the mixing time. From (2.8) we get

$$
\begin{aligned}
\tau_{\text{mix}} \geq T &= \frac{\log\Phi(X_0) - \frac{1}{2}\log\frac{4R}{\gamma\varepsilon}}{-\log(1-\gamma)} \\
&\geq \frac{\log C_a n - \frac{1}{2}\log((1+o(1))8n)}{-\log(1-(1+o(1))\frac{2a^4x_0}{(a-1)^2n^4})} \\
&= (1+o(1))\frac{(a-1)^2n^4}{2a^4x_0}\left(\log n - \frac{1}{2}\log n\right) \\
&= (1+o(1))\frac{(a-1)^2}{4a^4x_0}n^4\log n.
\end{aligned}
$$

Where we have used that $\frac{1}{-\log(1-\gamma)} = \frac{1+o(1)}{\gamma}$ as $\gamma \to 0$. Thus we have a lower bound on the mixing time of neighbour transpositions on lollipop graphs of order $n^4\log n$.

□

3.2 The comparison lemma

We will use comparison with shuffling on the complete graph G_n to establish an upper bound on the mixing time, τ_{mix}, for random neighbour transpositions on lollipop graphs. The same method will be used also to bound the mixing time on random graphs from above, in section 4.2. To this end we will need following definitions and lemmas, which are valid for random walks on any finite group. Lemma 2 and 3 were stated by Diaconis and Saloff-Coste, [6]. First we define the Dirichlet form.

Definition 5 (Dirichlet form). *The Dirichlet form, $\mathcal{E}(\varphi, \varphi)$, associated with a measure ν on the state space S, of a function φ on S is in discrete time given by*

$$\mathcal{E}(\varphi, \varphi) = \sum_{x \in S} \sum_{y \in S} \nu(y)(\varphi(x) - \varphi(xy))^2.$$

In Chapter 2 we defined the L^2-norm of measures on a finite state space S. For a function φ on S, we define the L^2-norm as follows.

Definition 6 (L^2-norm of function). *Let φ be a function on the state space S of the measure π. Then the L^2-norm of φ with respect to π is*

$$\|\varphi\|_2 = \left(\sum_{s \in S} |\varphi(s)|^2 \pi(s) \right)^{1/2}.$$

(Note that then the L^2-norm of the measure ν with respect to π can equivalently to Definition 2 be defined as the L^2-norm of the function $\varphi(s) = \nu(s)/\pi(s)$.)

Consider two symmetric probability measures on a group S, μ and μ^b, with Dirichlet forms $\mathcal{E}$ and $\mathcal{E}^b$. Let $n := |S|$. Furthermore, let $\{X_t\}_{t=0}^{\infty}$ and $\{X_t^b\}_{t=0}^{\infty}$ be the random walks generated by these, with uniform stationary distribution π. Let $1 = \kappa_1 \geq \kappa_2 \geq \ldots \geq \kappa_n \geq -1$ be the ordered eigenvalues of the discrete time transition matrix $\boldsymbol{P} = [p_{xy}]_{x,y \in S}$ for μ. Furthermore, let $0 = \lambda_1 \leq \lambda_2 \leq \ldots$, be the eigenvalues of $-\boldsymbol{Q} = -[q_{xy}]_{x,y \in S}$, where q_{xy}, $x \neq y$, is the intensity for a jump from x to y in the corresponding continuous time random walk. In the continuous version one lets the time between the steps be exponential with intensity 1, and $q_{xx} = -\sum_{y \neq x} q_{xy}$. Note that since $q_{xy} = p_{xy}$ for $x \neq y$, we have $\lambda_i = 1 - \kappa_i$.

The measure μ^b will work as a benchmark measure to compare with. The eigenvalues for the corresponding discrete and continuous random walks are denoted κ_i^b and λ_i^b.

The Dirichlet form can be used for the so called extremal characterization of the eigenvalues λ_i,

$$\lambda_i = \max_{\{W : \dim W = n-i+1\}} \min_{\varphi \in W} \frac{\mathcal{E}(\varphi, \varphi)}{\|\varphi\|_2^2}, \tag{3.13}$$

see for example Horn and Johnson [14], page 176.

Lemma 2 (Comparison Lemma)**.** *Suppose that* $A > 0$ *is a constant such that* $\mathcal{E}^b \leq A\mathcal{E}$*, then*

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 \leq n\kappa_n^{2t} + ne^{-t/A} + \|\mathbb{P}(X^b_{\lfloor t/2A \rfloor} \in \cdot) - \pi\|_2^2. \tag{3.14}$$

Proof. Since the stationary distribution for $\{X_t\}$ is uniform, the transition matrix $\boldsymbol{P}$ is symmetric. Thus $\boldsymbol{P}$ has the eigendecomposition

$$\boldsymbol{P} = \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}^{-1},$$

where $\boldsymbol{U}$ is the matrix with the orthonormal left eigenvectors $\phi_1, \phi_2, \ldots, \phi_n$ as columns, and $\boldsymbol{K}$ is the diagonal matrix with the eigenvalues $\kappa_1, \kappa_2, \ldots, \kappa_n$ on the diagonal, in the same order as the eigenvectors in $\boldsymbol{U}$. Note also that since $\boldsymbol{U}$ is orthonormal, so that $\boldsymbol{U}^{-1} = \boldsymbol{U}'$. Thus we can write powers of $\boldsymbol{P}$ like

$$\boldsymbol{P}^t = \boldsymbol{U}\boldsymbol{K}^t\boldsymbol{U}',$$

where $\boldsymbol{K}^t$ is the diagonal matrix with κ_i^t, $i = 1, 2, \ldots n$ as diagonal elements. Note also that the eigenvector corresponding to the first eigenvalue, $\kappa_1 = 1$, is the uniform vector $\phi_1 = \sqrt{n}\pi = \left(\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}\right)$. The starting distribution $\mathbb{P}(X_0 \in \cdot)$ is a vector with 1 at some position j, and 0's elsewhere. If we place ϕ_1 in column j of the eigenvector matrix $\boldsymbol{U}$, we get

$$\mathbb{P}(X_t \in \cdot) = \mathbb{P}(X_0 \in \cdot)\boldsymbol{P}^t = \mathbb{P}(X_0 \in \cdot)\boldsymbol{U}\boldsymbol{K}^t\boldsymbol{U}' = \phi_1\boldsymbol{K}^t\boldsymbol{U}'.$$

Also, $\phi_1\boldsymbol{K}^t$ is the row vector with entries $\frac{1}{\sqrt{n}}\kappa_i^t$, $i = 1, 2, \ldots, n$, ordered in the same way as the corresponding eigenvectors on the rows of $\boldsymbol{U}'$. Thus

$$\mathbb{P}(X_t \in \cdot) = \phi_1\boldsymbol{K}^t\boldsymbol{U}' = \sum_{i=1}^n \frac{1}{\sqrt{n}}\kappa_i^t\phi_i = \pi + \sum_{i=2}^n \frac{1}{\sqrt{n}}\kappa_i^t\phi_i.$$

We get the following for the left hand side of (3.14)

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 = \|\sum_{i=2}^n \frac{1}{\sqrt{n}}\kappa_i^t\phi_i\|_2^2 = \sum_{i=2}^n \kappa_i^{2t}. \tag{3.15}$$

Next step is to split the last sum in two:

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 = \sum_{i \geq 2: \kappa_i \leq 0} \kappa_i^{2t} + \sum_{i \geq 2: \kappa_i > 0} \kappa_i^{2t}.$$

We can bound the first sum of the right hand side by $n\kappa_n^{2t}$. Since by assumption $\mathcal{E}^b \leq A\mathcal{E}$, the extremal characterization of the eigenvalues, (3.13), yields that $\lambda_i \geq \lambda_i^b/A$. Hence

$$\kappa_i = 1 - \lambda_i \leq 1 - \frac{\lambda_i^b}{A} \leq e^{-\lambda_i^b/A},$$

where the last inequality follows from $1 - x \leq e^{-x}$, $\forall x$. Consequently,

$$\begin{aligned}
\|\mathbb{P}(X_t \in \cdot) - \pi\|_2^2 &\leq n\kappa_n^{2t} + \sum_{i\geq 2:\kappa_i>0} \kappa_i^{2t} \\
&\leq n\kappa_n^{2t} + \sum_{i\geq 2:\kappa_i>0} e^{-2\lambda_i^b t/A} \\
&\leq n\kappa_n^{2t} + \sum_{i\geq 2:\lambda_i^b\geq 1/2} e^{-2\lambda_i^b t/A} + \sum_{i\geq 2:\lambda_i^b<1/2} e^{-2\lambda_i^b/A} \\
&\leq n\kappa_n^{2t} + ne^{-t/A} + \sum_{i\geq 2:\lambda_i^b<1/2} (\kappa_i^b)^{t/A}.
\end{aligned}$$

The last inequality is subject to the fact that $e^{-2x} \leq 1-x$, for $x \leq 1/2$. Finally, observe that

$$\sum_{i\geq 2:\lambda_i^b<1/2} (\kappa_i^b)^{t/A} \leq \sum_{i=2}^{n} (\kappa_i^b)^{2\lfloor t/2A\rfloor} = \|\mathbb{P}(X^b_{\lfloor t/2A\rfloor} \in \cdot) - \pi\|_2^2.$$

The last equality in analogy with (3.15).

□

Before the next lemma we have to introduce some notation. Let E be a symmetric set of generators contained in the support of μ. For each $y \in S$, choose a representation $y = x_1x_2\cdots x_k$, where $x_j \in E$, $j = 1, 2, \ldots, k$. Write $|y| := k$. Further, denote by $N(x, y)$ the number of times that x appears in the chosen representation of y.

Lemma 3. *Let μ^b and μ be symmetric measures on S. Moreover, let E be a symmetric set of generators contained in the support of μ. If*

$$A = \max_{x\in E} \frac{1}{\mu(x)} \sum_{y\in S} |y| N(x, y)\mu^b(y).$$

Then $\mathcal{E}^b \leq A\mathcal{E}$.

Proof. Let φ be an arbitrary function on S, and $y, z \in S$. Choose the representation $y = x_1x_2\cdots x_{|y|}$ of y, where $x_j \in E$, $j = 1, 2, \ldots |y|$. Then we can write $\varphi(z) - \varphi(zy)$ as the following telescoping sum

$$\begin{aligned}
\varphi(z) - \varphi(zy) = (\varphi(z) - \varphi(zx_1)) + (\varphi(zx_1) - \varphi(zx_1x_2)) + (\varphi(zx_1x_2) - \varphi(zx_1x_2x_3)) \\
+ \ldots + (\varphi(zx_1\ldots x_{|y|-1}) - \varphi(zx_1\ldots x_{|y|})).
\end{aligned}$$

Since by Cauchy-Schwarz inequality $\left(\sum_{i=1}^{|y|} a_i\right)^2 \leq |y| \sum_{i=1}^{|y|} a_i^2$, squaring the expression above gives

$$(\varphi(z)-\varphi(zy))^2 \le |y|\sum_{i=1}^{|y|}(\varphi(zx_1\ldots x_{i-1})-\varphi(zx_1\ldots x_i))^2.$$

Note that the terms in the right sum are on the form $(\varphi(z')-\varphi(z'x_i))^2$, where $z' \in S$. Summing both sides over all generators in S, each term appears at most $N(x_i, y)$ times on the right hand side. Thus

$$\begin{aligned}\sum_{z\in S}(\varphi(z)-\varphi(zy))^2 &\le |y|\sum_{z\in S}\sum_{i=1}^{|y|}(\varphi(zx_1\ldots x_{i-1})-\varphi(zx_1\ldots x_i))^2\\ &\le |y|\sum_{z\in S}\sum_{i=1}^{|y|}(\varphi(z)-\varphi(zx_i))^2 N(x_i,y)\\ &\le |y|\sum_{z\in S}\sum_{x\in E}(\varphi(z)-\varphi(zx))^2 N(x,y). \qquad (3.16)\end{aligned}$$

Multiply the left hand side of (3.16) with $\mu^b(y)$ and sum over all $y \in S$. This gives

$$\sum_{y\in S}\sum_{z\in S}\mu^b(y)(\varphi(z)-\varphi(zy))^2 = \mathcal{E}^b(\varphi,\varphi).$$

The same operation on the right hand side of (3.16) yields

$$\begin{aligned}&\sum_{z\in S}\sum_{x\in E}(\varphi(z)-\varphi(zx))^2\sum_{y\in S}|y|N(x,y)\mu^b(y) \le\\ &\sum_{z\in S}\sum_{x\in E}(\varphi(z)-\varphi(zx))^2\mu(x)A \underset{E\subseteq S}{\le}\\ &\mathcal{E}(\varphi,\varphi)A.\end{aligned}$$

Consequently $\mathcal{E}^b(\varphi,\varphi) \le A\mathcal{E}(\varphi,\varphi)$ for any function φ on S.

□

The following lemma will also be useful, when bounding the term $n\kappa_n^{2t}$ in Lemma 2.

Lemma 4. *For a discrete time random walk on S generated by μ, we have the following inequality for the smallest eigenvalue of the transition matrix*

$$\kappa_{|S|} \ge 2\mu(id) - 1. \qquad (3.17)$$

Proof. The result is true if $\mu(\mathrm{id}) = 0$. If $\mu(\mathrm{id}) > 0$, let $\mu' = \frac{\mu(x)-\mu(\mathrm{id})\delta_{\mathrm{id}}}{1-\mu(\mathrm{id})}$, where δ is the Dirac delta function. Some algebra shows that the transition matrix of the random walk generated by μ' has eigenvalues

$$\kappa_i' = \frac{\kappa_i - \mu(\mathrm{id})}{1 - \mu(\mathrm{id})}.$$

Since $\kappa_{|S|}' \geq -1$, (3.17) follows from rearranging the expression above with $i = |S|$.

□

3.3 Upper bound on the mixing time on lollipop graphs

Lemma 5. *For random neighbour transpositions on the $\left(\left(1 - \frac{1}{a}\right) n, \frac{n}{a}\right)$-lollipop graph, $a > 1$, we have*

$$\tau_{mix} \leq (1 + o(1)) \frac{(a-1)^2(2a-1)^2}{2a^6} n^4 \log n.$$

Proof. We will use Lemma 2 and 3. Let μ and μ^b be the measures that generate the random neighbour transposition process on the lollipop graph $L_{n,a} = (V_{L_{n,a}}, E_{L_{n,a}})$ and the complete graph $G_n = (V_{G_n}, E_{G_n})$ respectively. Further, let $\{X_t\}$ and $\{X_t^b\}$ be the random walks on the symmetric group generated by these measures. The benchmark shuffle, ordinary random transpositions, corresponding to shuffling on the complete graph G_n, is defined by $\mu^b(i\ \ j) = 2/n^2$, for $i, j = 1, 2, \ldots, n$, $i \neq j$ and $\mu^b(\mathrm{id}) = 1/n$. For this shuffle, we know that $\|\mathbb{P}(X^b_{\lfloor (1/2 + C_b/\log n) n \log n \rfloor} \in \cdot) - \pi\|_2^2 \leq 1/4$, for some large enough constant C_b. This was proved by Diaconis and Shahshahani [8], see also [17]. First we will derive an upper bound on A in Lemma 3, and then use this bound on A to find $t = t(n)$ such that the right hand side of (3.14) converges to something less than $1/4$. This t will by (2.3) be an upper bound on the mixing time τ_{mix} for $\{X_t\}$.

We use the notation $\mathfrak{E}(G_n) = \{(i\ \ j) | (i, j) \in E_{G_n}\}$ and $\mathfrak{E}(L_{n,a}) = \{(i\ \ j) | (i, j) \in E_{L_{n,a}}\}$. We know that $\mu(x) = \frac{1}{m}$ for all $x \in \mathfrak{E}(L_{n,a})$, and $\mu^b(y) = \frac{2}{n^2}$ for all $y \in G_n$. Thus

$$A = \frac{2m}{n^2} \max_{x \in \mathfrak{E}(L_{n,a})} \sum_{y \in \mathfrak{E}(G_n)} |y| N(x, y).$$

This leaves us with the task to maximize

$$g(x) := \sum_{y \in \mathfrak{E}(G_n)} |y| N(x, y)$$

over all $x \in \mathfrak{E}(L_{n,a})$.

For $x \in \mathfrak{E}(L_{n,a})$ we can choose x itself as representation. Look at a vertex i in the linear part and vertex j in the complete part of the graph. Choose this representation of $(i\ \ j)$:

$$y = (i\ \ j) = (i\ \ i+1)(i+1\ \ i+2) \cdots (\frac{n}{a} - 1\ \ \frac{n}{a})(\frac{n}{a}\ \ j)(\frac{n}{a} - 1\ \ \frac{n}{a}) \cdots (i\ \ i+1).$$

Consider the transposition $x = (\frac{n}{a}\ \ j)$ in the clique. We have $N(x, y) = 1$ and $|y| = 2\left(\frac{n}{a} - i\right) + 1$.

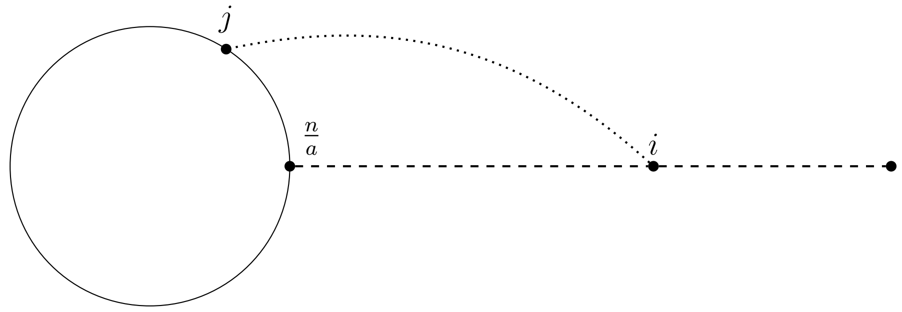

Figure 3.3: The representation of $(i \quad j)$ involves $x = (\frac{n}{a} \quad j)$.

Thus, summing over all vertices i in the linear part (including $i = \frac{n}{a}$, where $|y| = 1$ and $N(x,y) = 1$), we get for any transposition $x = (\frac{n}{a} \quad j)$ in the clique

$$g(x) = \sum_{y \in \mathfrak{E}(G_n)} |y| N(x,y) = \sum_{i=1}^{\frac{n}{a}} \left(2\left(\frac{n}{a} - i\right) + 1\right) \cdot 1 = \frac{n^2}{a^2}. \tag{3.18}$$

Furthermore, for vertices i and j both in the linear part, i.e. $i < j \leq \frac{n}{a}$, choose the following representation for the transposition $(i \quad j)$

$$y = (i \quad j) = (i \quad i+1)(i+1 \quad i+2) \cdots (j-1 \quad j)(j-2 \quad j-1) \cdots (i \quad i+1),$$

so that $|y| = 2(j-i) - 1$.

Suppose $b \in [a, n]$ and that $\frac{n}{b}$ is an integer. Let x be the transposition $\left(\frac{n}{b} \quad \frac{n}{b} + 1\right)$. Then x occurs in three types of representations, see Figure 3.4 below.

Type I is representations of $y = (i \quad j)$, where vertex i is in the linear part to the right of vertex $\frac{n}{b}$, and vertex j is in the clique. For each i there are $\left(1 - \frac{1}{a}\right) n$ such representations. Moreover, with to our choice of representations above, $|y| = 2\left(\frac{n}{a} - i\right) + 1$ and $N(x,y) = 2$.

Type II is representations of $y = (i \quad k)$, where vertex k is in the linear part, but to the left of vertex $\frac{n}{b} + 1$. Here $|y| = 2(k-i) - 1$ and $N(x,y) = 2$.

Type III is representations of $y = \left(i \quad \frac{n}{b} + 1\right)$. Here $|y| = 2\left(\frac{n}{b} + 1 - i\right) - 1$ and $N(x,y) = 1$.

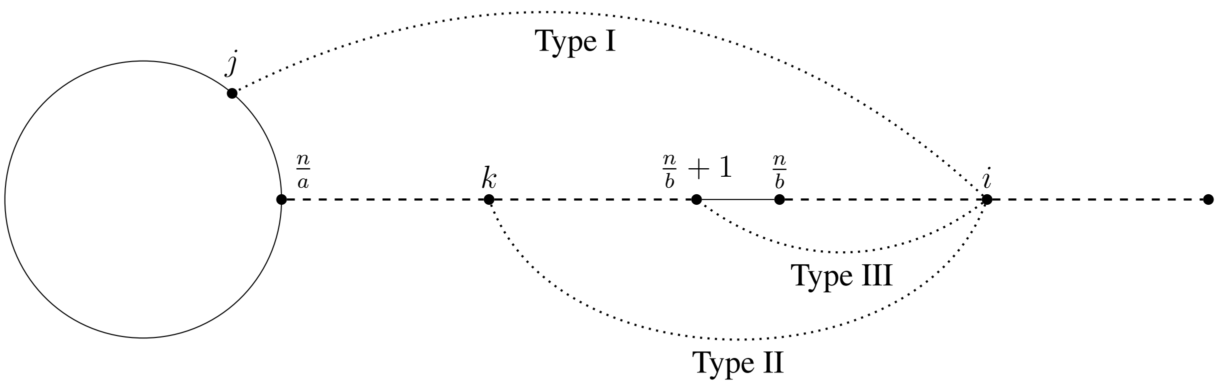

Figure 3.4: The transposition $x = \left(\frac{n}{b} \quad \frac{n}{b}+1\right)$ occurs in the representations of $(i \quad j)$, $(i \quad k)$ and $\left(i \quad \frac{n}{b}+1\right)$.

Summing over all vertices i to the right of vertex $\frac{n}{b}$ we get for $x = \left(\frac{n}{b} \quad \frac{n}{b}+1\right)$

$$
\begin{aligned}
&g(x) = \sum_{y\in\mathfrak{E}(G_n)} |y|N(x,y) = \\
&\sum_{i=1}^{\frac{n}{b}} \Bigg(\left(1-\frac{1}{a}\right) n \cdot \left(2\left(\frac{n}{a}-i\right)+1\right)\cdot 2 + \\
&\qquad + \sum_{k=\frac{n}{b}+2}^{\frac{n}{a}} \left((2(k-i)-1)\cdot 2\right) + \left(2\left(\frac{n}{b}+1-i\right)-1\right)\cdot 1 \Bigg) = \\
&\sum_{i=1}^{\frac{n}{b}} \left(4\left(1-\frac{1}{a}\right) n\left(\frac{n}{a}-i\right) + \left(\frac{2n^2}{a^2} - \frac{4n}{a}i - \frac{2n^2}{b^2} + \frac{4n}{b}i\right) + O(n)\right) = \\
&\left(\frac{4}{ab} - \frac{2}{a^2 b} - \frac{2}{b^2}\right) n^3 + O(n^2),
\end{aligned} \tag{3.19}
$$

where in the first equality the first term is for the representation of transpositions of Type I. The second sum is for representations of Type II, and the last term is for the representation of Type III. Furthermore, $\frac{4}{ab} - \frac{2}{a^2b} - \frac{2}{b^2}$ reaches its maximum, $\frac{(2a-1)^2}{2a^4}$, when $\frac{1}{b} = \frac{1}{a} - \frac{1}{2a^2}$. From (3.18) and (3.19) we can conclude that $g(x) = \sum_{y\in\mathfrak{E}(G_n)} |y|N(x,y)$ takes its maximum for $x = x_a = \left(n\left\lfloor\frac{1}{a} - \frac{1}{2a^2}\right\rfloor \quad n\left\lfloor\frac{1}{a} - \frac{1}{2a^2}\right\rfloor + 1\right)$. Note that for a close to 1, yielding a lollipop with just a small clique, x_a will be about half way into the linear part, similar to the path-graph. Whereas for large a corresponding to a lollipop with a short "stick", x_a will correspond to one of the edges closest to the complete part, just as our intuition might tell us. From (3.19) we obtain the following bound on the maximum of $g(x)$;

$$\max_{x\in\mathfrak{E}(L_{n,a})} g(x) = \max_{x} \sum_{y\in\mathfrak{E}(G_n)} |y|N(x,y) \leq (1+O(n^{-1}))\frac{(2a-1)^2}{2a^4}n^3.$$

Altogether, using that $m = (1+O(n^{-1}))\frac{1}{2}\left(1-\frac{1}{a}\right)^2 n^2$, we have

$$\begin{aligned} A &\leq (1+O(n^{-1}))\frac{2m}{n^2}\frac{(2a-1)^2}{2a^4}n^3 \\ &= (1+O(n^{-1}))\frac{(a-1)^2(2a-1)^2}{2a^6}n^3. \end{aligned}$$

According to Lemma 3 we have that

$$\mathcal{E}^b \leq (1+O(n^{-1}))\frac{(a-1)^2(2a-1)^2}{2a^6}n^3\mathcal{E},$$

where $\mathcal{E}^b$ and $\mathcal{E}$ are the Dirichlet forms associated with the measures generating the neighbour transposition shuffle on the complete graph G_n and the lollipop graph $L_{n,a}$ respectively.

Suppose κ_n is the smallest eigenvalue of the measure μ. Then Lemma 4 yields $|\kappa_n| \leq 1-\frac{2}{m} = 1-\Theta(n^{-2})$. Inserting $t = \left\lceil\left(1+\frac{4C_b}{\log n}\right)\frac{(a-1)^2(2a-1)^2}{2a^6}n^4\log n\right\rceil$ into (3.14) we get

$$\begin{aligned} &\|\mathbb{P}(X_{\left\lceil\left(1+\frac{4C_b}{\log n}\right)\frac{(a-1)^2(2a-1)^2}{2a^6}n^4\log n\right\rceil} \in \cdot) - \pi\|_2^2 \leq \\ &\underbrace{n\left(1-\Theta(n^{-2})\right)^{\Theta(n^4\log n)}}_{\to 0} + \underbrace{ne^{-\Theta(n\log n)}}_{\to 0} + \underbrace{\|\mathbb{P}(X^b_{\left\lfloor\left(\frac{1}{2}+\frac{2C_b+o(1)}{\log n}\right)n\log n\right\rfloor} \in \cdot) - \pi\|_2^2}_{<1/4 \text{ as } n\to\infty}. \end{aligned}$$

From (2.3) we see that for random neighbour transpositions on the $\left(\left(1-\frac{1}{a}\right)n, \frac{n}{a}\right)$-lollipop graph $t = (1+o(1))\frac{(a-1)^2(2a-1)^2}{2a^6}n^4\log n \geq \hat{\tau} \geq \tau_{\text{mix}}$, thus we have found an lower upper bound on the mixing time τ_{mix}.

□

We have proved Lemma 1 and Lemma 5, and consequently Theorem 2. Note that the ratio between the upper and lower bound, $\frac{2x_0(2a-1)^2}{a^2}$, will always be in the range $(2\pi, 4\pi)$, approaching 2π as a approaches 1 and close to 4π for large a.

Chapter 4

Neighbour transpositions on random graphs

Consider a realization $G_{n,p} = (V_{G_{n,p}}, E_{G_{n,p}})$ of $\mathcal{G}(n,p)$, Gilbert's random graph model on n vertices, where each pair of vertices are connected by an *open edge* with probability p independently. If there is no edge present, we say that the edge is *closed*.

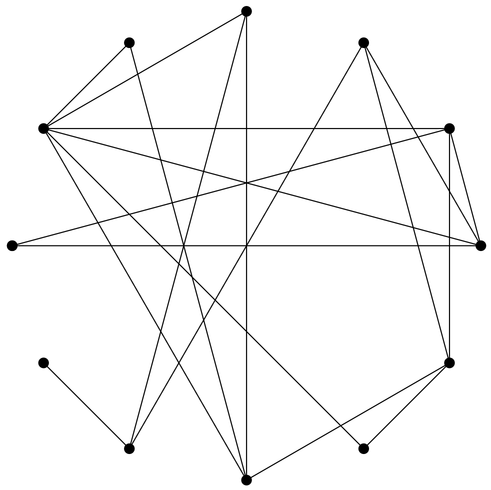

Figure 4.1: Realization of Gilbert's random graph model with $n = 12$ vertices, and edge probability $p = 0.3$.

Write $m := |E_{G_{n,p}}|$, for the total number of open edges, and d_i for the *degree* of vertex i, i.e. the number of incident edges. The set of generators of the neighbour transposition shuffle on $G_{n,p}$ is denoted by $\mathfrak{E}(G_{n,p}) = \{(i\ \ j) | (i,j) \in E_{G_{n,p}} \text{ or } i = j\}$. The updating measure for the card shuffling chain is given by $\mu(x) = \frac{1}{m+n}$, $x \in \mathfrak{E}(G_{n,p})$. Note that the transpositions $(i\ \ i)$, $i = 1, 2, \ldots, n$, are allowed, yielding $\mu(\mathrm{id}) = \frac{n}{m+n}$. Let $\{X_t^n\}_{t=0}^{\infty}$ be the Markov chain on S_n induced by this random process. Note that $m \sim \mathrm{Bin}(\binom{n}{2}, p)$ and $d_i \sim \mathrm{Bin}(n-1, p)$.

Since we are dealing with irreducible Markov chains, the graph has to be connected. Erdős and Rényi [11] proved that edge probability $\frac{\log n}{n}$ is a sharp threshold for a.a.s. connectedness of

$G_{n,p}$, thus we will restrict ourselves to the case $p > \frac{\log n}{n}$.

In section 4.1 and section 4.2 we will prove the following theorem on random graphs.

Theorem 3. *For random neighbour transpositions on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$, the mixing time τ_{mix} asymptotically almost surely has the following lower bound for the given ranges of $p < 1$.*

For $p = \omega\left(\frac{\log n}{n}\right)$,

$$\tau_{mix} \geq \frac{1 - o(1)}{2} n \log n.$$

For $p = \frac{c \log n}{n}$, $c > 1$,

$$\tau_{mix} \geq \frac{1 - o(1)}{2\left(1 + \sqrt{\frac{2}{c}}\right)} n \log n.$$

Moreover, the mixing time a.a.s. has the following upper bound for p such that $1 > p > n^{\delta - 1}$, for some $\delta > 0$,

$$\tau_{mix} \leq C n \log n,$$

for some large enough constant C.

4.1 Lower bound on the mixing time on random graphs

We will need Chernoff's inequality for binomial random variables.

Lemma 6 (Chernoff's inequality)**.** *Let Y be a binomially distributed random variable with parameters n and p, and let $\epsilon > 0$ be a constant which may depend on n or p. Then it holds that*

$$\mathbb{P}\left(\frac{Y}{n} \geq p + \epsilon\right) \leq \exp\left(-n\psi_p(p+\epsilon)\right), \tag{4.1}$$

$$\mathbb{P}\left(\frac{Y}{n} \leq p - \epsilon\right) \leq \exp\left(-n\psi_p(p-\epsilon)\right), \tag{4.2}$$

where

$$\psi_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}.$$

Note that, if $\frac{\epsilon}{p} \to 0$ and $\frac{\epsilon}{1-p} \to 0$, we get using the power series for $\log(\cdot)$

$$\psi_p(p+\epsilon) = \epsilon^2 \left(\frac{1}{2p} + \frac{1}{2(1-p)} \right) (1+o(1)), \tag{4.3}$$

$$\psi_p(p-\epsilon) = \epsilon^2 \left(\frac{1}{2p} + \frac{1}{2(1-p)} \right) (1+o(1)). \tag{4.4}$$

The classical way to calculate lower bounds on the mixing time τ_{mix}, for the Markov chains $\{X_t^n\}_{t=0}^{\infty}$ with stationary distributions π^n is to find events $A = A^n$ and times $t = t(n)$, such that $\pi^n(A^n) \to 0$ and $\mathbb{P}(X_{t(n)}^n \in A^n) \to 1$ as $n \to \infty$. Then, from the definition of mixing time, (2.2), we know that $\tau_{\text{mix}} \geq t(n)$. We will divide the range of p for which $G_{n,p}$ a.a.s. is connected, i.e. $p > \frac{\log n}{n}$, into two cases. We treat $p = \omega\left(\frac{\log n}{n}\right)$ and $p = c\frac{\log n}{n}$, $c > 1$, separately.

Lemma 7. *Suppose $1 > p = \omega\left(\frac{\log n}{n}\right)$. Then, for random neighbour transpositions on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$, we have the following a.a.s. lower bound on the mixing time.*

$$\tau_{\textit{mix}} \overset{\textit{a.a.s.}}{\geq} \frac{1-o(1)}{2} n \log n$$

Proof. Consider the event, $A = A^n$, that at least $\log n$ cards are in their starting positions. Note that the expected number of such cards at stationarity is 1. Markov's inequality thus yields that

$$\pi^n(A^n) \leq \frac{1}{\log n} \to 0.$$

We want to find $t(n)$, as large as possible, such that $\mathbb{P}(X_{t(n)}^n \in A^n) \to 1$. Use the following simple checking procedure. A card is checked when it is transposed with another card or itself for the first time. Let $T_0 = 0$ and let T_i be the time when the ith cards is checked. Note that $T_{\lfloor n-\log n \rfloor} \geq t$ implies $X_t^n \in A^n$. Thus, if we can find $t = t(n)$ such that $\mathbb{P}(T_{\lfloor n-\log n \rfloor} \geq t(n)) \to 1$, then $\tau_{\text{mix}} \overset{\text{a.a.s.}}{\geq} t(n)$. When i cards are checked, let D_i be the sum of the degrees of the vertices with yet unchecked cards. Then, no matter what time this has taken, the probability that another card is checked is at most $\frac{D_i+n-i}{m+n}$, and at least $\frac{D_i/2+n-i}{m+n}$, it depends on whether the checked cards are neighbours or not. The term $n-i$ comes from the possibility of transposing a card with itself. Since at most two cards are checked at each shuffle the time between check $i-2$ and i, $T_i - T_{i-2}$, stochastically dominates $V_i \sim \text{Geo}\left(p_i = \frac{D_{i-2}+n-i+2}{m+n}\right)$. Choose the V_i's sequentially so that they are independent of each other, and such that for all $j \leq \lfloor n/2 \rfloor$

$$\sum_{i=1}^{j} T_{2i} - T_{2i-2} \succeq \sum_{i=1}^{j} V_{2i}.$$

Then we have

$$T_{\lfloor n-\log n \rfloor} \geq \sum_{i=1}^{\left\lfloor \frac{n-\log n}{2} \right\rfloor} T_{2i} - T_{2i-2} \succeq \underbrace{\sum_{i=1}^{\left\lfloor \frac{n-\log n}{2} \right\rfloor} V_{2i}}_{=:V}.$$

To bound the sum V, we first derive an upper bound on the p_i's. Put $d_{\max} = \max_{j\in[1,n]} d_j$, then

$$\begin{aligned}
&\mathbb{P}\left(d_{\max} \geq np\left(1+2\sqrt{\frac{\log n}{np}}\right)\right) = \\
&\mathbb{P}\left(\bigcup_{j=1}^{n}\left\{d_j \geq np\left(1+2\sqrt{\frac{\log n}{np}}\right)\right\}\right) \leq \\
&\sum_{j=1}^{n}\mathbb{P}\left(\frac{d_j}{n-1} \geq p+2p\sqrt{\frac{\log n}{np}}(1+o(1))\right) \underset{(4.1),\,(4.3)}{\leq} \\
&n\exp\left(-(n-1)4p^2\frac{\log n}{np}\left(\frac{1}{2p}+\frac{1}{2(1-p)}\right)(1+o(1))\right) \leq \\
&n\exp\left(-(2+o(1))\log n\right) \xrightarrow[n\to\infty]{} 0.
\end{aligned} \tag{4.5}$$

Using Chebyshev's inequality, and the fact that $\mathbf{Var}(m) = \binom{n}{2}p(1-p)$, we can also bound m from below.

$$\mathbb{P}\left(|m-\binom{n}{2}p| \leq \sqrt{p}n\log\log n\right) \overset{\text{Chebyshev}}{\leq} \frac{\binom{n}{2}p(1-p)}{pn^2\log\log^2 n} \xrightarrow[n\to\infty]{} 0 \tag{4.6}$$

Thus $d_{\max} \overset{\text{a.a.s.}}{\leq} np\left(1+2\sqrt{\frac{\log n}{np}}\right)$ and $m \overset{\text{a.a.s.}}{\geq} \binom{n}{2}p - \sqrt{p}n\log\log n$. Now, since $D_{i-2} \leq (n-i+2)d_{\max}$ we get, uniformly for all i,

$$\begin{aligned}
p_i &\leq \frac{(n-i+2)(d_{\max}+1)}{m+n} \\
&\overset{\text{a.a.s.}}{\leq} \frac{(n-i+2)\left(np+2np\sqrt{\frac{\log n}{np}}+1\right)}{\binom{n}{2}p-\sqrt{p}n\log\log n} \\
&= \frac{(n-i+2)np\left(1+2\sqrt{\frac{\log n}{np}}(1+o(1))\right)}{\frac{n^2p}{2}\left(1-\frac{\log\log n}{\sqrt{n^2p}}(1+o(1))\right)} \\
&= \frac{2(n-i+2)}{n}\left(1+2\sqrt{\frac{\log n}{np}}(1+o(1))\right).
\end{aligned}$$

Since $V_{2i} \sim \text{Geo}\left(p_{2i}\right)$ we get the following lower bound on $\mathbb{E}[V]$.

$$
\begin{aligned}
\mathbb{E}[V] &= \mathbb{E}\left[\sum_{i=1}^{\lfloor\frac{n-\log n}{2}\rfloor} V_{2i}\right] \\
&= \sum_{i=1}^{\lfloor\frac{n-\log n}{2}\rfloor} \frac{1}{p_{2i}} \\
&\geq \left(1+2\sqrt{\frac{\log n}{np}}(1+o(1))\right)^{-1} \cdot \sum_{i=1}^{\lfloor\frac{n-\log n}{2}\rfloor} \frac{n}{2(n-2i+2)} \\
&\geq \left(1-2\sqrt{\frac{\log n}{np}}(1+o(1))\right)\frac{n}{2}\Bigg(\log n-(1+o(1))\log\log n\Bigg) \\
&= \frac{n\log n}{2}\left(1-\underbrace{\left(2\sqrt{\frac{\log n}{np}}+\frac{\log\log n}{\log n}\right)(1+o(1))}_{:=h_1(n)}\right) \\
&= \frac{n\log n}{2}(1-h_1(n))
\end{aligned}
$$

Next step is to bound the variance of V. The independence of the V_{2i}'s yields

$$
\begin{aligned}
\mathbf{Var}(V) &= \mathbf{Var}\left(\sum_{i=1}^{\lceil\frac{n-\log n}{2}\rceil} V_{2i}\right) = \sum_{i=1}^{\lceil\frac{n-\log n}{2}\rceil} \mathbf{Var}(V_{2i}) \\
&\leq \sum_{i=1}^{\lfloor\frac{n-\log n}{2}\rfloor} \frac{1}{(p_{2i})^2}.
\end{aligned}
$$

Put $d_{\min} = \min_{j\in[1,n]} d_j$. Analogously to (4.5) we can prove that $d_{\min} \overset{\text{a.a.s.}}{\geq} np\left(1-2\sqrt{\frac{\log n}{np}}\right) = np(1-o(1))$ and from (4.6) we have that $m \overset{\text{a.a.s.}}{\leq} \binom{n}{2}p+\sqrt{p}n\log\log n = \binom{n}{2}p(1+o(1))$. Consequently, uniformly for all i we have

$$
\begin{aligned}
p_i &\geq \frac{(n-i+2)(d_{\min}+1)}{m+n} \\
&\overset{\text{a.a.s.}}{\geq} \frac{(n-i+2)np}{\binom{n}{2}p}(1-o(1)) \\
&= \frac{2(n-i+2)}{n}(1-o(1)).
\end{aligned}
$$

Thus, since $\sum_i \frac{1}{i^2}$ converges

$$\mathbf{Var}(V) \leq \sum_{i=1}^{\lceil \frac{n-\log n}{2} \rceil} \frac{1}{(p_{2i})^2} \leq \sum_{i=1}^{\lceil \frac{n-\log n}{2} \rceil} \frac{n^2}{(n-2i)^2}(1+o(1)) = B_0 n^2,$$

for some constant B_0. To sum up we have $\mathbb{E}[V] \geq \frac{1-h_1(n)}{2} n \log n$ and $\mathbf{Var}(V) \leq B_0 n^2$. Using that $h_1(n) \to 0$, but $h_1(n) \log n \to \infty$, we get

$$\begin{aligned}
&\mathbb{P}\left(V < \frac{1-\sqrt{h_1(n)}}{2} n \log n\right) \leq \\
&\mathbb{P}\left(V - \mathbb{E}[V] \leq \frac{1-\sqrt{h_1(n)}}{2} n \log n - \frac{1-h_1(n)}{2} n \log n\right) \leq \\
&\mathbb{P}\left(|V - \mathbb{E}[V]| \geq (1-o(1))\frac{\sqrt{h_1(n)}}{2} n \log n\right) \underset{\text{Chebyshev}}{\leq} \\
&\frac{B_0 n^2}{\left((1-o(1))\frac{\sqrt{h_1(n)}}{2} n \log n\right)^2} = \frac{4B_0(1+o(1))}{h_1(n)\log^2 n} \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Since $T_{\lfloor n-\log n \rfloor} \succeq V$, the above expression yields that $\mathbb{P}\left(T_{\lfloor n-\log n \rfloor} \geq \frac{1-\sqrt{h_1(n)}}{2} n \log n\right) \to 1$. Thus, we have an a.a.s. lower bound on $T_{\lfloor n-\log n \rfloor}$, which in turn also bounds τ_{mix} from below. Hence

$$\tau_{\text{mix}} \overset{\text{a.a.s.}}{\geq} \frac{1-\sqrt{h_1(n)}}{2} n \log n = \frac{1-o(1)}{2} n \log n.$$

□

Next we deal with the case $p = c\frac{\log n}{n}$. Note that for $c > 1$ the random graphs $G_{n,p}$ is a.a.s. connected.

Lemma 8. *Suppose $p = c\frac{\log n}{n}$, for some constant $c > 1$. Then, for random neighbour transpositions on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$, we have the following a.a.s. lower bound on the mixing time.*

$$\tau_{mix} \overset{a.a.s.}{\geq} \frac{1-o(1)}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n.$$

Proof. Again we consider the event A^n, that at least $\log n$ cards are in their starting positions. Remember that $\pi^n(A^n) \xrightarrow[n\to\infty]{} 0$. We use the same notation as in the proof of Lemma 7, and will establish an asymptotically almost sure lower bound on $T_{\lfloor n-\log n\rfloor}$, and at the same time bound the mixing time from below. For this range of p we have

$$\begin{aligned}
&\mathbb{P}\left(d_{\max} \geq np\left(1+\sqrt{\frac{2}{c}}+\frac{1}{\log\log n}\right)\right) = \\
&\mathbb{P}\left(\bigcup_{j=1}^{n}\left\{d_j \geq np\left(1+\sqrt{\frac{2}{c}}+\frac{1}{\log\log n}\right)\right\}\right) \leq \\
&\sum_{j=1}^{n}\mathbb{P}\left(\frac{d_j}{n-1} \geq p+p\left(\sqrt{\frac{2}{c}}+\frac{1}{\log\log n}\right)(1+o(1))\right) \underset{(4.1),\,(4.3)}{\leq} \\
&n\exp\left(-(n-1)p^2\left(\sqrt{\frac{2}{c}}+\frac{1}{\log\log n}\right)^2\left(\frac{1}{2p}+\frac{1}{2(1-p)}\right)(1+o(1))\right) \underset{np=c\log n}{=} \\
&n\exp\left(-\frac{c\log n}{2}\left(\frac{2}{c}+\frac{2}{\log\log n}\sqrt{\frac{2}{c}}\right)(1+o(1))\right) = \\
&\exp\left(-\sqrt{2c}\frac{\log n}{\log\log n}(1+o(1))\right) \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Thus $d_{\max} \overset{\text{a.a.s.}}{\leq} np\left(1+\sqrt{\frac{2}{c}}+\frac{1}{\log\log n}\right)$ and since (4.6) is still valid, $m \overset{\text{a.a.s.}}{\geq} \binom{n}{2}p-\sqrt{p}n\log\log n$. Mimicking the proof of Lemma 7 we get, with the same notation as before

$$p_i \overset{\text{a.a.s.}}{\leq} \frac{2(n-i+2)}{n}\left(1+\sqrt{\frac{2}{c}}+\frac{1+o(1)}{\log\log n}\right),$$

uniformly for all i. Since $V_{2i} \sim \text{Geo}\,(p_{2i})$ we get the following lower bound on $\mathbb{E}[V]$.

$$
\begin{aligned}
\mathbb{E}[V] &= \mathbb{E}\left[\sum_{i=1}^{\lfloor \frac{n-\log n}{2} \rfloor} V_{2i}\right] \\
&\geq \sum_{i=1}^{\lfloor \frac{n-\log n}{2} \rfloor} \frac{1}{p_{2i}} \\
&\geq \frac{n}{2\left(1+\sqrt{\frac{2}{c}}+\frac{1+o(1)}{\log\log n}\right)} \sum_{i=1}^{\lfloor \frac{n-\log n}{2} \rfloor} \frac{1}{(n-2i+2)} \\
&\geq \frac{n}{2\left(1+\sqrt{\frac{2}{c}}\right)} \left(1-\frac{1+o(1)}{2\left(1+\sqrt{\frac{2}{c}}\right)\log\log n}\right)\Big(\log n-(1+o(1))\log\log n\Big) \\
&= \frac{n\log n}{2\left(1+\sqrt{\frac{2}{c}}\right)} \left(1-\underbrace{\frac{1+o(1)}{2\left(1+\sqrt{\frac{2}{c}}\right)\log\log n}}_{:=h_2(n)}.\right) = \frac{n\log n}{2\left(1+\sqrt{\frac{2}{c}}\right)}\left(1-h_2(n)\right),
\end{aligned}
$$

And in analogy with the previous lemma

$$
\begin{aligned}
\mathbf{Var}(V) &= \mathbf{Var}\left(\sum_{i=1}^{\lceil \frac{n-\log n}{2} \rceil} V_{2i}\right) \leq \sum_{i=1}^{\lfloor \frac{n-\log n}{2} \rfloor} \frac{1}{(p_{2i})^2} \\
&\leq B_1 n^2,
\end{aligned}
$$

for some constant B_1. Thus $\mathbb{E}[V] \geq \frac{1-h_2(n)}{2\left(1+\sqrt{\frac{2}{c}}\right)} n\log n$ and $\mathbf{Var}(V) \leq B_1 n^2$. Finally, Chebyshev's inequality, together with the fact that $h_2(n) \to 0$ and $\sqrt{h_2(n)}\log n \to \infty$, implies

$$\mathbb{P}\left(V < \frac{1-\sqrt{h_2(n)}}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n\right) \leq$$
$$\mathbb{P}\left(|V - \mathbb{E}[V]| \geq (1-o(1)) \frac{\sqrt{h_2(n)}}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n\right) \leq$$
$$\frac{B_1 n^2}{\left((1-o(1))\frac{\sqrt{h_2(n)}}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n\right)^2} \xrightarrow[n\to\infty]{} 0.$$

This entails $\tau_{\text{mix}} \overset{\text{a.a.s.}}{\geq} \frac{1-\sqrt{h_2(n)}}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n = \frac{1-o(1)}{2\left(1+\sqrt{\frac{2}{c}}\right)} n \log n.$

□

4.2 Upper bound on the mixing time on random graphs

In this section we will derive an upper bound of order $n \log n$ on the mixing time for neighbour transpositions on random graphs with bounded diameter. Note that $G_{n,p}$ a.a.s. has bounded diameter if $p > n^{\delta-1}$ for some $\delta > 0$. Besides, the diameter will then be bounded by $1 + 1/\delta$, see for example [5].

We will reach the upper bound via comparison with ordinary random transpositions, using Lemma 2. The reasoning will also rely upon a paper by Broder et al. about edge-disjoint paths in random graphs, [4].

Lemma 9. *Suppose we have p such that $1 > p > n^{\delta-1}$ for some $\delta > 0$. Then the mixing time of the random neighbour transposition shuffle on a realization $G_{n,p}$ of $\mathcal{G}(n,p)$ a.a.s. has the following upper bound*

$$\tau_{mix} \overset{a.a.s.}{\leq} Cn \log n,$$

for some large enough constant C.

Proof. We will use Lemma 2 and 3. Let μ and μ^b be the measures that generate the transposing neighbours shuffle on $G_{n,p} = (V_{G_{n,p}}, E_{G_{n,p}})$ and $G_n = (V_{G_n}, E_{G_n})$ respectively, and let $\{X_t\}$ and $\{X_t^b\}$ be the random walks on the symmetric group generated by these measures. The benchmark shuffle, ordinary random transpositions, corresponds to neighbour transpositions on the complete graph G_n. For this shuffle, we know that $\|\mathbb{P}(X^b_{(1/2+C_b/\log n)n \log n} \in \cdot) - \pi\|_2^2 \leq 1/4$, for some constant C_b, see [8]. First we will establish an upper bound on A, and then find $t = t(n)$ such that the right hand side of (3.14) converges to something less than $1/4$. This t will be an upper bound on the mixing time τ_{mix}, according to expression (2.3).

Remember that for the neighbour transposition shuffle on $G_{n,p}$ we have $\mu(i\ \ j) = 1/(m+n)$ for $i \neq j$ and $\mu(\text{id}) = n/(m+n)$, where $m = |E_{G_{n,p}}|$ is the number of open edges in $G_{n,p}$. From the expression for A in Lemma 3 we see that

$$\begin{aligned} A &= \max_{x\in\mathfrak{E}(G_{n,p})} \frac{1}{\mu(x)} \sum_{y\in\mathfrak{E}(G_n)} |y| N(x,y)\mu^b(y) \\ &\leq (m+n)\frac{2}{n^2} \max_{x\in\mathfrak{E}(G_{n,p})} \sum_{y\in\mathfrak{E}(G_n)} |y| N(x,y). \end{aligned} \tag{4.7}$$

Using that $m \sim \text{Bin}(\binom{n}{2}, p)$, we can also bound $m+n$.

$$\begin{aligned} &\mathbb{P}\left(m+n > \left(1+\frac{1}{\log n}\right)p\binom{n}{2}\right) = \mathbb{P}\left(m > \left(1+\frac{1+o(1)}{\log n}\right)p\binom{n}{2}\right) = \\ &\mathbb{P}\left(\frac{m}{\binom{n}{2}} > p + \frac{1+o(1)}{\log n}p\right) \underset{(4.1),\,(4.3)}{\leq} \exp\left(-\binom{n}{2}\frac{(1+o(1))p^2}{\log^2 n}\left(\frac{1}{2p}+\frac{1}{2(1-p)}\right)\right) = \\ &\exp\left(-\frac{(1+o(1))n^2 p}{4\log^2 n}\right) \xrightarrow[n\to\infty]{} 0 \end{aligned} \tag{4.8}$$

Together with (4.7) this yields

$$\begin{aligned} A &\overset{\text{a.a.s.}}{\leq} \left(1+\frac{1}{\log n}\right)p\binom{n}{2}\frac{2}{n^2}\max_{x\in\mathfrak{E}(G_{n,p})} \sum_{y\in\mathfrak{E}(G_n)} |y| N(x,y) \\ &= (1+o(1))p \cdot \max_{x\in\mathfrak{E}(G_{n,p})} \sum_{y\in\mathfrak{E}(G_n)} |y| N(x,y). \end{aligned} \tag{4.9}$$

Next step is to bound $N = \max_{x\in\mathfrak{E}(G_{n,p})} \sum_{y\in\mathfrak{E}(G_n)} |y|N(x,y)$, where $N(x,y)$ is the number of times x is utilized in the chosen representation of y, and $|y|$ is the length of the representation. Consider a transposition $y = (a\ \ b) \in \mathfrak{E}(G_n)$. If $y \in \mathfrak{E}(G_{n,p})$, we can use y itself as representation in $\mathfrak{E}(G_{n,p})$. If $y \notin \mathfrak{E}(G_{n,p})$, i.e. the edge between vertex a and b is closed, then the idea is to find an open path between v_a and v_b and use this to form a representation of y in $\mathfrak{E}(G_{n,p})$. To bound N tightly we would like these paths to be short, and that no edge is used too many times. To this end we are interested in finding edge-disjoint paths joining as many of the vertex pairs in $G_{n,p}$ as possible. Broder et al. [4] proved the following theorem.

Theorem 4. *The graph $G_{n,p}$, has with probability $1-o(1)$ the following property as $n \to \infty$: there exist positive constants α and β s.t. for all sets of pairs of vertices $\{(a_j, b_j)|j = 1, 2, \dots, \eta\}$ satisfying:*

(i) $\eta = \left\lceil \alpha n^2 p \frac{\log np}{2\log n} \right\rceil$,

(ii) for each vertex v, $|\{j|a_j = v\}| + |\{j|b_j = v\}| \le \min\{d_v, \beta np\}$,

there exist edge-disjoint paths in $G_{n,p}$, joining a_j to b_j, for each $j = 1, 2, \ldots, \eta$.

For the range of p we consider, β will be small enough to ensure that $\min\{d_v, \beta np\} = \beta np$. We also have $\alpha < \beta < 1$.

We can use Theorem 4 to form paths that we can use in the representations of all transpositions $y \notin \mathfrak{E}(G_{n,p})$. Consider a pair of vertices v_a and v_b. Suppose we have an open path between v_a and v_b of length l: $(a, v_1, v_2, v_3, \ldots, v_{l-2}, v_{l-1}, b)$. We can use the following representation of the transposition $(a\ \ b)$:

$$(a\ \ b) = y = (a\ \ v_1)(v_1\ \ v_2)\cdots(v_{l-2}\ \ v_{l-1})(v_{l-1}\ \ b)(v_{l-2}\ \ v_{l-1})\cdots(a\ \ v_1) \tag{4.10}$$

Broder et al. create edge-disjoint paths between the vertex pairs (a_j, b_j) in the multiset $U := [a_1, b_1, a_2, b_2, \ldots, a_\eta, b_\eta]$ in the following way. First $G_{n,p} = (V_{G_{n,p}}, E_{G_{n,p}})$ is partitioned into five edge-disjoint graphs $G_k = (V_k, E_k)$, $k = 1, \ldots 5$. What is relevant here is that $V_1 = V_{G_{n,p}}$ and $V_2 \subseteq V_{G_{n,p}}$ such that $|V_2| = n - o(n)$. Moreover, this splitting algorithm places each edge of $E_{G_{n,p}}$ independently with probability at least $\frac{5}{6}$ in E_1.

Next, a multiset W, of 2η vertices is chosen from V_2 uniformly at random with replacement. Broder et al. then show that the 2η vertices in U can be connected to those in W, in a one-to-one correspondence, via edge-disjoint paths in E_1. This is done to spread out the endpoint of the vertex pairs over G_2. The vertices in W connected to a_i and b_i are denoted $\tilde{a}_i$ and $\tilde{b}_i$ respectively, so that $W = [\tilde{a}_1, \tilde{b}_1, \tilde{a}_2, \tilde{b}_2, \ldots, \tilde{a}_\eta, \tilde{b}_\eta]$.

After that the authors prove that any set of η vertex pairs from W, $\{(\tilde{a}_j, \tilde{b}_j)|j = 1, 2, \ldots, \eta\}$, can be connected with edge-disjoint paths of lengths $O\left(\frac{\log n}{\log np}\right) = O(1)$ in $E_{G_{n,p}} \setminus E_1$. Thus, there are edge-disjoint paths $(a_j, \tilde{a}_j, \tilde{b}_j, b_j)$ in $G_{n,p}$ connecting the vertex pairs $\{(a_j, b_j)|j = 1, 2, \ldots, \eta\}$.

From (4.10) we see that $|y| = 2l - 1$, where l is the length of the path between a and b. Thus, to bound $N = \max\limits_{x \in \mathfrak{E}(G_{n,p})} \sum_{y \in \mathfrak{E}(G_n)} |y| N(x, y)$ it is essential to find an upper bound for the path length l. We denote this upper bound by $l_{\max}$. Broder et al. does not give any upper bound on the length of the paths from a_j to $\tilde{a}_j$ and from $\tilde{b}_j$ to b_j. So this is our next task.

There are $\binom{n}{2}$ separate vertex pairs in $V_{G_{n,p}}$. Our aim is to connect all of them with short paths that use the same edge as few times as possible. First, to get shorter paths, reduce the number of vertex pairs to connect with edge-disjoint paths from η to $\lceil \frac{\eta}{4} \rceil$. Partition the $\binom{n}{2}$ vertex pairs in $V_{G_{n,p}}$ into sets of at most $\lceil \frac{\eta}{4} \rceil$ pairs. We denote the number of such sets by $Q := \left\lceil \binom{n}{2} \Big/ \lceil \frac{\eta}{4} \rceil \right\rceil$. This yields multisets $U_r^i = [a_1^i, b_1^i, a_2^i, b_2^i, \ldots, a^i_{\lceil \frac{\eta}{4} \rceil}, b^i_{\lceil \frac{\eta}{4} \rceil}]$, $i = 1, 2, \ldots, Q$, of endpoints of the vertex pairs. (For some theory on multisets, see [19].)

An upper bound on Q will be $\frac{4+o(1)}{\alpha\delta p}$, since

$$Q = \left\lceil \frac{\binom{n}{2}}{\lceil \frac{\eta}{4} \rceil} \right\rceil \le \frac{8\binom{n}{2}}{\alpha n^2 p \frac{\log n^\delta}{\log n}} = \frac{4 + o(1)}{\alpha\delta p}. \tag{4.11}$$

For each U_r^i select a multiset W^i of 2η vertices in V_2 as described above. Denote the multiplicity of a member u in U_r^i by r_u^i. Likewise, denote the multiplicity of a member v in W^i by s_v^i. We will show that there exist edge-disjoint paths from U_r^i to a sub-multiset $W_r^i := [\tilde{a}_1^i, \tilde{b}_1^i, \tilde{a}_2^i, \tilde{b}_2^i, \dots, \tilde{a}^i_{\lceil \frac{\eta}{4} \rceil}, \tilde{b}^i_{\lceil \frac{\eta}{4} \rceil}] \subseteq W^i$, such that all of them has length 1, or 0 (then the path consists of a single vertex), for $i = 1, 2, \dots, Q$.

Further, we use the following notation, $U_{r1}^i := [u \in U_r^i | u \in V \setminus V_2]$ and $U_{r2}^i := [u \in U_r^i | u \in V_2]$. We will deal with the paths from U_{r1}^i to W^i and from U_{r2}^i to W^i separately. First, consider the members of U_{r2}^i. We will see that a.a.s. the multiplicity in W^i of such a member u, that is s_u^i, is greater than its multiplicity r_u^i in U_{r2}^i, provided that the U_r^i's are properly chosen. This means that U_{r2}^i is a sub-multiset of W^i. In other words, U_{r2}^i can be connected directly, via "zero-edge paths", to W^i.

The number r_u^i will on average take the value $\frac{2}{n}\lceil \frac{\eta}{4} \rceil$. We can partition the edge pairs in $V_{G_{n,p}}$ so that $\max_{v \in V, i \in [1,Q]} r_v^i \leq \frac{\eta}{n}$. Note that since $\frac{\eta}{n} = \frac{\lceil \alpha n^2 p \frac{\log np}{2 \log n} \rceil}{n} \leq \left\lceil \frac{1}{2} \alpha np \frac{\log np}{\log n} \right\rceil < \beta np$, condition (ii) of Theorem 4 will then be satisfied.

Since W^i is chosen uniformly from V_2, s_u^i will be Bin$(2\eta, \frac{1}{|V_2|})$-distributed for each $u \in V_2$. Using that $|V_2| = n - o(n)$ we can derive an a.a.s. lower bound on $\min_{u \in V_2} s_u^i$ over all Q multisets.

$$
\begin{aligned}
&\mathbb{P}\left(\bigcup_{i=1}^{Q} \left\{ \min_{u \in V_2} s_u^i \leq \frac{2\eta}{|V_2|} \left(1 - \frac{1}{\log n}\right) \right\} \right) \leq \\
&\frac{4 + o(1)}{\alpha \delta p} \mathbb{P}\left(\min_{u \in V_2} s_u^1 \leq \frac{2\eta}{|V_2|} \left(1 - \frac{1}{\log n}\right) \right) \underset{p > n^{\delta - 1}}{\leq} \\
&\frac{(4 + o(1)) n^{2-\delta}}{\alpha \delta} \mathbb{P}\left(s_1^1 \leq \frac{2\eta}{|V_2|} \left(1 - \frac{1}{\log n}\right) \right) \underset{(4.2),\,(4.4)}{\leq} \\
&\frac{(4 + o(1)) n^{2-\delta}}{\alpha \delta} \exp\left(-\frac{\eta}{|V_2| \log^2 n} (1 - o(1)) \right) = \\
&\frac{(4 + o(1)) n^{2-\delta}}{\alpha \delta} \exp\left(-\frac{\alpha np \log np}{2 \log^3 n} (1 - o(1)) \right) \underset{p > n^{\delta - 1}}{\leq} \\
&\frac{(4 + o(1)) n^{2-\delta}}{\alpha \delta} \exp\left(-\frac{\alpha \delta n^{\delta}}{2 \log^2 n} (1 - o(1)) \right) \xrightarrow[n \to \infty]{} 0
\end{aligned} \tag{4.12}
$$

Thus, a.a.s. all members in W^i has multiplicity at least $\frac{2\eta}{|V_2|}\left(1 - \frac{1}{\log n}\right) = \frac{2\eta}{n}(1 - o(1))$, for $i = 1, 2, \dots, Q$. For a member u of U_{r2}^i we have

$$
r_u^i \leq \frac{\eta}{n} < \frac{2\eta}{n}(1 - o(1)) \underset{\text{a.a.s}}{\leq} s_u^i, \quad i = 1, 2, \dots, Q.
$$

Hence, for all $i = 1, 2, \dots, Q$, we can connect the multiset U_{r2}^i directly to the same multiset in W^i via single vertex paths.

It remains to connect U_{r1}^i to $W_1^i := W^i \setminus U_{r2}^i$. We will show that there are such connections in G_1 consisting of only one-edge paths.

Since each edge in $E_{G_{n,p}}$ independently with probability at least $\frac{5}{6}$ is in E_1, each member of U_{r1}^i is a.a.s. neighbour in G_1 to at least $\frac{5np}{6}(1-o(1))$ members of W_1^i. This can be proved with Chernoff's inequality, similar to (4.12).

Remember that the multiplicity of a member u in U_{r1}^i is r_u^i. If we evenly spread out the one-edge connections from u on the neighbour members of W_1^i, no member will receive more than 1 connection, since

$$\frac{r_u^i}{\frac{5np}{6}(1-o(1))} \leq \frac{\eta/n}{\frac{5np}{6}(1-o(1))} \leq \frac{3\alpha\delta}{5}(1+o(1)) < 1.$$

In the last inequality we use the fact that $\alpha < 1$. Furthermore, a member v in W^i is neighbour to at most $|V \setminus V_2| p(1+o(1)) = o(np)$ members in U_{r1}^i. Thus, a member of W^i get at most $o(np)$ connections from U_{r1}^i. Consequently, the total number of connections to a member v in W^i from U_{r1}^i and U_{r2}^i will be less than $r_v + o(np)$. Since

$$r_v + o(np) \leq \frac{\eta}{n} + o(np) \underset{\text{a.a.s}}{<} s_v,$$

we can form edge-disjoint paths in E_1 of maximum length 1 from U_r^i to a sub-multiset W_r^i of W^i. To sum up, we have proved that there are paths in E_1 of maximum length 1 connecting the vertex pairs $\{(a_j^i, \tilde{a}_j^i) | a_j^i \in U_r^i, \tilde{a}_j^i \in W^i, j = 1, 2, \ldots \lceil \frac{\eta}{4} \rceil\}$ and $\{(b_j^i, \tilde{b}_j^i) | b_j^i \in U_r^i, \tilde{b}_j^i \in W^i, j = 1, 2, \ldots \lceil \frac{\eta}{4} \rceil\}$, for $i = 1, 2, \ldots Q$. In addition, according to the results of Broder et al. there are edge-disjoint paths of length $O(1)$ in $E_{G_{n,p}} \setminus E_1$ connecting any set of $\{(\tilde{a}_j^i, \tilde{b}_j^i) | j = 1, 2, \ldots \lceil \frac{\eta}{4} \rceil\}$ of vertex pairs in W^i. This means that for $i = 1, 2, \ldots, Q$ there are edge-disjoint paths in $E_{G_{n,p}}$ of length $O(1)$ connecting the vertex pairs $\{(a_j^i, b_j^i) | j = 1, 2, \ldots \lceil \frac{\eta}{4} \rceil\}$. Thus $l_{\max} = O(1)$.

From (4.10) we can see that a transposition x is used at most twice in each representation, that is

$$\max_{x \in \mathfrak{E}(G_{n,p})} N(x, (a \;\; b)) = 2.$$

In addition, since the paths within each of the Q sets of vertex pairs are edge-disjoint, x can occur in at most Q different representations. Thus, since $|y| \leq 2l_{\max} - 1$ for all $y \in \mathfrak{E}(G_n)$,

$$N = \max_{x \in \mathfrak{E}(G_{n,p})} \sum_{y \in \mathfrak{E}(G_n)} |y| N(x, y) \leq (2l_{\max} - 1)2Q.$$

Moreover, from (4.9) and (4.11), and fact that $l_{\max} = O(1)$, we get the following upper bound on A.

$$A \leq (1+o(1))p \cdot N \leq (1+o(1))p \cdot (2l_{\max} - 1)2Q \leq C_A,$$

for some large enough constant C_A.

Finally, Lemma 3 now yields that

$$\mathcal{E}^b \leq A\mathcal{E} \leq C_A\mathcal{E},$$

where $\mathcal{E}^b$ and $\mathcal{E}$ are the Dirichlet forms associated with the measures generating the neighbour transposition shuffle on the complete graph G_n and the random graph $G_{n,p}$ respectively.

Since $\mu(\mathrm{id}) = \frac{n}{m+n}$, Lemma 4 implies the following inequality for smallest eigenvalue κ_n of the transition matrix for μ.

$$|\kappa_n| \leq 1 - 2\frac{n}{m+n} \overset{\text{a.a.s.}}{\leq} 1 - \frac{n}{p\binom{n}{2}}$$

With $t = \lceil 2C_A n \log n \rceil$, Lemma 2, then entails

$$\begin{aligned}
&\|\mathbb{P}(X_{\lceil 2C_A n \log n \rceil} \in \cdot) - \pi\|_2^2 \leq \\
&n\kappa_n^{2\lceil 2C_A n \log n\rceil} + ne^{-\lceil 2C_A n \log n \rceil / A} + \|\mathbb{P}(X^b_{\lfloor \lceil 2C_A n \log n \rceil / 2A \rfloor} \in \cdot) - \pi\|_2^2 \overset{\text{a.a.s.}}{\leq} \\
&\underbrace{n\left(1 - \frac{n}{p\binom{n}{2}}\right)^{4C_A n \log n}}_{\to 0} + \underbrace{ne^{-2n \log n}}_{\to 0} + \underbrace{\|\mathbb{P}(X^b_{\lfloor n \log n \rfloor} \in \cdot) - \pi\|_2^2}_{<1/4 \text{ as } n \to \infty}.
\end{aligned}$$

In conclusion, together with (2.3), this inequality implies that the random walk $\{X_t\}$, corresponding to the neighbour transposition shuffle on $G_{n,p}$, for p s.t. $1 > p > n^{\delta - 1}$, for some $\delta > 0$, has mixing time $\tau_{\text{mix}} \overset{\text{a.a.s.}}{\leq} \lceil 2C_A n \log n \rceil < Cn \log n$, for $C = 2C_A + 1$.

□

Finally, we observe that Lemma 7, Lemma 8, and Lemma 9 together proves Theorem 3.

Bibliography

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

[2] D. BAYER AND P. DIACONIS. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, **2** (1992), 294-313.

[3] G. BRIGHTWELL AND P. WINKLER. Maximum hitting times for random walks on graphs. *Random Structures and Algorithms* , **1** (1990), 263-276.

[4] A. Z. BRODER, A. M. FRIEZE, S. SUEN, AND E. UPFAL. Optimal construction of edge-disjoint paths in Random graphs. *Siam J. Comput.*, **28** (1998), 541-573.

[5] F. CHUNG AND L. LU. The diameter of sparse random graphs. *Adv. in Appl. Math.*, **26** (2001), 257-279.

[6] P. DIACONIS AND L. SALOFF-COSTE . Comparison techniques for random walk on finite groups. *Ann. Probab.*, **21** (1993), 2131-2156.

[7] P. DIACONIS AND L. SALOFF-COSTE . Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, **3** (1993), 696-730.

[8] P. DIACONIS AND M. SHAHSHAHANI. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, **57** (1981), 159-179.

[9] P. DIACONIS. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences*, **93** (1996), 1659-1664.

[10] P. ERDŐS AND A. RÉNYI. On random graphs I. *Publicationes Mathematicae*, **6** (1959), 290-297.

[11] P. ERDŐS AND A. RÉNYI. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, **5** (1960), 17-61.

[12] U. FEIGE. A tight upper bound on the cover time for random walks on graphs. *Random Structures and Algorithms* , **6** (1995), 51-54.

[13] E. N. GILBERT. Random graphs. *Ann. Math. Stat.*, **30** (1959), 1141-1144.

[14] R. HORN AND C. JOHNSON. Matrix Analysis. Cambridge Univ. Press, (1985).

[15] J. JONASSON. Lollipop graphs are extremal for Commute times. *Random Structures and Algorithms*, **16** (2000), 131-142.

[16] J. JONASSON. The Overhand shuffle mixes in $\Theta(n^2 \log n)$ steps. *Ann. Appl. Probab.*, **16** (2006), 231-243.

[17] P. MATTHEWS. A strong uniform time for random transpositions. *J. Theoret. Probab.*, **1** (1988), 411-423.

[18] R. PEMANTLE. Randomization time for the overhand shuffle. *J. Theoret. Probab.*, **2** (1989), 37-49.

[19] A. SYROPOULOS. Mathematics of multisets. Department of Civil Engineering, Democritus University of Thrace (2001).

[20] D. B. WILSON. Mixing time of the Rudvalis shuffle. *Electron. Commun. Probab.*, **8** (2003), 77-85.

[21] D. B. WILSON. Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.*, **14** (2004), 274-325.