Glauber Dynamics for Sampling an Edge Colouring of Regular Trees

POON, Chun Yeung

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy

in

Computer Science and Engineering

The Chinese University of Hong Kong

September 2016

Abstract

We study the problem of sampling a graph colouring using Glauber Dynamics. This is an interesting problem as Jerrum showed that we can approximate the number of proper colouring if we can sample a colouring nearly uniformly. Therefore we want a sampler with polynomial running time. Glauber Dynamics is one natural Markov Chain for sampling a graph colouring and it has been studied extensively. The state space of it is the set of proper colourings. In each step, we sample a colour c and a node u randomly. Then we update u to colour c if the new colouring is still proper, otherwise we stay at the current colouring. For a graph with maximum degree d, let q be the number of colours. One important goal on this problem is proving polynomial mixing time if $q \geq d + 2$. For general graphs, the best result is polynomial mixing time if $q \geq 11d/6$ by Vigoda. For some classes of graphs, we can sample a colouring for fewer number of colours. One example are graphs with large girth and large maximum degree and there are many results for this kinds of graphs.

In the thesis, we focus on the mixing time of Glauber Dynamics for sampling an edge colouring of a d-regular tree. This is equivalent to sampling a proper vertex colouring of the line graph of the tree. We consider this special case as the line graph has small girth so previous results and techniques does not apply directly. The best previous result is polynomial mixing time if $q \geq 11d/3$ by Vigoda. Our main result is polynomial mixing time if $q \geq 2d$. Our proof is based on the multicommodity flow argument by Sinclair.

摘要

我們研究用 Glauber Dynamics 隨機抽出圖頂點著色的問題。這是個有趣的問題因為 Jerrum 證明如果能隨機抽出頂點著色，則可以找出圖形著色數目的近似值。我們希望有多項式時間內的取樣器。Glauber Dynamics 用馬可夫鏈抽出圖形著色並被大量研究。它的狀態空間是頂點著色的集合。每一步我們隨機抽出一種顏色 c 和頂點 u，並嘗試把 u 的顏色改成 c。如果有衝突則停留在現在的顏色。假設圖的最大度是 d，顏色數目是 q。這問題其中一個目標是証明當 q 不少於 d + 2 時，混合時間是多項式時間。現在最好的成果是 Vigoda 提出的對於所有圖，混合時間是多項式時間如果 q 不少於 11d/6。對於某一種類的圖，我們可以用更少的顏色。其中一個例子是有大周長和大的最大度的圖。

在這篇論文，我們集中在用 Glauber Dynamics 隨機抽出正則樹邊著色的問題。這相當於用 Glauber Dynamics 隨機抽出正則樹的線圖的頂點著色。我們研究這特殊例子因為線圖的周長小因此不能直接應用之前的結果。目前最好的結果是 vigoda 的多項式混合時間如果 q 不少於 11d/3。我們的主要成果是多項式混合時間如果 q 不少於 2d。

Contents

Chapter 1

Introduction

1.1 Background

For a graph $G = (V, E)$, a (proper) vertex colouring of G is a colouring of vertex of G such that no pair of adjacent vertices share the same colour. It is known that counting the number of vertex colourings of general graphs is #-P complete. A related question of interest is to approximately count the number of colourings. Jerrum [11] shows that if we can construct a sampler that return a proper colouring nearly uniformly, then we can approximately count the number of colourings.

One approach for constructing the sampler is using Markov Chain Monte Carlo. The idea is to define an ergodic Markov Chain with uniform stationary distribution. Then we run the chain until it is close to stationary distribution. For sampling a graph colouring, there is a simple Markov Chain called Glauber Dynamics. The state space of it is the set of proper colourings. In each step, we sample a colour c and a node u randomly and independently. Then we change the colour of u to c if the new colouring is still proper. Otherwise we stay at the current colouring.

The running time of the counting algorithm and sampler depends on the mixing time of the Glauber Dynamics. Therefore, it is important to bound the mixing time. There is a natural conjecture[11] about the mixing time of the Glauber Dynamics.

Conjecture 1.1. Let $G = (V, E)$ be a graph with maximum degree d. Let Ω be the set of q-colourings of G. For $q \geq d + 2$, the Glauber Dynamics for sampling a q-colouring of G has mixing time in $O(|V|log|V|)$.

The conjecture is still unproven but steady progress on it has been made throughout the years. Our work focus on a special case of the conjecture.

For background knowledge, this book [13] and this survey [8]provide an introduction to the theory of Markov Chain and Mixing time. This survey [6] summarizes results on the mixing time of the Glauber Dynamics.

1.2 Previous results

We first list the results for general graphs. Let d be the maximum degree of the graph and q be the number of colours. Jerrum [11] used the coupling method to prove $O(q|V|\log|V|)$ mixing time for $q > 2d$ and $O(|V|\log|V|)$ mixing time for $q > 3d$. His coupling method was simplified using path coupling proposed by Bubley and Dyer [3]. Then Vigoda [19] introduced another Markov Chain and proved $O(q|V|\log|V|)$ mixing time of that chain for $q > 11d/6$. His result also implies $O(q|V|^2\log|V|)$ mixing time of Glauber Dynamics for $q > 11d/6$. However, $11d/6$ is still the best bound for the number of colours. We do not know whether the chain has polynomial mixing time or not if we use fewer colours.

The mixing time of the Glauber Dynamics in some restricted classes of graphs are also studied. One common restriction is that the graph has large girth and large maximum degree. Under this restriction, we can sample a colouring for a much smaller number of colours. Dyer and Frieze [4] proved that for graph with $d = \Omega(\log n)$ and girth $g = \Omega(\log d)$, the Glauber Dynamics has $O(n\log n)$ mixing time if $q > \alpha d$ for a constant $\alpha \approx 1.763$. Then there are many results to reduce the number of colours, lower the girth requirement and lower the degree requirement and we will list some of them. Hayes and Vigoda [10] proved $O(n\log n)$ mixing time for $q > (1+\varepsilon)d$ for all $\varepsilon > 0$ if girth is at least 11 and $d = \Omega(\log n)$. Dyer, Frieze, Hayes and Vigoda [5] proved $O(n\log n)$ mixing time if $q > (1+\varepsilon)\beta d$ for a constant $\beta \approx 1.489$ for graphs with girth at least 7 and d at least a constant that grows with $1/\varepsilon$.

The other kind of graphs people studied are planar graphs and trees. A notable result are from Hayes, Vera and Vigoda [9], they reduce the number of colours to $d/\log d$ for planar graphs. They proved polynomial mixing time for planar graphs with $q = \Omega(d/\log d)$. Then two more results [7, 14]studied the mixing time for trees if we use even fewer colours.

1.3 Our work

Our work focus on the Glauber Dynamics for sampling an edge colouring of a d-regular tree T. It is equivalent to Glauber Dynamics for sampling a vertex colouring of the line graph of T. Let q be the number of colours. We prove that if $2d \leq q \leq 4d$, then the Glauber Dynamics will have polynomial mixing time. Our proof is based on the multicommodity flow argument.

Previous results on edge colouring of regular tree

There are not many previous results on this kind of graphs. The only relevant results are the results on general graphs. Let $|T|$ denote the number of edge of T. For $q \geq 4d$, Jerrum's result [11] show that the mixing time is $O(q|T| \log |T|)$. For $q \geq 11d/3$, Vigoda's result [19] show that the mixing time is $O(|T|^2)$.

Motivation

The first motivation of our work is studying the use of multicommodity flow [18] on bounding the mixing time of Glauber Dynamics. Both multicommodity flow and coupling are powerful methods for bounding mixing time of Markov Chains. However, many previous results are based on coupling and we want to see what we can get with multicommodity flow. Also, many previous results based on coupling seems to be weak against locally dense graph. The line graph of a regular tree is a simple example of locally dense graph so we start with it. We also want to give a better bound on the number of colours. We want to show polynomial mixing time if we only have $2d$ colours. Note that the line graph has maximum degree $2d - 2$, $2d$ is the number mentioned in the conjecture.

Our main result is the following theorem.

Theorem 1.2. *Let q be the number of colours. Let M be the Glauber Dynamics for sampling an edge colouring of a d-regular tree T. Let h be the height of T. If $4d \geq q \geq 2d$, then*

$$\tau_{mix} \leq 2^{21h} q |T|^8 \ln^{2h}(2q) \ln(2|T|)$$

We remark that the mixing is in polynomial of $|T|$ since $|T| = d^h$ and $q \leq 4d$.

1.4 Future work

It would be interesting to extend the results in the following directions. We bound the mixing time of Glauber Dynamics for sampling an edge colouring of regular trees. It would be nice if we can extend the results to edge colouring of general trees. One part of our analysis use induction by decomposing the tree into several subtrees and treat the subtrees as smaller instances. So we decompose the tree recursively and the mixing time is of the form c^l where the l is the depth of recursion. Currently we decompose the tree by removing the root, so the result can be exponential if the tree is not balance. If we use a better way to decompose the tree then we may be able to extend the results to general trees.

Another possible work is improving the results. For example, can we prove the mixing time is still polynomial if we use fewer colours? If we use fewer colours, then we may be unable to update some edges for some colouring. And we need to prove the chain is still ergodic. Also, we did not try to minimize the mixing time, so one possible work is to find a better upper bound on the mixing time. It is also nice to find the lower bound of mixing time.

1.5 Organization

In Chapter 2 we give background material for this thesis. We define Markov Chain and present some methods for bounding the mixing time of a Markov Chain. We also introduce the problem of sampling a graph colouring and give background knowledge.

In Chapter 3 we present our main work. We study the Glauber Dynamics for sampling an edge colouring of regular tree. We give an upper bound on the mixing time of the Glauber Dynamics.

Chapter 2

Background

In this chapter, we present background knowledge for this thesis. We first introduce the Markov Chain, a famous tool for Monte Carlo simulation. Then we present some methods for bounding the mixing time of a Markov Chain. We talk about bounding mixing time using eigenvalue gap of transition matrix, multicommodity flow, and also coupling. Finally, we introduce the problem of approximate counting of graph colouring.

2.1 Markov Chain

In this section we introduce Markov Chain.

Definition 2.1. Let M be a discrete stochastic process on a state space Ω. Let X_t be the state of M at time t. M is called a Markov Chain if it has the following property for all time t and $x, y \in \Omega$.

$$\Pr(X_{t+1} = y | X_t = x, X_{t-1} = x_{t-1} ..., X_0 = x_0) = \Pr(X_{t+1} = y | X_t = x)$$

The above property is also called Markov property.

Let Ω be the state space, we can write down a $|\Omega| \times |\Omega|$ matrix P where $P_{x,y} = \Pr(X_{t+1} = y | X_t = x)$. We call P a probability transition matrix. A probability distribution π on Ω is a stationary distribution if and only if $\pi P = \pi$. A chain is irreducible if for every x, y, there is a time t such that $P^t_{x,y} > 0$. A chain is irreducible means that no matter which state we start from, we can go to every other states. A chain is aperiodic if for every x, $\gcd\{t : P^t_{x,x} > 0\} = 1$. We called a chain is ergodic if it is both aperiodic and irreducible. One of the most important property of an ergodic chain is that it has an unique stationary distribution.

Lemma 2.2. *If a Markov Chain is irreducible and aperiodic, then there exists an unique stationary distribution π.*

Proof. Let P be the transition matrix of the Markov Chain. Since the chain is irreducible and aperiodic, then there exists a time t such that P^t is a positive matrix. Let $P' = P^t$, We will argue that P' has a unique stationary distribution.

Suppose there are two stationary distribution u, v. Then there is a state x such that $\frac{u(x)}{v(x)} = \min_y \frac{u(y)}{v(y)}$. Then

$$\begin{aligned} u(x) &= \sum_y u(y)P'(y,x) \\ &= \sum_y \frac{u(y)}{v(y)} v(y)P'(y,x) \\ &\geq \frac{u(x)}{v(x)} \sum_y v(y)P'(y,x) \\ &= \frac{u(x)}{v(x)} v(x) \\ &= u(x) \end{aligned}$$

Since P' is positive and both sides are equal, the inequality is actually an equality. We will have $\frac{u(x)}{v(x)} = \frac{u(y)}{v(y)} \forall y$. Since u and v are probability distributions, we must have $u = v$. So P' has a unique stationary distribution, it implies P also must has a unique stationary distribution.

□

In the remaining part, we will only consider finite ergodic chain with symmetric transition matrix. However, many results can also be extended to non-symmetric case.

2.1.1 Mixing time

An important measure of a Markov Chain is its convergence rate. It tells us how fast the Markov Chain will converge to stationary distribution. We first introduce the metric for probability distribution. Then we prove a Markov Chain will converge to its stationary distribution and introduce mixing time.

Definition 2.3. Let p, q be two probability distributions on state space Ω. The total variation distance is defined by

$$\|p - q\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |p(x) - q(x)|$$

The total variation distance also equals to l_1-norm of $p - q$ divided by two.

Consider an ergodic Markov Chain M with symmetric transition matrix P. Let π be the stationary distribution. Let P_x^t be the distribution after t steps if we start at state x. We will show that for any starting state x, P_x^t converges to stationary distribution π when t tends to infinity.

We first make a claim that tells us that P_x^t converges to π if P is a positive matrix. We will prove it later in 2.5.

Claim 2.4. For any finite ergodic chain with transition matrix P, let $\delta_j = \min_i P(i,j)$ be the smallest entry of j^{th} column. Let $\delta = \sum_i \delta_i$, then for any starting position x and time t,

$$\|P_x^t - \pi\|_{TV} \leq (1-\delta)^t$$

If P is a positive matrix, it is easy to see that δ will be positive and P_x^t converges to π. Now we will show that the probability distribution of any finite ergodic chain converges to π.

Lemma 2.5. *For any finite ergodic Markov Chain M with probability transition matrix P. Let π be the stationary distribution. Then for any state x,*

$$\lim_{t\to\infty} \|P_x^t - \pi\| = 0$$

Proof. Since the chain is ergodic, from definition, there exists a time N such that for $t > N$, $P^t(x,y) > 0$ for every $x, y \in \Omega$. Therefore P^N is a positive matrix and define an ergodic chain. Then by the previous claim, there exists $0 < \delta < 1$ such that $\|P_x^{Nt} - \pi\|_{TV} \leq (1-\delta)^t$ for any state x. So when t tends to infinity, P_x^t will converge to π. $\square$

In the end of this section we define mixing time and fast mixing.

Definition 2.6. The mixing time of a Markov Chain is defined by

$$\tau(\varepsilon) = \max_x \min\{t : \left\|P_x^t - \pi\right\|_{TV} < \varepsilon\}$$

which is time needed to guarantee the distance between the distribution and π is at most ε no matter which state we started at. Also, we define

$$\tau_{mix} = \tau(\frac{1}{4}) = \max_x \min\{t : \left\|P_x^t - \pi\right\|_{TV} < \frac{1}{4}\}$$

A chain is fast mixing or rapidly mixing if τ_ε is in polynomial in $\ln(n)$ and $\ln(\varepsilon^{-1})$. Alternatively, showing that τ_{mix} is in polynomial in $\ln(n)$ is also enough for fast mixing by the following lemma.

Lemma 2.7. $\tau(\varepsilon) \leq \lceil \log_2(\varepsilon^{-1}) \rceil \tau_{mix}$

We have introduced the terms and definitions for convergence rate. We will introduce techniques for bounding mixing time of a Markov Chain.

2.2 Spectral theory

The first technique is about spectral theory. Note that $\pi^T P = \pi$ implies π is a left eigenvector of P with eigenvalue 1. It is natural that we study the eigenvalues and eigenvectors to learn more about the chain. In fact, we get information about the mixing time.

We need the Perron-Frobenius Theorem [17] for the prove in this part.

Theorem 2.8 (Perron-Frobenius)**.** *For any finite ergodic Markov Chain with transition matrix P.*

- *The first eigenvalue λ_1 of P has multiplicity* 1*.*
- *The other eigenvalues of P satisfies $|\lambda_i| < \lambda_1$*

Let $\{\lambda_i\}_{i=1}^n$ be eigenvalue of P with corresponding eigenvector $\{v_i\}_{i=1}^n$. We first show some properties of them.

Lemma 2.9. *P has n real eigenvalue* $1 = \lambda_1 > \lambda_2 \geq ...\lambda_n > -1$.

Proof. We first prove all eigenvalues are real using contradiction. Suppose P has a complex eigenvalue λ with eigenvector v. Since P is symmetric, $v^T(\lambda v) = v^T P v = (Pv)^T v$. Then by property of inner product, $(Pv)^T v = \overline{v^T P v} = \overline{\lambda} v^T v$. Since $v \neq 0$, we have $\lambda = \overline{\lambda}$. Therefore the eigenvalue must be a real number.

Now we give bound to the eigenvalue. It is easy to check that the stationary distribution π is an eigenvector of P with eigenvalue 1. Consider an eigenvalue λ with eigenvector v. Let k be the index with largest absolute entry such that $|v(k)| \geq |v(i)| \forall i$. Then $|\lambda v(k)| = |(Pv)(k)| = \sum_j P(k,j)|v(j)| \leq |v(k)| \sum_j P(k,j) = |v(k)|$. Therefore $|\lambda_i| \leq 1$ with $\lambda_1 = 1$.

Finally, the multiplicity of λ_1 is 1 and $\lambda_n > -1$ by the Perron-Frobenius theorem.

□

Let $\lambda_{max} = \max\{|\lambda_2|, |\lambda_n|\}$. Define the spectral gap $g = 1 - \lambda_{max}$ to be the gap between largest eigenvalue and second largest eigenvalue. Now we will show the relationship between mixing time and spectral gap. Roughly speaking, mixing time is bounded by inverse of spectral gap. So large spectral gap implies fast mixing.

Lemma 2.10. $\tau(\varepsilon) \leq \frac{1}{g} \ln \frac{n}{2\varepsilon}$

Proof. Let u be the starting distribution. We can decompose it to $u = \sum_i c_i v_i$ where $c_i = u^T v_i$. In particular, $c_1 = 1/\sqrt{n}$, so $c_1 v_1 = \pi$.

After t step, the distribution we have is $u^T P^t = \sum_i c_i \lambda_i^t v_i$. The square of l_2 norm between uniform distribution and $u^T P^t$ is

$$\begin{aligned}
\left\|\pi - u^T P^t\right\|_2^2 &= \left\|\sum_{i \geq 2} c_i \lambda_i^t v_i\right\|_2^2 \\
&= \left\|\sum_{i \geq 2} \lambda_i^t c_i v_i\right\|_2^2 \\
&= \sum_{i \geq 2} \lambda_i^{2t} u^T v_i v_i^T u \\
&\leq \lambda_2^{2t} u^T (\sum_{i \geq 1} v_i v_i^T - v_1 v_1^T) u \\
&= \lambda_2^{2t} (u^T P u - \frac{1}{n}) \\
&\leq \lambda_2^{2t} (1 - \frac{1}{n})
\end{aligned}$$

Using l_2 norm, we can get an upper bound of l_1 norm,

$$\begin{aligned}
\left\|\pi - u^T P^t\right\|_{TV} &= \frac{1}{2} \left\|\pi - u^T P^t\right\|_1 \\
&\leq \frac{1}{2} n^{1/2} \left\|\pi - u^T P^t\right\|_2 \\
&= \frac{1}{2} \lambda_2^t (n-1)^{1/2}
\end{aligned}$$

To upper bound the distance by ε, we get

$$\frac{1}{2} \lambda_2^t (n-1)^{1/2} \leq \varepsilon$$

Solving it and we will get

$$t \geq \frac{1}{g} ln \frac{n}{2\varepsilon}$$

□

Therefore we can also prove a chain is rapid mixing by showing the spectral gap is large.

Usually, we can ignore λ_n by considering the lazy chain $P' = \frac{1}{2}(I + P)$. All eigenvalue of P' is non-negative so spectral gap is $1 - \lambda_2$. Also the mixing time of lazy chain is only twice of original chain. This allows us to focus on bounding λ_2 from 1.

2.2.1 Characterization of λ_2

We will show two characterizations of the second largest eigenvalue of symmetric transition matrix.

Lemma 2.11. *For any real symmetric probability transition matrix P, let $1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ be the eigenvalues of P with orthonormal eigenvectors $\overrightarrow{1} = v_1, v_1, ..., v_n$. Then $\lambda_2 = \max_{x : x \perp \overrightarrow{1}} \frac{x^T P x}{x^T x}$*

Proof. For any $x \perp v_1$, we can write x as $\sum_{i \geq 2} c_i v_i$.

So

$$\frac{x^T P x}{x^T x} = \frac{\sum_{i \geq 2} c_i^2 \lambda_i}{\sum_{i \geq 2} c_i^2} \leq \lambda_2 \frac{\sum_{i \geq 2} c_i^2}{\sum_{i \geq 2} c_i^2} = \lambda_2$$

And the equality holds when $x = v_2$, so the lemma is true. □

Lemma 2.12. *For any real symmetric probability transition matrix P, let $1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ be the eigenvalues of P with orthonormal eigenvectors $\overrightarrow{1} = v_1, v_1, ..., v_n$. Let π be the stationary distribution of P. Then λ_2 satisfies*

$$1 - \lambda_2 = \min_{x : x \perp \overrightarrow{1}} \frac{\sum_{i,j} (x(i) - x(j))^2 \pi(i) P(i,j)}{\sum_{i,j} (x(i) - x(j))^2 \pi(i) \pi(j)}$$

Proof. By 2.11, we have $\lambda_2 = \max_{x : x \perp \overrightarrow{1}} \frac{x^T P x}{x^T x}$. So, $1 - \lambda_2 = \min_{x : x \perp \overrightarrow{1}} \frac{x^T (I - P) x}{x^T x}$

We first look at denominator. Let X be the random variable that takes value $x(i)$ with probability $\pi(i) = 1/n$, then

$$\begin{aligned} x^T x &= n\,\mathrm{E}(X^2) \\ &= n\,\mathrm{Var}(X) \text{ Since } x \perp \overrightarrow{1} \\ &= \frac{n}{2}\sum_{i,j}(x(i) - x(j))\pi(i)\pi(j) \end{aligned}$$

For nominator,

$$\begin{aligned} x^T(I-P)x &= \sum_i x(i)^2 - \sum_{i,j} P(i,j)x(i)x(j) \\ &= \sum_i \sum_{j>i} (x(i) - x(j))^2 P(i,j) \\ &= \frac{1}{2}\sum_{i,j}(x(i) - x(j))^2 P(i,j) \end{aligned}$$

So

$$1 - \lambda_2 = \min_{x:x\perp\overrightarrow{1}} \frac{x^T(I-P)x}{x^T x} = \min_{x:x\perp\overrightarrow{1}} \frac{\sum_{i,j}(x(i)-x(j))^2\pi(i)P(i,j)}{\sum_{i,j}(x(i)-x(j))^2\pi(i)\pi(j)}$$

□

2.3 Conductance

The second technique is about viewing the Markov Chain as a graph and use Cheeger's inequality [1] to bound the spectral gap. For any Markov Chain M with finite state space and symmetric transition matrix, we can construct a weighted undirected graph $G = (\Omega, E)$. The vertex set of G is the state space Ω and edge set is $E = \{(u,v) : P(u,v) > 0\}$. The weight of edge (u,v) is $P_{u,v}$.

First we define the term expansion.

Definition 2.13. The expansion of a set $S \subset \Omega$ is defined as

$$\phi(S) = \frac{\sum_{(u,v)\in S\times\overline{S}} P(u,v)}{|S|} \tag{2.1}$$

And the expansion of graph G is defined by

$$\phi = \phi(G) = \min_{S \subset \Omega, 0 < |S| \leq 1/2} \phi(S) \tag{2.2}$$

It is natural to consider expansion when bounding mixing time. If the expansion is small, then there is a set S such that we have high probability to stay at S once we enter S. Hence the mixing time may be large. If the expansion is large, we leave S with high probability for any set S so we may appear at any state after few steps. Therefore the mixing time may be small.

Cheeger's inequality tells us the relationship between expansion and λ_2.

Lemma 2.14.

$$1 - 2\phi \leq \lambda_2 \leq 1 - \frac{\phi^2}{2} \tag{2.3}$$

By considering lazy chain, we know that spectral gap is $1 - \lambda_2$. Cheeger's inequality allows us to bound spectral gap using expansion. Therefore we can obtain an upper bound of mixing time with expansion.

Lemma 2.15. $\tau_{mix}(M) \leq \frac{2\ln(2n)}{\phi^2}$

Proof. By lemma 2.10,

$$\tau_{mix}(M) \leq \frac{1}{1 - \lambda_2} \ln(2n)$$

Then applying lemma 2.14,

$$1 - \lambda_2 \geq \frac{\phi^2}{2}$$

So,

$$\tau_{mix}(M) \leq \frac{2\ln(2n)}{\phi^2}$$

□

2.4 Multicommodity flow

In this section we will introduce the multicommodity flow argument by Sinclair [18] for bounding the mixing time of Markov Chain. This is another technique that bounds mixing time by giving a lower bound on spectral gap.

Definition

Let M be a ergodic Markov Chain on state space Ω with symmetric transition matrix P. Suppose the stationary distribution π is uniform. Let $E = \{(X,Y) : P(X,Y) > 0\}$ be the set of transitions of M. For any pair of distinct states X, Y, let $\mathcal{P}_{X,Y}$ be the set of simple paths from X to Y using only transitions in E. Let $\mathcal{P} = \cup_{X \neq Y} \mathcal{P}_{X,Y}$. A multicommodity flow of M is a fractional flow that send 1 unit of flow from X to Y for every pair of distinct X, Y. We define the flow by a function $f : \mathcal{P} \to \mathbb{R}^+ \cup \{0\}$. The input is a path p and f returns a non negative real number which is the size of the flow on p. For every distinct pair of states X, Y, the demand of f is $D(X,Y) = 1$. In other word, f must satisfy the constraint for every distinct X, Y,

$$\sum_{p \in \mathcal{P}_{X,Y}} f(p) = 1$$

The "quality" of a flow f can be measured by its congestion $\rho(f)$

$$\rho(f) = \max_{(X,Y) \in E} \frac{1}{|\Omega| P(X,Y)} \sum_{p \in \mathcal{P} : (X,Y) \in p} f(p)$$

And we define the elongated congestion by considering the length.

$$\overline{\rho}(f) = \max_{e(x,y) \in E} \frac{1}{|\Omega| P(e)} \sum_{p : e \in p} f(p)|p| \tag{2.4}$$

The following lemma relates second largest eigenvalue and congestion.

Lemma 2.16. *For any ergodic Markov Chain with symmetric transition matrix P, and any multicommodity flow f, the second largest eigenvalue satisfies $\lambda_2 \leq 1 - \frac{1}{\overline{\rho}(f)}$*

Proof. Note that the stationary distribution will be uniform since the transition matrix is symmetric.a

By lemma 2.12, λ_2 satisfies

$$1 - \lambda_2 = \min_{x : x \perp \vec{1}} \frac{\sum_{i,j} (x(i) - x(j))^2 \pi(i) P(i,j)}{\sum_{i,j} (x(i) - x(j))^2 \pi(i) \pi(j)}$$

For any flow f and x, we can rewrite the denominator as,

$$\sum_{i,j} (x(i) - x(j))^2 \pi(i)\pi(j) \quad = \quad \sum_{i,j} \pi(i)\pi(j) \sum_{p \in \mathcal{P}_{i,j}} f(p)(x(i) - x(j))^2$$

Let e^+ and e^- denotes the start and end of a directed edge e respectively, We can write the equation as

$$\begin{aligned}
\sum_{i,j}(x(i)-x(j))^2\pi(i)\pi(j) &= \sum_{i,j}\pi(i)\pi(j)\sum_{p\in\mathcal{P}_{i,j}} f(p)\left(\sum_{e\in p} x(e^+)-x(e^-)\right)^2 \\
&\leq \sum_{i,j}\pi(i)\pi(j)\sum_{p\in\mathcal{P}_{i,j}} f(p)|p|\sum_{e\in p}(x(e^+)-x(e^-))^2 \\
&= \sum_{e}(x(e^+)-x(e^-))^2\sum_{i,j}\sum_{p\in\mathcal{P}_{i,j}:e\in p}\pi(i)\pi(j)f(p)|p| \\
&= \sum_{e}(x(e^+)-x(e^-))^2\pi(i)P(i,j)\overline{\rho}(e) \\
&\leq \overline{\rho}(f)\sum_{e}(x(e^+)-x(e^-))^2\pi(i)P(i,j)
\end{aligned}$$

where the first inequality comes from Cauchy-Schwartz inequality. So we have

$$\frac{\sum_{i,j}(x(i)-x(j))^2\pi(i)P(i,j)}{\sum_{i,j}(x(i)-x(j))^2\pi(i)\pi(j)} \geq \frac{1}{\overline{\rho}(f)}$$

and the lemma follows. □

Let $l(f) = \max_{p:f(p)>0}|p|$ be the longest path length in f. We can have the following corollary.

Corollary 2.17. *For any Markov Chain with symmetric transition matrix, and any multicommodity flow f, the second largest eigenvalue of P satisfies $\lambda_2 \leq 1 - \frac{1}{\rho(f)l(f)}$, where $l(f)$ is the length of the longest path p with $f(p) > 0$.*

Now we will introduce the relation between congestion and mixing time. We have obtained an upper bound of second largest eigenvalue, if we apply lemma 2.10, we can bound the mixing time of M using $\rho(f)$.

Theorem 2.18. *For any Markov Chain M with symmetric transition matrix, and any multicommodity flow f of M. Let n be the number of states, the mixing time of M is bounded by*

$$\tau_{mix}(M) \quad\leq\quad \rho(f)l(f)\ln(2n)$$

2.5 Coupling

The last technique we introduce is coupling. It is a simple but powerful technique for bounding mixing time.

Let M be a Markov Chain on a state space Ω. Let P be the probability transition matrix. A coupling is a Markov Chain $Z_t = (X_t, Y_t)$ on the state space $\Omega \times \Omega$ such that

$$\begin{aligned} \Pr(X_{t+1} = x_{t+1} | Z_t = (x_t, y_t)) &= P(x_t, x_{t+1}) \\ \Pr(Y_{t+1} = y_{t+1} | Z_t = (x_t, y_t)) &= P(y_t, y_{t+1}) \end{aligned}$$

So a coupling is a joint process of two copies of the original chain such that if we look at only one chain, it behaves exactly as the original chain. However, they do not need to be independent. The following lemma tells us the power of coupling.

Lemma 2.19 (Coupling Lemma). *Let M be a finite ergodic Markov Chain on a state space Ω. Let $Z_t = (X_t, Y_t)$ be a coupling of M. If $Pr(X_t \neq Y_t | Z_0 = (x_0, y_0)) < \varepsilon \; \forall x_0, y_0 \in \Omega$, then $\tau(\varepsilon) \leq t$.*

Proof. Suppose initially, $X_0 = x$ and Y_0 is chosen according to stationary distribution. Let P_x^t be the probability distribution of chain X at time t. We are going to bound $\max_x \left\| P_x^t - \pi \right\|_{TV}$. For any set $S \in \Omega$, the probability that $X_t \in S$ is

$$\begin{aligned} \Pr(X_t \in S) &\geq \Pr(X_t \in S \text{ and } Y_t \in S) \\ &\geq \Pr(Y_t \in S) - \Pr(X_t \neq Y_t) \\ &= \pi(S) - \varepsilon \end{aligned}$$

Similarly, we can obtain the bound $\Pr(X_t \in \overline{S}) \geq \pi(\overline{S}) - \varepsilon$. So $\max_{S \subseteq \Omega} |Pr(X_t \in S) - \pi(S)| \leq \varepsilon$, which implies $\max_x \left\| P_x^t - \pi \right\|_{TV} \leq \varepsilon$ and $\tau(\varepsilon) \leq t$.

□

For any integer-value metric d on Ω. For $\varepsilon > 0$, we say a chain is ε distance decreasing if there exists a coupling such that for every x, y, $\mathrm{E}(d(X_1, Y_1 | X_0 = x, Y_0 = y)) < (1 - \varepsilon) d(x, y)$.

We can obtain a bound on mixing time by showing that the chain is ε distance decreasing.

Lemma 2.20. *Let $d_{max} = \max_{x,y \in \Omega} d(x, y)$. If a Markov Chain is ε distance decreasing for $\varepsilon > 0$. Then $\tau(1/4) \leq \varepsilon^{-1} \ln(4 d_{max})$*

Proof. If we apply the strategy t times, probability that X and Y is different at time t is

$$\begin{aligned}
\Pr(X_t \neq Y_t|X_0, Y_0) &\leq \mathrm{E}(d(X_t, Y_t)|X_0, Y_0) && (2.5)\\
&\leq (1-\varepsilon)\,\mathrm{E}(d(X_{t-1}, Y_{t-1})|X_0, Y_0) && (2.6)\\
&\leq (1-\varepsilon)^t d_{max} && (2.7)\\
&\leq \exp(-\varepsilon t) d_{max} && (2.8)
\end{aligned}$$

When $t \geq \varepsilon^{-1} \ln(4d_{max})$, we have $\exp(-\varepsilon t)d_{max} \leq 1/4$. So $\tau_{mix} \leq \varepsilon^{-1} \ln(4d_{max})$ by previous lemma. □

Using coupling, we can prove claim 2.4.

Proof. Consider two copies of the Markov Chain M. By definition $\delta = \min_{i,j} P(i, j)$. Therefore we can construct a coupling that the copies of the chain move to the same state with probability at least δ. So the probability they are at some different states is at most $(1 - \delta)^t$ after t steps. Then we have the claim by the coupling lemma. □

2.5.1 Path coupling

Although coupling is a powerful tool, sometimes it is hard to find a coupling strategy for every pair of states. Bubley and Dyer [3] introduced an idea called "path coupling". In path coupling, we only need to define coupling strategy for adjacent states. Consider a connected graph $G = (\Omega, E)$ where $E \subseteq \Omega \times \Omega$. Let $d(X, Y)$ be the shortest distance between X and Y in G.

Lemma 2.21. *If there exists a coupling such that* $\mathrm{E}(d(X_1, Y_1)|X_0 = x, Y_0 = y) < (1 - \varepsilon)d(x, y)\forall(x, y) \in E$, *then the chain is* ε *distance decreasing.*

Proof. The idea is to construct a coupling for all pair x, y by using coupling on states of edges of the shortest path. Let $l = d(x, y)$. Let the x, y shortest path be $x = z_0, z_1, ..., z_l = y$. Our goal will be finding a way to sample x' and y' such that $\Pr(X_1 = x'|X_0 = x, Y_0 = y)) = P(x, x')$, $\Pr(Y_1 = y'|X_0 = x, Y_0 = y)) = P(y, y')$ and $\mathrm{E}(d(x', y')|x, y) < (1 - \varepsilon)d(x, y)$.

We already have a coupling strategy for $l = 1$. The coupling can be defined as a stochastic process C_t on $\Omega \times \Omega$. And for $C_0(x, y) \in E$, let $C_1 = (x', y')$ be the next state, we expect $d(x', y') < (1 - \varepsilon)d(x, y)$.

For $l > 1$, we will sample $x' = z'_0, z'_1, ..., z'_l = y'$ using the follow way. We first sample z'_0 and z'_1 by applying the coupling C on (z_0, z_1). We can do this because $(z_0, z_1) \in E$. Then for $i \geq 1$, we sample z'_{i+1} by applying coupling C on $(X_0, Y_0) = (z_i, z_{i+1})$ and condition on $X_1 = z_i$.

Then we prove that $\Pr(z'_i = v|z_i) = P(z_i, v)$ by induction. It is true for z'_0, z'_1 as we sample them from valid coupling. Assume it is true for z'_i, consider z'_{i+1},

$$\begin{aligned}
\Pr(z'_{i+1} = v|z_{i+1}) &= \sum_u \left[\Pr(z'_i = u|z_i) \Pr(C_1 = (z'_i, v)|C_0 = (z_i, z_{i+1}), z'_i = u)\right] \\
&= \sum_u \left[P(z_i, u) \Pr(C_1 = (z'_i, v)|C_0 = (z_i, z_{i+1}), z'_i = u)\right] \\
&= \sum_u P(z_i, u) \frac{\Pr(C_1 = (u, v)|C_0 = (z_i, z_{i+1})))}{\sum_{v'} \Pr(C_1 = (u, v')|C_0 = (z_i, z_{i+1}))} \\
&= \sum_u P(z_i, u) \frac{\Pr(C_1 = (u, v)|C_0 = (z_i, z_{i+1})))}{P(z_i, u)} \\
&= \sum_u \Pr(C_1 = (u, v)|C_0 = (z_i, z_{i+1}))) \\
&= P(z_{i+1}, v)
\end{aligned}$$

So we can see that the coupling is valid and well defined. If we apply this coupling, then the expected distance at 1 step is

$$\begin{aligned}
\mathrm{E}(d(X_1, Y_1)|X_0 = x, Y_0 = y) &= \mathrm{E}(d(z'_0, z'_l)) \\
&= \sum_{0 \leq i \leq l-1} \mathrm{E}(d(z'_i, z'_{i+1})) \\
&< \sum_{0 \leq i \leq l-1} (1 - \varepsilon) d(z_i, z_{i+1}) \\
&= (1 - \varepsilon) d(x, y)
\end{aligned}$$

Therefore the chain is ε distance decreasing. □

We will have the following corollary on mixing time if we apply lemma 2.20 directly.

Corollary 2.22. *If there exists a coupling such that* $\mathrm{E}(d(X_1, Y_1)|X_0 = x, Y_0 = y) < (1 - \varepsilon) d(x, y) \forall (x, y) \in E$*, then the mixing time of the chain is at most* $\varepsilon^{-1} \ln(4 d_{max})$.

2.6 Approximate sampling proper colouring

In this section we introduce the problem of approximate counting and sampling of graph colouring. To begin, we first give some definitions about approximate counting and sampling. We will use the definition from this book[16].

2.6.1 Approximate sampling and approximate counting

Definition 2.23. A randomized algorithm gives an ε-approximate for a value V if the output X of the algorithm satisfies

$$\Pr(|X - V| \leq \varepsilon V) \geq \frac{3}{4}$$

Definition 2.24. A fully polynomial randomized approximation scheme (FPRAS) for a function f is a randomized algorithm for which, given an input x and any parameters ε with $0 < \varepsilon < 1$, the algorithm outputs an ε-approximation to $f(x)$ with probability at least $3/4$ and runs in time that is polynomial in $1/\varepsilon$ and the size of input x.

Definition 2.25. Let Ω be the state space. Let π be the uniform distribution on Ω. An algorithm generate an ε-uniform sample of Ω if it returns a state x from a distribution u and $\|u - \pi\|_{TV} \leq \varepsilon$.

Definition 2.26. A fully polynomial almost uniform sampler (FPAUS) for a problem is an algorithm for which, given an input x and a parameter $\varepsilon > 0$, it returns an ε-uniform sample from $\Omega(x)$ and runs in time that is polynomial in $\ln(1/\varepsilon)$ and size of the input x.

Now we will introduce the problem of counting the number of proper colourings.

Let $G = (V, E)$ be a simple graph. Let $[q] = \{0, 1, ..., q-1\}$ be the set of q colours. A (proper) vertex q-colouring of G is a colouring of vertex of G such that no pair of adjacent vertices share the same colour. Let d be the maximum degree of G. It is not hard to see that we can get a colouring if $q \geq d + 2$. A natural problem is how many q-colourings does G has.

However counting the number of q-colourings is #P-complete. Hence we focus on approximate count the number. Jerrum [11] showed that if we can sample a colouring near uniformly, then we can approximate the number of colourings.

Lemma 2.27. *For any graph G with maximum degree d. There is a FPRAS for counting the number of q-colourings of a graph G if there exists a FPAUS for sampling a q proper colouring for $q \geq d + 2$.*

Proof.

Constructing an approximate scheme

We first show how we get an approximate scheme. Let n be the number of nodes in G. Let m be the number of the edges in G. Let d be the maximum degree of G. To avoid trivialities, we assume $n \geq 3$ and $d \geq 2$. Then we can construct a sequence of graph $G = G_m, G_{m-1}, ..., G_1, G_0 = (V, \emptyset)$ where G_i are obtained by removing a single edge from G_{i+1}. Let $\Omega(G_i)$ denote the set of q-colourings of G_i. Then we can write down the number of colourings as a product of ratios.

$$|\Omega(G)| = |\Omega(G_0)| \prod_{i=0}^{m-1} \frac{|\Omega(G_{i+1})|}{|\Omega(G_i)|}$$

Clearly, $|\Omega(G_0)| = q^n$. Therefore, our goal will be finding a good estimate of the ratio

$$r_i = \frac{|\Omega(G_{i+1})|}{|\Omega(G_i)|}$$

for $0 \leq i < m$ in polynomial time.

Suppose that we remove edge(u, v) from G_{i+1} to have G_i. It is not hard to see that $\Omega(G_{i+1}) \subseteq \Omega(G_i)$. By assumption, we have a FPAUS to generate a ε-uniform colouring of G_i, with run time $p(n, \varepsilon)$ for some polynomial p. Therefore we can get an estimation $\widetilde{r_i}$ using the following way.

We sample $s = \lceil 37\varepsilon^{-2}m \rceil$ independent $(\varepsilon/6m)$-uniform copies of colouring of G_i. Let Z_i^k be the random variable that $Z_i^k = 1$ if the k^{th} sample is also a colouring of G_{i+1} and $Z_k = 0$ otherwise. Finally, we return the ratio

$$\widetilde{r_i} = \frac{\sum_{k=1}^{s} Z_i^k}{s}$$

In other word, we return the experimental probability of having a colouring of G_{i+1}. Once we have all $\widetilde{r_i}$, we return $q^n \prod_{i=0}^{m-1} \widetilde{r_i}$ as an approximation of $|\Omega(G)|$.

We have constructed the scheme, the remaining work will be bounding the total run time and proving the scheme will return a good approximation.

Bounding error

We will bound the error in $\widetilde{r_i}$ to bound the error in our approximation. Suppose G_i and G_{i+1} only differ on edge(u, v). Note that every colouring in $\Omega(G_i) \setminus \Omega(G_{i+1})$ has the same colouring for u and v, and can become a colouring in $\Omega(G_{i+1})$ by giving u a colour that its neighbours don't use. There are at least $q - d \geq 2$ possible colours. So we can lower bound r_i,

$$r_i \geq \frac{2}{3}$$

Since we sample a $(\varepsilon/6m)$-uniform colouring for Z_i^k, So,

$$|\operatorname{E}(Z_i^k) - r_i| \leq \frac{\varepsilon}{6m} \leq \frac{\varepsilon}{4m} r_i$$

Then we can bound the expected value of $\widetilde{r_i}$

$$|\operatorname{E}(\widetilde{r_i}) - r_i| \quad \leq \quad \frac{\varepsilon}{4m} r_i \tag{2.9}$$

For variance of $\widetilde{r_i}$,

$$\operatorname{Var}(\widetilde{r_i}) \quad = \quad s^{-1} \operatorname{E}(\widetilde{r_i})(1 - \operatorname{E}(\widetilde{r_i}))$$

Since $1 \geq \operatorname{E}(\widetilde{r_i}) \geq \frac{2}{3}(1 - \frac{\varepsilon}{4m}) \geq 1/2$, we have

$$\frac{\operatorname{Var}(\widetilde{r_i})}{\operatorname{E}(\widetilde{r_i})^2} \leq s^{-1}(\frac{1}{\operatorname{E}(\widetilde{r_i})} - 1) \leq s^{-1}$$

So,

$$\begin{aligned}
\frac{\operatorname{Var}(\prod_{i=0}^{m-1} \widetilde{r_i}))}{\prod_{i=0}^{m-1} \operatorname{E}(\widetilde{r_i})^2} \quad &= \quad \prod_{i=0}^{m-1} (1 + \frac{\operatorname{Var}(\widetilde{r_i})}{\operatorname{E}(\widetilde{r_i})^2}) - 1 \\
&\leq \quad (1 + s^{-1})^m - 1 \\
&\leq \quad \frac{\varepsilon^2}{36}
\end{aligned}$$

Then by Chebyshev's inequality, with probability at least 3/4, we have

$$(1 - \frac{\varepsilon}{3}) \prod_{i=0}^{m-1} \operatorname{E}(\widetilde{r_i}) \leq \prod_{i=0}^{m-1} \widetilde{r_i} \leq (1 + \frac{\varepsilon}{3}) \prod_{i=0}^{m-1} \operatorname{E}(\widetilde{r_i})$$

Also by inequality 2.9, we have

$$(1 - \frac{\varepsilon}{2}) \prod_{i=0}^{m-1} r_i \leq \prod_{i=0}^{m-1} \operatorname{E}(\widetilde{r_i}) \leq (1 + \frac{\varepsilon}{2}) \prod_{i=0}^{m-1} r_i$$

Combining the above two inequalities, we can have an approximate of $\Omega(G)$ with multiplicative error of ε and probability at least 3/4.

Bounding running time

Finally we bound the total run time. Every time we estimate $\widetilde{r_i}$, we call the sampler s times. In each time the sample run in time $p(n, 6m\varepsilon^{-1})$. To estimate all r_i, we call the sampler ms times. The total run time of our approximate scheme will be $O(msp(n, 6m\varepsilon^{-1}))$ which is polynomial of size of G and $1/\varepsilon$. So the scheme also satisfies the run time requirement. □

Therefore, we need to find a way to sample a colouring nearly uniformly. However, this is still an open problem.

2.6.2 Glauber Dynamics

One attempt of constructing a sampler is using the Markov Chain Monte Carlo method. The idea of this method is to construct a Markov Chain with the desired stationary distribution and fast mixing time. Then we run the chain until it mixed and return the current state. The following is a natural Markov Chain for sampling a graph colouring.

For any graph G, let q be the number of colours. Consider a Markov Chain M with state space Ω is the set of proper colourings of G. Let X_t be the state of M at time t. X_{t+1} are found by the following procedure

1. Sample a vertex v and a colour c out of q colours both uniformly and independently.
2. Recolour $X_t(v)$ to c to obtain a new colouring X'.
3. If X' is proper, let $X_{t+1} = X'$. Otherwise let $X_{t+1} = X_t$.

This chain is called "Glauber Dynamics" in the field of statistical physics.

Stationary distribution of Glauber Dynamics

It is not difficult to see the unique stationary distribution of M is the uniform distribution for $q \geq d + 2$. It is aperiodic because there is non-zero probability to stay at current state in each step. For $q \geq d+2$, we can construct a path from X to Y for every pair of colourings X, Y using the following strategy. For every node u such that $X(u) \neq Y(u)$, we first recolour the neighbour of u such that none of them use colour $Y(u)$. This step is always possible because $q \geq d + 2$. Then we can safely recolour $X(u)$ to $Y(u)$. So M is also irreducible and it has a unique stationary distribution. Since the transition matrix is symmetric, the stationary distribution must be uniform.

Conjecture on mixing time

Using Glauber Dynamics, we can construct a sampler for a graph colouring by running the chain for a long enough time. One important question is the mixing time of the Glauber Dynamics. For $q < d+2$, the chain might be frozen at some colourings because we can't change the colour of any node. For example, it happens for $q = |V|$ and G is a complete graph. Therefore we want to find the mixing time for $q \geq d + 2$. There is a conjecture[11] that the chain mixes in polynomial time in this case.

Conjecture 2.28. Let $G = (V, E)$ be a graph with maximum degree d. Let Ω be the set of q-colourings of G. For $q \geq d + 2$, the Glauber Dynamics for sampling a q-colouring of G has mixing time in $O(|V| \log |V|)$

2.6.3 Jerrum Coupling

In the end of this chapter, we present the path coupling version of Jerrum's coupling [11]. Although our proof does not use the coupling method, we present it as many results on the Glauber Dynamics is based on a very careful study of it.

Theorem 2.29. *Let $G = (V, E)$ be a graph with maximum degree d. Let d be the maximum degree of G. Let q be the number of colours. The Glauber Dynamics for sampling a proper q-colouring of G has mixing time in $O(nk \log n)$ if $q > 2d$.*

Proof. Let Ω be the set of proper colourings. We first define an integer-value metric d on Ω. For any pair of colourings X, Y, we let $d(X, Y) = |\{u \in V : X(u) \neq Y(u)\}|$ be the number of nodes where they have different colour in X and Y.

Now we will define the coupling strategy. Let X_t, Y_t be two copies of the Glauber Dynamics at time t. Suppose $d(X_t, Y_t) = 1$. Let $cx = X_t(u)$ and $cy = Y_t(u)$. We first sample a node v and a colour c both uniformly and independently. Then we update both chains using the following strategy.

- If v is a neighbour of u and $c = cx$, we update $X(v)$ to cx and $Y(v)$ to cy.
- If v is a neighbour of u and $c = cy$, we update $X(v)$ to cy and $Y(v)$ to cx.
- Otherwise, we update both $X(v)$ and $Y(v)$ to c.

The updates that will increase distance are the updates in the second case. In this case, we will update v to different colours in both chains. The updates that will decrease distance are the successful update of u. Since there are at most d colours in the neighbours of v, we can update u if we sample the other $q - d$ colours.

So the expected change of distance after one step is

$$\begin{aligned} \mathrm{E}(d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)) &\leq \frac{1}{q|V|}(d - (q - d)) \\ &\leq -\frac{1}{q|V|} \end{aligned}$$

Using the path coupling 2.22, the mixing time is $O(q|V| \log |V|)$. □

Chapter 3

Glauber Dynamics for sampling an edge colouring of regular trees

3.1 Introduction

In this chapter, we study the use of Glauber Dynamics for sampling an edge colouring of a d-regular tree. It is equivalent to sampling a vertex colouring of the line graph of the tree. We show that the chain mixes in polynomial time for $q \geq 2d$.

We use the multicommodity flow argument proposed by Sinclair. To apply the argument we need to construct a flow that send 1 unit of flow between every pair of distinct colourings using Glauber Dynamics transitions. We construct the flow by induction on the tree height. The base case is a special case where the tree is a single edge. The construction of the flow is trivial in this case. Then for a tree with height h, the height of the subtrees of the root is $h-1$. By induction assumption, we know how to send a flow between two colourings if they only disagree at one subtree. We show that is enough to construct the flow by studying another type of Markov Chain called Block Dynamics.

3.2 Preliminary

3.2.1 Edge colouring on tree

Let $T = (V, E)$ be an undirected rooted tree with edge set E. Let d be the maximum degree of T. Let $[q] = \{0, ..., q-1\}$ denote a set of $q > d$ colours where we use an integer to label a colour.

A proper edge q-colouring of a tree is a colouring of edges with element of $[q]$ such that edges sharing same node as an endpoint do not receive the same colour. We will define it as a function $f : E \to [q]$ that maps edges to colours such that if edges e_1 and e_2 shares an endpoint, $f(e_1) \neq f(e_2)$. In the remaining part, we may omit the word "proper", so an edge colouring actually means a proper edge colouring.

3.2.2 Planted tree

In this chapter, we will also study edge colouring on regular planted trees. A planted tree is a rooted tree and the root has degree 1. We use the root-edge to denote the only edge connecting the root of a planted tree.

A d-regular planted tree is a planted tree with the following two property,

- Every non-root and non-leaf vertex has the same degree d.
- The distance from the root to any leaf vertices are the same.

We study the planted tree because we will consider the Glauber Dynamics on subtrees of the root. Then the subtrees is exactly a planted tree with a forbidden set of colours of size $d-1$ for the root-edge. The set of forbidden colours is formed by the set of colours occupied by other edges connecting the root.

3.2.3 Glauber Dynamics for edge colouring

Usually, people use the Glauber Dynamics for sampling a proper vertex colouring of a graph. In this chapter we use Glauber Dynamics for sampling an edge q-colouring for the following two types of trees

- d-regular trees
- d-regular planted trees with a forbidden set of colours of size $d-1$ for the root-edge

Let T be any tree of the above types. Let Ω be the set of edge colourings. The Glauber Dynamics for T is a Markov Chain M with state space Ω. Let X_t be the state of M at time t, the chain does the following transition to find X_{t+1}.

1. Let $Y = X_t$.

2. Sample an edge e uniformly, and a colour c out of the q colours uniformly and independently.
3. Set $Y(e) = c$.
4. If Y is proper, set $X_{t+1} = Y$. Otherwise, set $X_{t+1} = X_t$.

In other words, in each step we sample an edge e and colour c uniformly at random. Then we update the colour of e to c if the resulting colouring is still proper. Otherwise we stay at the current colouring.

3.2.4 Multicommodity flow

The multicommodity flow argument [18] will be our main tool for bounding the mixing time of Glauber Dynamics.

Definition

Let M be a reversible Markov Chain on a state space Ω with a transition matrix P. Suppose the stationary distribution π is uniform. Let $E = \{(X,Y) : P(X,Y) > 0\}$ be the set of transitions of M. For any pair of distinct states X, Y, let $\mathcal{P}_{X,Y}$ be the set of simple paths from X to Y using only transitions in E. Let $\mathcal{P} = \cup_{X \neq Y} \mathcal{P}_{X,Y}$. A multicommodity flow of M is a fractional flow that send 1 unit of flow from X to Y for every pair of distinct X, Y. We denote the flow by a function $f : \mathcal{P} \to \mathbb{R}^+ \cup \{0\}$. And for $X \neq Y$, f needs to satisfy the constraint

$$\sum_{p \in \mathcal{P}_{X,Y}} f(p) = 1$$

The congestion of f is defined as

$$\rho(f) = \max_{(X,Y) \in E} \frac{1}{|\Omega| P(X,Y)} \sum_{p \in \mathcal{P} : (X,Y) \in p} f(p)$$

And the congestion of M is the minimum value over all flow f.

$$\rho(M) = \min_f \rho(f)$$

Let $l(f) = \max_{p : f(p) > 0} |p|$ be the longest path length in f.

The congestion and mixing time of M are related. The following two propositions are proven by Sinclair. Firstly, having a multicommodity flow with low congestion implies small mixing time.

Proposition 3.1. *Suppose M is a reversible ergodic Markov Chain on a state space Ω with a transition matrix P and an uniform stationary distribution π. Let f be a multicommodity flow of M. Let $\tau_{mix} = \tau_{1/4}(M)$. Then,*

$$\tau_{mix}(M) \leq \rho(f)l(f)\ln(2|\Omega|)$$

Secondly, small mixing time implies the existence of a multicommodity flow with low congestion.

Proposition 3.2. *Suppose M is a reversible ergodic Markov Chain on a state space Ω with a transition matrix P and uniform stationary distribution π. Let $\tau_{mix} = \tau_{1/4}(M)$. Then there exists a flow f with $\rho(f) \leq 16\tau_{mix}$ and $l(f) \leq 2\tau_{mix}$.*

3.3 Problem

Now we state the main problem and our result. Consider a d-regular tree T. Let $|T|$ be the number of edges of T. Let $[q]$ be the set of colours. We want to know the mixing time if we use the Glauber Dynamics M to sampling a proper edge q-colouring for T.

Previous results

Let $L(T)$ be the line graph of T, then M is actually the Glauber Dynamics for sampling a vertex colouring of $L(T)$. Note that maximum degree of $L(T)$ is $2d-2$. There is a folklore conjecture [11] that the mixing time of M is $O(|T|\log|T|)$ if $q \geq 2d$. Using Jerrum coupling [11], Jerrum proved the mixing time of M is $O(q|T|\log|T|)$ if $q \geq 4d$. If we use Vigoda's result [19], the mixing time of M is $O(q|T|^2\log|T|)$ if $q \geq 11d/3$. We are not aware of any previous result that can be applied on this special class of graph.

Our results

We show that the mixing time of M is polynomial in $|T|$ if $2d \leq q \leq 4d$. Our result is as follows.

Theorem 3.3. *Let q be the number of colours. Let M be the Glauber Dynamics for sampling an edge colouring of a d-regular tree T. Let h be the height of T. If $4d \geq q \geq 2d$, then*

$$\tau_{mix} \leq 2^{21h} q|T|^8 \ln^{2h}(2q)\ln(2|T|)$$

Note that $2^{21h} = O(poly(|T|))$ since $|T| = d^h$. Our results shows that the Glauber Dynamics for sample an edge colouring of a d-regular tree will mix in polynomial time if the number of colours $q \geq 2d$. The number of colours needed to guarantee polynomial mixing time matches the number mention in the conjecture. However, the mixing time is much larger than the time in the conjecture. It may be possible to get a better time because we did not try to minimize the mixing time.

3.4 Special case

A special case is the tree height $h = 1$. We will first study the Markov Chain for this special case. The result of this special case gives bound on mixing time of Block Dynamics which we will consider in later part. In the special case, we will consider using Glauber Dynamics for sampling an edge colouring for the following two types of trees:

- regular tree with height 1
- regular planted tree with height 2 and a forbidden set of colours of size $d-1$ for the root-edge

We remark that the tree is a star graph in both cases. The only difference is the colour constraint in the second case.

We will prove the following for this special case.

Theorem 3.4. *Let q be the number of colours. Let M be the Glauber Dynamics for sampling an edge colouring of a tree T of height 1. Let d be the number of edges of T. If $q \geq 2d$ and at most one edge of T has a forbidden set of colours of size $d-1$, then* $\tau_{mix} \leq 4d^2q^2ln(2q)$

Notations

We will use the following notations. Let T be the d-regular tree or planted tree that we consider in this section. Let $\{e_i\}_{i=1}^{d}$ be the set of edges of T. If one edge of T has a forbidden set of colours of size $d-1$, we let e_1 to denote that edge. For other edges, we label the edges arbitrarily. Let M be the Glauber Dynamics for sampling a q colouring of T. We use P_d^q to denote $\frac{q!}{(q-d)!}$.

Number of colouring

First we lower bound the number of proper edge colourings.

Lemma 3.5. *The number of edge q-colourings for T is at least* $\frac{q-d+1}{q}\ P_d^q$.

Proof. There will be 2 cases.

Case 1: e_1 can choose any colours. In this case, the number of colourings is P^q_d.

Case 2: There is a set C where e_1 cannot choose colour from C, and $|C| = d-1$. We would choose colour for e_i from $i = 1$ to d one by one. There are $d-1$ colours that e_1 cannot choose from. The number of choices for e_1 is $q-d+1$. And for $i > 1$, the number of choice for e_i is $q-i+1$. The total number of colourings is $(q-d+1)\prod_{i=2}^{d}(q-i+1) = \frac{q-d+1}{q} P^q_d$.

Combining both cases, the number of colourings is at least $\frac{q-d+1}{q} P^q_d$. □

Proof technique

We use the canonical path argument to bound the mixing time. Canonical path argument is a special case of multicommodity flow where we send the flow from X to Y on a single path only.

For every pair of colourings X, Y, we will construct a path $\mathcal{P}_{X,Y}$ from X to Y using transition of M. Then we can construct a multicommodity flow f where f send 1 unit of flow from X to Y on path $\mathcal{P}_{X,Y}$ for every pair of distinct colourings X, Y.

We will upper bound $\rho(f)$ and $l(f)$ to our desired quantity in the lemma 3.6 which we will prove later.

Lemma 3.6. *There exists a multicommodity flow f of M such that $\rho(f) \le \frac{2dq^2}{q-d+1}$ and $l(f) \le 2d$.*

Then we will apply the proposition 3.1 to obtain an upper bound on mixing time.

Proof of Theorem 3.4. By lemma 3.6, We can construct a multicommodity flow f on the state space of the Glauber Dynamics. The congestion $\rho(f)$ is bounded by $\frac{2dq^2}{q-d+1}$ and $l(f)$ is bounded by $2d$. Then by proposition 3.1, $\tau_{mix} \le \frac{4d^2q^2}{q-d+1}\ln(4|\Omega|)$. Note that $|\Omega| \le q^d$ and $q-d+1 > d$ since $q \ge 2d$, we have $\tau_{mix} \le 4d^2q^2 ln(2q)$.

□

3.4.1 Canonical Path

In this section, we will define a path from X to Y for any pair of colourings X, Y

The path consists of d rounds. The goal of round i is to recolour e_i to $Y(e_i)$. There will be two possible cases in round i. The first case is that no other e_j is using the colour $Y(e_i)$. The second case is that some e_j is using the colour $Y(e_i)$.

For the first case, the path recolour e_i to $Y(e_i)$ directly, or do nothing if e_i already have colour $Y(e_i)$.

For the second case, the path first choose a colour c and change the colour of e_j to c. Then the path change the colour of e_i to $Y(e_i)$. c is chosen using the following way. Let X' be the colouring at beginning of this round. Let S be the set of unused colour in X'. The path choose the largest c such that $c \in S$ and $c < X'(e_i)$. If every colour in S is larger than $X'(e_i)$, the path choose the largest colour in S to be c.

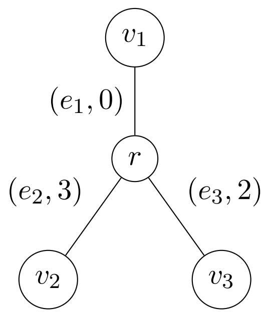
The following is the pseudo code.

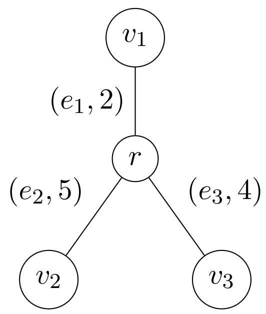
for *i = 1 to d* **do**
    if *there is e_j has colour $Y(e_i)$* **then**
        let S be the set of unused colour
        if *there is $a \in S$ and $a < X(e_i)$* **then**
            let $c = \max(\{a \in S : a < X(e_i)\})$
        end
        else
            let $c = \max(S)$
        end
        update $X(e_j)$ to c
    end
    update $X(e_i)$ to $Y(e_i)$
end

Algorithm 1: Recolour(X, Y)

Example

We will show one example of the path here. Suppose we have 6 colours.

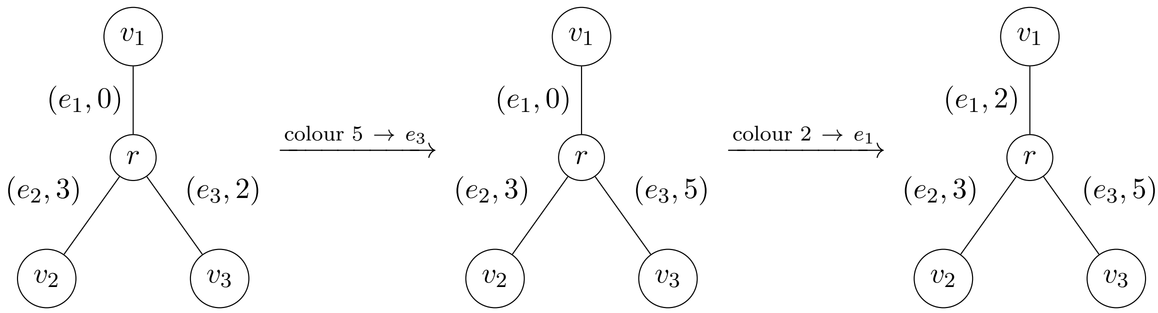
Let $X =$ 

$, Y =$ 

For each edge, we label it with (edge id, edge colour). For example, $(e_2, 3)$ means edge e_2 has colour 3 in X. We will show an example that go from X to Y.

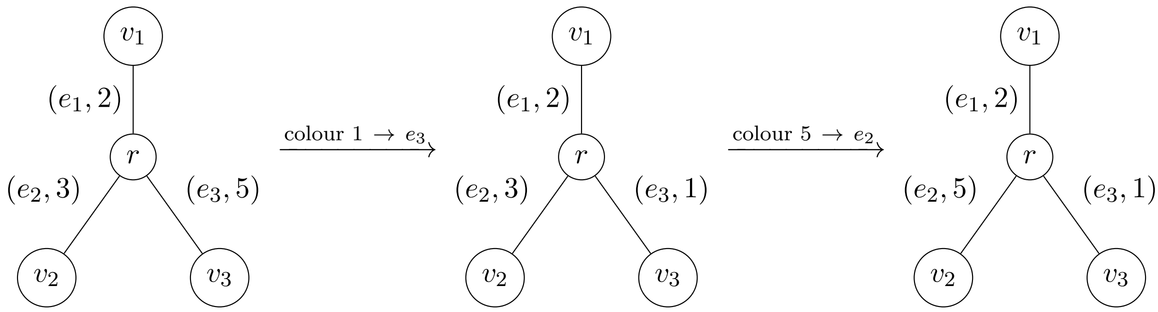
Round 1

In this round we want to change colour of e_1 to 2. However, e_3 is blocking it. We need to recolour e_3 to a colour that is the largest unused colour and smaller than current colour of e_1, which is 0. Since all unused colour is larger than 0, we change to recolour e_3 to largest unused colour, which is 5. Then we can recolour e_1 to 2.



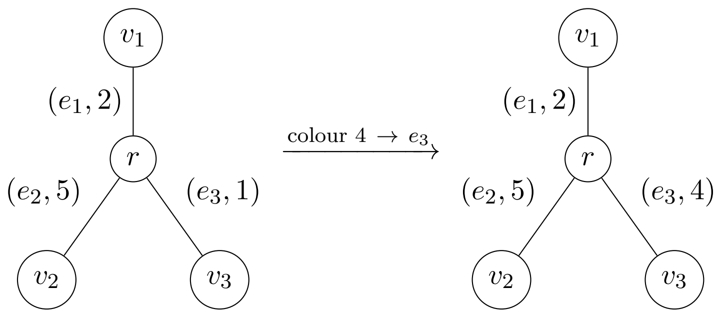
Round 2

Again we need to free the colour from e_3. The largest unused colour that is smaller than colour of e_2 is 1. So we first change colour of e_3 to 1. Then change colour of e_2 to 5.



Round 3

Finally, we change colour of e_3 to 4.



3.4.2 Analysis of canonical path

In this section, we analyse the canonical path to bound the congestion in lemma 3.6. For any colourings X, Y, let $\mathcal{P}_{X,Y}$ denote the path generated by *Recolour*(X, Y). Let $\mathcal{P}_{X,Y}(i)$ denote the colouring after round i. Also let $\mathcal{P}_{X,Y}(0)$ denote the start colouring X. So the path is like the following

$$X = \mathcal{P}_{X,Y}(0) \xrightarrow{\text{round } 1} \mathcal{P}_{X,Y}(1) ... \xrightarrow{\text{round } i} \mathcal{P}_{X,Y}(i) \xrightarrow{\text{round } i+1} ... \xrightarrow{\text{round } d} \mathcal{P}_{X,Y}(d) = Y$$

We need the following lemma to prove lemma 3.6.

Lemma 3.7. *For any transition t, the number of pairs X, Y such that $t \in \mathcal{P}_{X,Y}$ is at most* $2\ P_d^q$.

Proof of lemma 3.6. We construct the multicommodity flow f by sending 1 unit of flow along $\mathcal{P}_{X,Y}$ for every distinct X, Y. For any X, Y, the length of $\mathcal{P}_{X,Y}$ is at most $2d$, so $l(f) \leq 2d$.

By definition,

$$\rho(f) = \max_{(A,B)\in E} \frac{1}{|\Omega| P(A,B)} \sum_{(X,Y):(A,B)\in\mathcal{P}_{X,Y}} 1$$

Since $\Omega \geq \frac{q-d+1}{q} P^q_d$ and $P(A,B) = 1/dq$,

$$\begin{aligned}
&\leq \frac{dq^2}{(q-d+1)\ P^q_d} \max_{A,B} |\{(X,Y) : (A,B) \in \mathcal{P}_{X,Y}\}| \\
&\leq \frac{dq^2}{(q-d+1)\ P^q_d} (2\ P^q_d) \text{ by lemma 3.7} \\
&\leq \frac{2dq^2}{q-d+1}
\end{aligned}$$

□

To proof lemma 3.7, we will need the following two lemma.

Lemma 3.8. *Suppose t is a transition that change colour of edge e_j, the number of X such that there exists Y and $\mathcal{P}_{X,Y}$ use t in round k is at most* $\begin{cases} P^q_{k-1} & \text{if } j > k \\ (d-k+1)\ P^q_{k-1} & \text{if } j = k \end{cases}$

Lemma 3.9. *Suppose t is a transition that change colour of edge e_j, the number of Y such that there exists X and $\mathcal{P}_{X,Y}$ use t in round k is at most P^{q-k}_{d-k}*

Proof of Lemma 3.7. Suppose t is a transition that change colour of edge e_j, then it can only appear in first j rounds.

By multiplying the results in lemma 3.8 and lemma 3.9, the number of colourings X, Y such that $\mathcal{P}_{X,Y}$ use t in round k is at most $\begin{cases} \frac{1}{q-k+1}\ P^q_d & \text{if } j > k \\ \frac{d-k+1}{q-k+1}\ P^q_d & \text{if } j = k \end{cases}$

So,

$$\begin{aligned}
\text{number of } X, Y \text{ s.t. } t \in \mathcal{P}_{X,Y} &= \sum_{1\leq i\leq j} \text{number of } X, Y \text{ s.t. } \mathcal{P}_{X,Y} \text{ use } t \text{ in round } i \\
&\leq \sum_{1\leq i<j} \frac{1}{q-k+1}\ P^q_d + \frac{d-k+1}{q-k+1}\ P^q_d \\
&\leq P^q_d + P^q_d \\
&\leq 2\ P^q_d
\end{aligned}$$

□

For the remaining lemma 3.8 and 3.9, we will prove lemma 3.9 first since it is easier.

Proof of Lemma 3.9. For any X, Y such that $\mathcal{P}_{X,Y}$ use t in round k, we know what $Y(e_i)$ is for $i \le k$. For the remaining $d-k$ edges, there are $q-k$ colours for them. So the number of Y is at most P_{d-k}^{q-k}. □

To prove lemma 3.8, we need the following lemma.

Lemma 3.10. *For any colouring B, if $B = \mathcal{P}_{X,Y}(i)$ for some unknown X, Y and $i > 0$, then the number of possible $\mathcal{P}_{X,Y}(i-1)$ is at most $q-i+1$.*

Proof of Lemma 3.10. For any path $\mathcal{P}_{X,Y}$, in round i it recolour at most 2 edges. So there will be three cases.

If it didn't recolour any edges, then $\mathcal{P}_{X,Y}(i-1) = \mathcal{P}_{X,Y}(i)$. In this case, the number of $\mathcal{P}_{X,Y}(i-1)$ is 1.

If it recolour only 1 edge, then this edge must be e_i. In this case, the old colour of e_i must be in the unused colour of $\mathcal{P}_{X,Y}(i)$, which has $q-d$ possibilities. So the number of possible $\mathcal{P}_{X,Y}(i-1)$ is $q-d$.

If it recolour 2 edges, the first edge must be e_j for some $j > i$ and the second edge must be e_i. Suppose we know what j and $\mathcal{P}_{X,Y}(i)$ are, we can recover $\mathcal{P}_{X,Y}(i-1)$. It is because $\mathcal{P}_{X,Y}(i-1)(e_j) = \mathcal{P}_{X,Y}(i)(e_i)$. For $\mathcal{P}_{X,Y}(i-1)(e_i)$, let S be the set of unused colour of $\mathcal{P}_{X,Y}(i)$. If every colour in S is smaller than $\mathcal{P}_{X,Y}(i)(e_j)$, $\mathcal{P}_{X,Y}(e_i)$ is the smallest in S. Otherwise $\mathcal{P}_{X,Y}(i-1)(e_i)$ is the smallest colour in S that is greater than $\mathcal{P}_{X,Y}(i)(e_j)$. So the number of possible $\mathcal{P}_{X,Y}(i-1) = d-i$.

Summing for three cases, the number of possible $\mathcal{P}_{X,Y}(i-1)$ is at most $q-i+1$. □

Finally we prove lemma 3.8.

Proof of Lemma 3.8. Let $t = (t_{start}, t_{end})$ be any transition that change colour of e_j. We first count the size of set $S = \{A : \exists X, Y, t \in \mathcal{P}_{X,Y} \text{ and } \mathcal{P}_{X,Y}(k-1) = A\}$.

If $j > k$, then in round k the path will take 2 transitions and t is the first transition. Since t is the first transition, $\mathcal{P}_{X,Y}(k-1) = t_{start}$ and $|S| = 1$.

If $j = k$, there are 2 possibilities, then t is either the first transition or the second transition. Again, if t is the first transition, then $|S| = 1$. If t is the second transition, the first transition is freeing colour $Y(e_k)$ from some e_i and t is the transition that recolour e_k to $Y(e_k)$. If we know what e_i is, then we can recover $\mathcal{P}_{X,Y}(k-1)$ by giving back e_i the colour $Y(e_k)$. So $|S| = d - k$ which is the number of possible e_i.

By repeatedly applying lemma 3.10, for any colouring A, the number of X such that there exists Y and $\mathcal{P}_{X,Y}(k) = A$ is at most P_k^q. So the number of X such that there exists Y and $\mathcal{P}_{X,Y}$ use t in round k is at most $P_k^q|S| = \begin{cases} P_{k-1}^q & \text{if } j > k \\ (d-k+1)\ P_{k-1}^q & \text{if } j = k \end{cases}$. □

3.5 General case

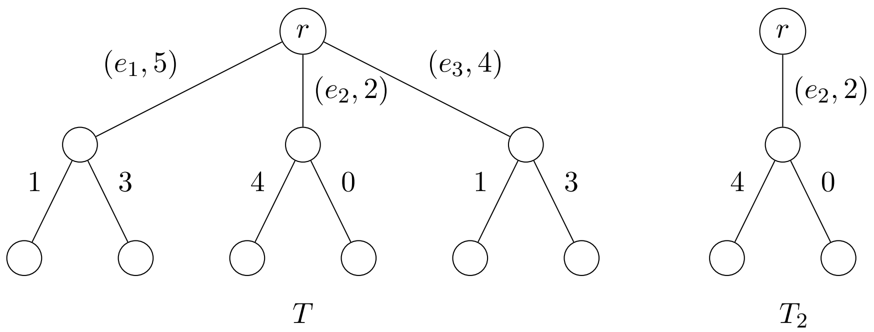
Now we will consider the general cases where the tree can have any height. We first introduce some notations.

Notation of tree

We will use the following notation for the d-regular tree / planted tree we considered. Let T be the tree.

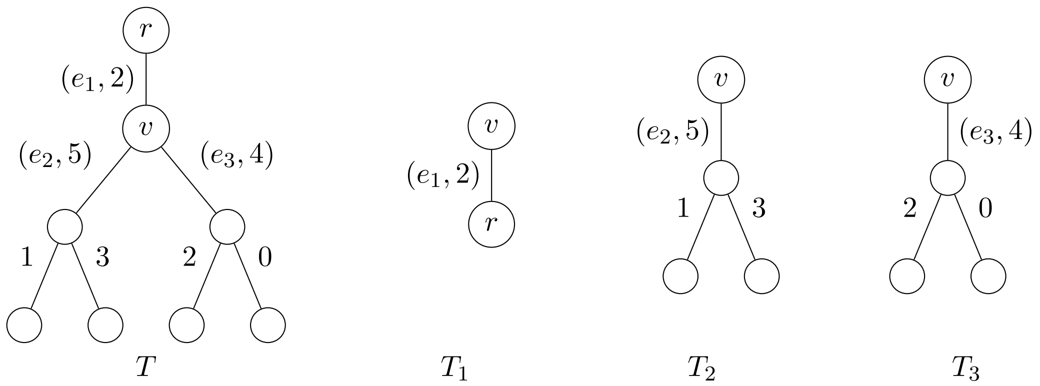
***d*-regular tree**

If T is a d-regular tree, we use e_i for $1 \leq i \leq d$ to denote the d edges connecting the root. We don't label other edges. We use T_i to label the planted tree with the same root of T and e_i begin the root edge. We put a number near every edge to denote the colour of the edge. The following is an example of a 3-regular tree of height 2. The right side is T_2 in this example.



***d*-regular planted tree**

If T is a d-regular planted tree, we use e_1 to denote the root-edge. For the $d-1$ edges that share an endpoint with e_1, we use e_i for $2 \leq i \leq d$ to label them. Let v be the only child of the root. We use T_i to denote the planted tree rooted at v with e_i being the rooted edge. The following is an example of a 3-regular planted tree T and T_1, T_2, T_3 of T.



3.5.1 Block Dynamics

In this section we introduce and analyse another type of Markov Chain which we call it Block Dynamics. It is another Markov Chain for sampling a colouring. Two previous results [2, 14] about mixing time of Glauber Dynamics for sampling a vertex colouring of a tree also use Block Dynamics in their analysis. In our setting of sampling an edge colouring, each Block Dynamics transition (X, Y) is a pair of colourings where X and Y disagree at one subtree only.

Our main objective is not about the mixing time of Block Dynamics. However we study them because it tells us how to construct a flow for Glauber Dynamics. Recall that fast mixing implies a flow with good congestion. It tells us how to send the flow between distinct pair of colourings if we know how to send a flow between two colourings that only disagree at one subtree.

Definition

Let T be tree rooted at some node r. Let q be the number of colours. Let Ω be the set of edge q-colourings of T. A Block Dynamics is a Markov Chain with state space Ω.

In each step the chain does the following to obtain the next state. The chain first samples a colour c and a planted tree T_k from $\{T_i\}_{i=1}^d$ randomly and independently. Then the chain samples a random edge colouring X for T_k with restriction that e_k must use colour c. If there is no conflict, the chain updates the colouring of T_k to X. Otherwise it stays at current colouring.

Clearly, the transition matrix of M is symmetric. So the stationary distribution is also uniform.

3.5.2 Mixing time of Block Dynamics

In this section we analyse the mixing time of Block Dynamics.

Lemma 3.11. *Let q be the number of colours. Let M be the Block Dynamics for sampling an edge q-colouring of a tree T. If T is a d-regular tree or d-regular planted tree, then $\tau_{mix}(M) \leq 16d^2q^2\ln(2q)$.*

Proof. Let S denote the set of d edges $\{e_i\}_{i=1}^{d}$. For any colourings X, let $X(S)$ denotes the colouring of S. Let X_t, Y_t be two copies of the Block Dynamics.

Our proof is based on coupling, we first couple the colouring on S. Once both chains have the same colouring on S, we can couple the colouring of the other parts easily.

We first make the following two claims and we will prove it later.

Claim 3.12. There exists a coupling strategy such that $\Pr[X_t(S) = Y_t(S)] \geq 4/5$ for $t \geq 14d^2q^2\ln(2q)$.

Claim 3.13. If $X_0(S) = Y_0(S)$, there exists a coupling strategy such that $\Pr[X_t = Y_t] \geq 31/32$ for $t \geq 2d^2q^2\ln(2q)$.

Let $t_1 = 14d^2q^2\ln(2q), t_2 = 2d^2q^2\ln(2q)$. We first apply the coupling in claim 3.12 for t_1 steps. If it succeeds, which means the colourings on S are same for both chains, then we apply the coupling in claim 3.13 for t_2 steps. By the two claims, $\Pr(X_{t_1+t_2} = Y_{t_1+t_2}) \geq \frac{4}{5} \cdot \frac{31}{32} \geq 3/4$. Then this lemma is true by the coupling lemma.

Now we prove claim 3.12. Let $X'_t = X_t(S)$ and $Y'_t = Y_t(S)$. Then X'_t and Y'_t are the Glauber Dynamics for sampling an edge colouring of the star graph formed by S. Let $T = 3.5\tau_{mix}(X'_t) = 14d^2q^2\ln(2q)$. Let u, v be the probability distribution of X'_T and Y'_T. By 2.7, $\|u - \pi\|_{TV} \leq \frac{1}{10}$ and $\|v - \pi\|_{TV} \leq \frac{1}{10}$. So $\|u - v\|_{TV} \leq \frac{1}{5}$.

Then we sample $Z = (Z_X, Z_Y)$ such that Z_X and Z_Y are distributed according to u and v and $\Pr(Z_X \neq Z_Y) = \|u - v\|_{TV} \leq \frac{1}{5}$. Then our goal will be constructing a path of length T from X_0 to Z_X and a path of length T from Y_0 to Z_Y.

We first show how to move from X_0 to Z_X. Let $X_0 = p_0, p_1, ..., p_T = Z_X$ denote a path of length T from X_0 to Z_X. We will move from X_0 to Z_X using path p with probability

$$\frac{\prod_{i=0}^{T-1} P(p_i, p_{i+1})}{P^T(X_0, Z_X)}$$

To see that the chain X_t still behaves like the original chain, let P be the transition matrix of the Block Dynamics X_t.

$$\begin{aligned}
\Pr(X_{i+1} = B|X_i = A) &= \frac{\Pr(X_i = A, X_{i+1} = B)}{\Pr(X_i = A)} \\
&= \frac{\sum_{Z'\in\Omega} \Pr(Z_X = Z')\Pr(X_i = A, X_{i+1} = B|Z_X = Z')}{\sum_{Z'\in\Omega} \Pr(Z_X = Z')\Pr(X_i = A|Z_X = Z')} \\
&= \frac{\sum_{Z'\in\Omega} P^T(X_0, Z')\frac{P^i(X_0,A)P(A,B)P^{T-i-1}(B,Z')}{P^T(X_0,Z')}}{\sum_{Z'\in\Omega} P^T(X_0, Z')\frac{P^i(X_0,A)P^{T-i}(A,Z')}{P^T(X_0,Z')}} \\
&= \frac{\sum_{Z'\in\Omega} P^i(X_0, A)P(A, B)P^{T-i-1}(B, Z')}{\sum_{Z'\in\Omega} P^i(X_0, A)P^{T-i}(A, Z')} \\
&= P(A, B)
\end{aligned}$$

Similarly we move from Y_0 to Z_Y. And $\Pr(X_T = Y_T) = \Pr(Z_X = Z_Y) \geq 4/5$.

Finally we prove claim 3.13. If $X_0(S) = Y_0(S)$ and we apply the same update to X_t and Y_t, they will succeed together or fail together. If an update on T_i succeed on both X_t and Y_t, both chains will have the same colouring on T_i. Therefore we let both chains receive the same update until all T_i are updated at least once.

Probability of sampling and updating T_i successfully in one step is at least $\frac{q-d+1}{dq}$. After t steps, probability of some T_i is not updated is at most

$$d(1 - \frac{q-d+1}{dq})^t \leq d\exp(\frac{-t(q-d+1)}{dq})$$

After $q\ln(32d)$ steps, all T_i are updated with probability at least $31/32$. Then the claim is true since $2d^2q^2\ln(2q) \geq q\ln(32d)$. □

We have bounded the mixing time of Block Dynamics. Then we can get the following corollary if we apply proposition 3.2 directly.

Corollary 3.14. *Let q be the number of colours. Let M be the Block Dynamics for sampling an edge q-colouring of a d-regular tree T. There exists a multicommodity flow f of M with $\rho(f) \leq 256d^2q^2\ln(2q)$ and $l(f) \leq 32d^2q^2\ln(2q)$.*

Which equals to following corollary.

Corollary 3.15. *Let q be the number of colours. Let M be the Block Dynamics for sampling an edge q-colouring of a d-regular tree T. Let P be transition matrix of M. Let Ω be the set of q colourings. Let E be the set of transitions of M. For every distinct colourings X, Y, we can construct a set of paths $\Gamma_{X,Y}$ such that every path consists of* *transition of E. Let $\Gamma = \cup_{X,Y:X\neq Y}\Gamma_{X,Y}$. And we can construct a flow $f : \Gamma \to \mathbb{R}^+ \cup \{0\}$ so the following constraints are satisfies.*

1. *For all pair of distinct colourings X, Y, $\sum_{\gamma\in\Gamma_{X,Y}} f(\gamma) = 1$*
2. $\forall (A, B) \in E, \sum_{p:(A,B)\in p} f(p) \leq 256d^2q^2|\Omega|P(A, B)\ln(2q)$
3. $\max_{\gamma\in\Gamma:f(\gamma)>0} |\gamma| \leq 32d^2q^2 \ln(2q)$

3.5.3 Mixing time of general case

Finally we bound the mixing time of the Glauber Dynamics for sampling an edge q-colouring of d-regular tree T of height h.

Proof technique

We use the multicommodity flow argument to bound the mixing time.

We will construct a flow f that send 1 unit of flow from X to Y for every pair of distinct colourings X, Y. Suppose the following lemma is true.

Lemma 3.16. *There exists a multicommodity flow f that send 1 unit of flow from X to Y for every distinct X, Y. And f satisfies the following constraints,*

1. *The amount of flow that a transition transfer is at most $256^hd^{2h}q^{2h}\frac{|\Omega|}{q|T|}\ln^h(2q)$ and*
2. *The longest path length is at most $32^hd^{2h}q^{2h}\ln^h(2q)$.*

Then we can prove theorem 3.3 by applying proposition 3.1 on lemma 3.16 directly.

Proof of theorem 3.3. Let q be the number of colours. Let M be the Glauber Dynamics for sampling an edge colouring for a d-regular tree T. Let E be the transition of M.

By lemma 3.16, we can construct a multicommodity flow f that satisfies following

$$\rho(f) = \max_{t\in E}\frac{q|T|}{|\Omega|}\sum_{p:t\in p} f(p) \leq 256^hd^{2h}q^{2h}\ln(2q) \tag{3.1}$$

and

$$l(f) \leq 32^hd^{2h}q^{2h}\ln(2q) \tag{3.2}$$

By proposition 3.1,

$$\begin{aligned}\tau_{mix}(M) &\leq \rho(f)l(f)\ln(2|T|^{q}) \\ &\leq 2^{13h}d^{4h}q^{4h}\ln^{2h}(2q)(q\ln(2|T|)) \text{ By 3.1 and 3.2} \\ &\leq 2^{21h}q|T|^{8}\ln^{2h}(2q)\ln(2|T|) \text{ Since } |T|\leq d^{h} \text{ and } q\leq 4d\end{aligned}$$

□

In the remaining part, we prove lemma 3.16.

Proof of lemma 3.16. We will proof the lemma by induction on the tree height h.

The base case is $h = 0$ where the tree is a single edge. In this case, for every distinct colourings X, Y, (X, Y) is a Glauber Dynamics transition. So we send the flow from X to Y directly. Then, $l(f) = 1$ and $\rho(f) = 1$ so the lemma is true in this case.

For any $h > 0$, suppose the lemma is true if the tree height is at most $h - 1$. Consider the case where the height is h.

We first give some notations.

Notation Let q be the number of colours. Let M be the Glauber Dynamics for sampling an edge colouring for T with transition P. Let M_B be the Block Dynamics for sampling an edge colouring for T with transition P_B. Note that both M and M_B have the same state space Ω and uniform stationary distribution π. Let E_B be the set of transitions of M_B. Let E_B^i be the set of transitions of M_B that change colouring of T_i. Similarly let E be the set of transitions of M and E^i be the set of transitions of M that change colour of some edge in T_i. For any T_i, let N_i be the number of colourings of T_i if we fix the colour of e_i.

We will construct the flow in two steps.

Constructing the flow using Block Dynamics transition

By corollary 3.15, we can construct a flow f_B that send 1 unit of flow from X to Y for every distinct X, Y. f_B will send the flow using Block Dynamics transition. And f_B satisfies the following constraints

$$\forall i\ \forall (A,B)\in E_B^i\ ,\quad \sum_{p:(A,B)\in p} f_B(p) \leq 256d^2q^2|\Omega|P_B(A,B)\ln(2q) \tag{3.3}$$

where

$$P_B(A,B) = \frac{1}{qdN_i} \text{ if } (A,B) \in E_B^i$$

and

$$l(f_B) \leq 32d^2q^2\ln(2q) \tag{3.4}$$

The first inequality tells us the maximum amount of flow a Block Dynamics transition will transfer. The second inequality tells us the maximum length of f_B.

Constructing Block Dynamics transitions using Glauber Dynamics transitions

Then we construct a flow that send 1 unit of flow from A to B for every $(A,B) \in E_B$. By definition, every Block Dynamics transition change the colouring of T_i for some i. Note that T_i is either a d-regular planted tree or d-regular tree of height at most $h-1$. So by induction assumption, we can construct a flow f' that 1 unit of flow from A to B for every $(A,B) \in E_B$ using only Glauber Dynamics transition. And f' satisfies the following constraints.

$$\forall i \ \forall (C,D) \in E^i, \sum_{p:(C,D)\in p} f(p) \leq 256^{h-1}d^{2h-2}q^{2h-2}\frac{qN_i}{q|T_i|}\ln^{h-1}(2q) \tag{3.5}$$

and

$$l(f') \leq 32^{h-1}d^{2h-2}q^{2h-2}\ln^{h-1}(2q) \tag{3.6}$$

Again, the first one bounds the amount of flow that a Block Dynamics transition will transfer and the second one bounds the length.

Then for every Block Dynamics transition (A,B) in f_B, we send the flow from A to B using the flow f'. So we have a new flow f that send 1 unit of flow from X to Y for distinct colourings X,Y using Glauber Dynamics transition.

In f, for any $t \in E$, the maximum amount of flow t will transfer is the amount of flow a Block Dynamics transition will transfer in f_B multiplies the flow t will transfer when we use t to construct Block Dynamics transition. For any any T_i,

$$\begin{aligned}\max_{t\in E^i}\sum_{p:t\in p} f(p) &\leq \left(\max_{(A,B)\in E_B^i}\sum_{p:(A,B)\in p} f_B(p)\right)\left(\max_{(C,D)\in E^i}\sum_{p:(C,D)\in p} f'(p)\right) \\ &\leq 256^h d^{2h}q^{2h}\frac{|\Omega|}{q|T|}\ln(2q) \text{ By 3.3 and 3.5}\end{aligned}$$

By 3.4 and 3.6, we can bound the length of f,

$$l(f) \leq l(f_B)l(f') \leq 32^h d^{2h} q^{2h} \ln(2q)$$

So the lemma is true in this case. By induction the lemma is true.

□

Bibliography

[1] Noga Alon and Vitali D Milman. $\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[2] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

[3] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 223–231. IEEE, 1997.

[4] Martin Dyer and Alan Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures & Algorithms*, 23(2):167–179, 2003.

[5] Martin Dyer, Alan Frieze, Thomas P Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 582–589. IEEE, 2004.

[6] Alan Frieze and Eric Vigoda. A survey on the use of markov chains to randomly sample colourings. *Oxford Lecture Series in Mathematics and its Applications*, 34:53, 2007.

[7] Leslie Ann Goldberg, Mark Jerrum, and Marek Karpinski. The mixing time of glauber dynamics for coloring regular trees. *Random Structures & Algorithms*, 36(4):464–476, 2010.

[8] Venkatesan Guruswami. Rapidly mixing markov chains: a comparison of techniques (a survey). *arXiv preprint arXiv:1603.01512*, 2016.

[9] Thomas P Hayes, Juan C Vera, and Eric Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 450–458. ACM, 2007.

[10] Thomas P Hayes and Eric Vigoda. A non-markovian coupling for randomly sampling colorings. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 618–627. IEEE, 2003.

[11] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–166, 1995.

[12] Lap Chi Lau and Michael Molloy. Randomly colouring graphs with girth five and large maximum degree. In *LATIN 2006: Theoretical Informatics*, pages 665–676. Springer, 2006.

[13] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[14] Brendan Lucier, Michael Molloy, and Yuval Peres. The glauber dynamics for colourings of bounded degree trees. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 631–645. Springer, 2009.

[15] Fabio Martinelli. *Lectures on Glauber dynamics for discrete spin models*. Springer, 1999.

[16] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[17] Oskar Perron. Zur theorie der matrices. *Mathematische Annalen*, 64(2):248–263, 1907.

[18] Alistair Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, probability and Computing*, 1(04):351–370, 1992.

[19] Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, 2000.