Color me converged: Surveying and experimenting on the Glauber dynamics for sampling k-colorings

Harvard CS 229R Final Project

Noah Singer and Grace Tian

December 20, 2020

Abstract

We explore the Glauber dynamics for approximately sampling k-colorings on an undirected graph. We begin with a survey of past results. We describe coupling techniques for bounding mixing time and use them to prove Jerrum's result [Jer95] that the Glauber dynamics mix rapidly on graphs of max-degree Δ such that $k/\Delta > 2$. We also outline results on rapid mixing of the flip dynamics [Vig00; Che+19] which yield better bounds for the Glauber dynamics and the use of methods from high-dimensional expanders in proving polynomial-time mixing of the Glauber dynamics in triangle-free graphs [Che+20]. Next, we experimentally investigate the Glauber dynamics via direct simulation on small input graphs, measure convergence time in total-variation distance, and approximate the spectral gap using the power method. We conclude that our experimental results are mostly consistent with the earlier-described theoretical bounds, and make two conjectures as to the monotonicity and limiting behavior of the mixing-time and spectral gap of the Glauber dynamics for a fixed graph as the number of colors grows.

1 Introduction

A standard greedy argument shows that a graph G with max-degree Δ is $(\Delta+1)$-colorable; a corresponding coloring may be found in linear time. *Counting* the number of such colorings, on the other hand, may be much more difficult. This problem is denoted #k-Colorings; it is complete for the class #**P**, and so it is highly unlikely for there to exist an efficient, exact algorithm computing it. Jerrum, Valiant, and Vazirani [JVV86] proved that for a certain wide class of #**P** problems, the *self-reducible* problems, the problem of approximately counting solutions reduces to and from the problem of *approximately sampling* solutions. #k-Colorings is such a problem, so to show that it is efficiently approximable, it is necessary and sufficient to show that there is an efficient algorithm that, given a graph G, samples a k-coloring from G from a distribution which is arbitrarily close (in total-variation distance) to the uniform distribution.

One beautiful and very successful method for developing approximate sampling algorithms for #**P** problems is the *Markov chain Monte Carlo (MCMC)* method. MCMC sampling algorithms for #**P** problems sample from a Markov chain whose state space is the space of all possible solutions and whose stationary distribution is the uniform distribution over solutions. Moreover, it is necessary that this Markov chain mix quickly (i.e., in polynomial time) as a function of n, the size of the input instance. Typically, such Markov chains are easy to define but very large (e.g. colorings in large, sparse graphs), so the mixing-time analyses can be very sophisticated. MCMC arguments have led to efficient, arbitrarily-precise approximation algorithms for various #**P**-complete problems, including matrix permanents (a.k.a. counting perfect matchings) [JSV04] and counting spanning forests [Ana+19].

In this paper, we study the potential polynomial-time mixing of MCMC algorithms for #k-Colorings. In particular, we consider the following Markov chain. Given a k-coloring x of G, where $x(v)$ denotes the color of a vertex v:

1. Randomly sample a vertex v from V and a color c from $\{1, \dots, k\}$.

2. If for each neighbor w of v, $x(w) \neq c$:

   (a) $x(v) \leftarrow c$.

The #k-Colorings sampling problem goes by another name in statistical physics, the *(anti-ferromagnetic) Potts model*, which consists of a system of particles which are assigned one of a particular set of k states, and the Hamiltonian penalizes adjacent particles which are assigned the same state. In the statistical physics setting, the goal is to understand whether and how quickly the system will reach thermodynamic equilibrium at a different temperature (wherein at higher temperatures, the particles' states are perturbed more often). Statistical physicists call the above Markov chain the *Glauber dynamics* for the Potts model, a term which denotes a Markov chain which simulates a physical model by making local, particle-specific updates.

If $k \leq \Delta + 1$, the Glauber dynamics may fail to converge to a stationary distribution; there may be k-colorings if $k \leq \Delta$, and in the case $k = \Delta + 1$, Lubin and Sokal [LS93] showed that the Glauber dynamics may still fail to be ergodic, and hence not converge to a stationary distribution. However, the Glauber dynamics is ergodic when $k \geq \Delta + 2$, leading naturally to the hope that it mixes rapidly in all such cases:

Conjecture 1. *Suppose that $k \geq \Delta + 2$. Then the Glauber dynamics for #k-Colorings mixes in* poly(n) *time.*

This would imply a fully polynomial-time randomized approximation algorithm for #k-Colorings as long as $k \geq \Delta + 2$.

Jerrum [Jer95], as well as Salas and Sokal [SS97], showed that the Markov chain mixes in polynomial time when $k > 2\Delta$; this analysis was subsequently simplified by Bubley and Dwyer [BD97] using a method called *path coupling*. Vigoda [Vig00] proved polynomial-time mixing for $k > \frac{11}{6}\Delta$. Recently, Chen, Delcourt, Moitra, Perarnou, and Postle [Che+19] proved polynomial-time mixing for $k \geq \left(\frac{11}{6} - \eta\right)\Delta$ for some small constant η, while also showing that one-step couplings analyzed as contractions in the Hamming distance — the type of proof used by Bubley and Dyer [BD97] and Vigoda [Vig00] — cannot show that the Glauber and other related dynamics mix quickly for $k < \frac{11}{6}$.

Organization and contributions

In this paper, we take two complementary approaches towards studying Conjecture 1 and the Glauber dynamics of colorings in general. Firstly, in §2, we review the known theoretical bounds; we describe in detail path-coupling methods and the path-coupling argument which gives efficient mixing for $k > 2\Delta$, and survey other bounds. Secondly, we implement an experimental analysis of the convergence of the Glauber dynamics in terms of total-variation distance and the spectral gap, produce various plots, and draw various conclusions about mixing in §3. We conclude with some observations in §4, and finally, include proofs of some theorems about couplings in Appendix A.

2 Theoretical bounds on mixing

In this section, we will show that the Glauber dynamics for #k-Colorings mixes quickly when $k > 2\Delta$, following the treatments of Frieze and Vigoda [FV05] and Moore and Mertens [MM11]. We will then discuss recent improvements to this bound and conclude by comparing the coupling method to spectral methods for the analysis of Markov chain convergence.

2.1 Preliminaries

A *(proper) k-coloring* of an undirected graph G is an assignment of a color in $\{1, \ldots, k\}$ to every vertex of G such that the endpoints of every edge in G have different colors. More generally, an *improper coloring* is any assignment of colors to vertices.

The *total variation (TV) distance* between two distributions X, Y on a finite space Ω is

$$\|X - Y\|_{TV} := \frac{1}{2}\|X - Y\|_1 = \frac{1}{2}\sum_{\omega\in\Omega}\left|\Pr_{x\sim X}[\omega = x] - \Pr_{y\sim Y}[\omega = y]\right|.$$

A *(finite) Markov chain* M is a stochastic matrix of transition probabilities associated to a finite state space Ω; the entry corresponding to $(b, a) \in \Omega^2$ gives the probability of transitioning to state b from state a. Given a starting distribution X_0 over Ω, we can define the t-th step distribution $X_t := M^t X_0$. We consider only Markov chains with stationary distributions, i.e., those for which there exists some distribution π on Ω such that for all starting distributions X_0, $\lim_{t\to\infty}\|X_t - \pi\|_{TV} = 0$. This occurs when M is ergodic (such as the Glauber dynamics of k-colorings).

The *mixing time* (to within ϵ TV-distance) of a Markov chain M with stationary distribution π is

$$t_{mix}(\epsilon) := \inf\{t > 0 : \forall \text{ distributions } X_0, \|X_t - \pi\|_{TV} \le \epsilon\};$$

that is, it is the smallest number of steps t such that regardless of starting distribution X_0, by the t-th step, the chain will be within TV-distance ϵ of stationarity.

2.2 Couplings

In this section, we define the notion of *couplings* and prove a theorem due to Bubley and Dyer [BD97] that gives a way to show rapid mixing by defining a partial coupling.

2.2.1 Intuition

We begin by providing an intuitive example of what a coupling is and how it can be used to prove fast mixing of a Markov chain. Our particular example is a card-shuffling coupling as presented by Moore and Mertens [MM11, §12.4].

Consider a deck of n uniquely labeled cards (e.g., a card might say "Six of clubs"), and the following shuffling mechanism: Given a deck of cards, randomly select one card, and move it to the top of the deck. After how many iterations of this process can we conclude that the deck is essentially random? (This shuffling algorithm may be viewed as an instance of Markov chain Monte Carlo for sampling randomly from the space of all possible decks.)

Consider the following thought experiment. Begin by placing *two* decks side by side and shuffling both in parallel, but with a "twist" — instead of just picking a random card in each deck and moving each to the top of their respective decks, pick a random *label* (e.g., "Six of clubs") and move that particular card in each deck to the top. If you focused solely on one of the two decks, this process would look identical to the one-deck random shuffle we are trying to analyze. *However*, these two decks are *coupled* in a very strong way: Once a particular card (e.g., "Six of clubs") is brought to the top of each deck, that card will stay in the same position in each deck forever! Hence, once every card has been shuffled, the two decks will be identical. The standard *coupon collectors' problem* argument shows that this happens after an expected $O(n \log n)$ steps.

How does this help us bound the mixing time for this chain? Firstly, note that using Markov's inequality, we can be confident (to any desired level of probability) that the two decks will be identical after $\Theta(n \log n)$ steps. Now switch back to the actual scenario where we are shuffling a single deck. After $\Theta(n \log n)$ steps, from a probabilistic standpoint, whatever particular ordering of cards we see no longer reveals any information about the deck we started with — it could just have likely resulted from shuffling from any other starting deck. Hence, the Markov chain has converged.

Indeed, this coupling has a stronger property that it is *distance-decreasing*: If we consider the Hamming distance between any two particular decks (i.e., the number of positions at which they differ), after one step of the coupling, the Hamming distance will never increase, and it has a decent probability of decreasing (in fact, this probability is proportional to the Hamming distance). Hence this coupling may be viewed as a contraction on the space of all decks which is compatible with the shuffling procedure. This intuitive picture

will be exceedingly useful for us in formally stating and analyzing couplings for the Glauber dynamics for k-colorings below.

2.2.2 Formally defining couplings

In this section, we formally define couplings and state a formal version of the theorem that "Couplings which bring initial states closer together imply fast mixing" (Theorem 5 below), based on Frieze and Vigoda's survey [FV05, §2] but using somewhat different notation. We omit the proofs here for brevity, since they only involve standard ingredients in probability; we include the proofs in Appendix A.

Definition 2 (Coupling)**.** *A* coupling *of a Markov chain M on a state space Ω is a joint Markov chain (X', Y') on state space Ω^2 such that X' and Y' each behave like M individually; that is, for every starting distribution $X_0 \times Y_0$ on Ω^2, if X_1 evolves from X_0 according to M, Y_1 evolves from Y_0 according to M, and (X'_1, Y'_1) evolves from $X_0 \times Y_0$ according to (X', Y'), then for all $\omega \in \Omega$,*

$$\Pr_{x \sim X_1}[x = \omega] = \Pr_{(x',y') \sim (X'_1, Y'_1)}[x' = \omega] \text{ and } \Pr_{y \sim Y_1}[y = \omega] = \Pr_{(x',y') \sim (X'_1, Y'_1)}[y' = \omega].$$

Lemma 3 (Coupling and TV-distance)**.** *Let M be a Markov chain on Ω, and (X', Y') a coupling of M. Let X_0, Y_0 be an arbitrary start distribution. Let $X_1, \ldots, X_t, \ldots$ evolve from X_0 according to M, $Y_1, \ldots, Y_t, \ldots$ evolve from Y_0 according to M, and $(X'_1, Y'_1), \ldots, (X'_t, Y'_t), \ldots$ evolve from $X_0 \times Y_0$ according to (X', Y'). Then for all t,*

$$\|X_t - Y_t\|_{TV} \leq \Pr_{(x',y') \sim (X'_t, Y'_t)}[x' \neq y'].$$

For instance, if by step t, the coupling has entirely *coalesced* — that is, X'_t and Y'_t are identical — then we can conclude that $\|X_t - Y_t\|_{TV} = 0$.

In what follows, fix an integer-valued metric d on Ω. Say that a coupling (X', Y') for M is *δ-distance decreasing* for a pair $(x_0, y_0) \in \Omega^2$ if when (X'_1, Y'_1) evolves from start state (x_0, y_0),

$$\mathbb{E}_{(x',y') \sim (X'_1, Y'_1)}[d(x', y')] \leq (1 - \delta) d(x_0, y_0).$$

Distance-decreasingness is enough to imply fast mixing:

Lemma 4 (Distance-decreasing implies fast mixing)**.** *Let M be a Markov chain. Suppose that there exists a coupling (X', Y') that is δ-distance decreasing for all pairs $(x_0, y_0) \in \Omega^2$. Then M mixes quickly:*

$$t_{mix}(\epsilon) \leq \frac{\log(d_{max}/\epsilon)}{\delta}$$

where $d_{max} := \max_{x,y \in \Omega} d(x, y)$ is the diameter of Ω.

A final preparation that we need is a theorem due to Bubley and Dyer [BD97] which states that it is sufficient to only check for δ-distance-decreasingness on special pairs of states, as long as there are enough of these pairs that it is possible to reach any state from any other state.

Theorem 5 (Path-coupling, [BD97])**.** *Let M be a Markov chain. Let $S \subseteq \Omega \times \Omega$ be a subset such that the graph (Ω, S) is connected (i.e., for any $x, y \in \Omega$, there is a path from x to y using "edges" in S). Suppose that there is a "partial coupling" $(\hat{X}', \hat{Y}')$ for S — that is, $(\hat{X}', \hat{Y}')$ only has transition probabilities defined for S — and that $(\hat{X}', \hat{Y}')$ is δ-distance decreasing for all pairs in S. Then $(\hat{X}', \hat{Y}')$ can be extended to a coupling (X', Y') for all of Ω^2 which is also δ-distance decreasing.*

In the next section, we will take advantage of Theorem 5 to show that the Glauber dynamics for #k-Colorings mix rapidly by defining couplings for pairs of "neighboring" colorings (i.e., colorings which differ only by one vertex).

2.3 Proofs of polynomial-time mixing

In this section, we present proofs of polynomial-time mixing as long as $k > 2\Delta$ (due to Jerrum [Jer95]); we begin with the case $k > 3\Delta$ and show how only a slight modification for $k > 2\Delta$ is necessary. Our proofs are hybrids of the exposition of Frieze and Vigoda [FV05, §3] and Moore and Mertens [MM11, §12.4]. These proofs will come from couplings of the Glauber dynamics for colorings. For convenience, we let the state space Ω of the dynamics be the space of all assignments of colors to vertices, including improper colorings.[1]

2.3.1 $k > 3\Delta$

We'll begin by using a simple coupling to show that as long as $k > 3\Delta$, the Glauber dynamics for #k-Colorings mixes rapidly. Throughout this section, let $\Omega := [k]^n$ be the space of *all* (possibly improper) colorings. Let $d(x, y)$ denote the Hamming distance between colorings x, y — that is, the number of vertices in which they differ.

Theorem 6. *Suppose that $k > 3\Delta$. Then the Glauber dynamic for #k-Colorings mixes within $O(nk \log n)$ steps.*

Proof. Let $S := \{(x, y) \in \Omega^2 : d(x, y) = 1\}$ be the set of (improper) colorings which differ at a single vertex. We will apply Theorem 5; it suffices to consider pairs of colorings $(x, y) \in S$ and show that they can be coupled in a distance-decreasing fashion. Note that S clearly connects Ω (as it includes improper colorings).

Consider the following coupling: Pick a random vertex v and a color c and attempt to set both $x(v)$ and $y(v)$ to c. This is a valid coupling for the Glauber dynamics for #k-Colorings, since it is clearly a Markov chain (it depends only on the current state) and marginally looks like the Glauber dynamics.

Suppose x and y are (improper) colorings that differ at exactly one vertex, v^*. Their Hamming distance is 1. What is the expected change in this distance when we apply our coupling? Let a *good* move (i.e., combination of vertex and new color) be one that decreases the distance, and a *bad* move be one that increases it; let p_{good} and p_{bad} be the respective probabilities of these events. Then since a good move will occur when we pick the vertex v^* and then pick any color that is not one of v^*'s neighbors' colors in either x or y, we have

$$p_{good} \geq \frac{k - \Delta}{kn}.$$

Similarly, a bad move will occur only if we pick a vertex w which is a neighbor of v^* and then set w's color to either $x(v^*)$ or $y(v^*)$; this occurs with probability

$$p_{bad} \leq \frac{2\Delta}{kn}.$$

Hence since $d(x, y) = 1$,

$$\mathbb{E}[d(x_1, y_1)] \leq 1 - \frac{3\Delta - k}{kn}.$$

So as long as $k > 3\Delta$, $\mathbb{E}[d(x_1, y_1)] \leq 1 - \frac{1}{kn}$; hence setting $\epsilon = 1/kn$, Theorem 5 gives us the desired bound. (Note that $d_{max} = n$ since the metric is the Hamming distance and our space includes improper colorings.) □

2.3.2 $k > 2\Delta$

A slight improvement to the above argument will allow us to decrease the probability of a bad move. This proof is due to Jerrum [Jer95].

Theorem 7. *Suppose that $k > 2\Delta$. Then the Glauber dyanmics for #k-Colorings mixes within $O(nk \log n)$ steps.*

[1] Note that once the Glauber dynamics reach a proper coloring, all future colorings will remain proper. Moreover, the Glauber dynamics converge towards proper colorings; the coloring will be (arbitrarily likely to be) proper after $\Theta(n \log n)$ steps by the coupon collector formula and Markov's inequality.

Proof. Let S, d, etc. be as above. We will slightly modify the coupling above: Pick a random vertex v and a color c and attempt to set both $x(v)$ and $y(v)$ to c. *However*, if v happens to be a neighbor of v^*, the neighbor at which x and y differ, *and* $c \in \{x(v^*), y(v^*)\}$, we'll do something different: Flip a coin. If it comes up heads, make no changes. If it comes up tails, set $x'(v) \leftarrow y(v^*)$ and $y'(v) \leftarrow x(v^*)$.[2] The reader can check that this is still a valid coupling (i.e., the marginal distributions still match the Glauber dynamics for $\#k$-Colorings).

Now the only way that a bad move can happen is if v is a neighbor of v^*, $c \in \{x(v^*), y(v^*)\}$, and the coin comes up tails; this happens only with probability

$$p_{bad} \leq \frac{1}{2} \cdot \frac{2\Delta}{kn} = \frac{\Delta}{kn}.$$

A bad move still only increases the distance of the two colorings by 1. Repeating the rest of the above argument, we reach the desired conclusion. □

2.4 Overview of better bounds

Vigoda [Vig00] proved polynomial-time mixing as long as $k > \frac{11}{6}\Delta$. Instead of analyzing the Glauber dynamics directly, Vigoda analyzed the *flip dynamics*, which at each step change the colors of a *cluster* of vertices — specifically, given a coloring x, we pick a random vertex v and color c, and then look at all the vertices which are reachable from v via paths whose colors alternate between c and v's current color $x(v)$, and with a certain probability, we flip the colors of every vertex in the cluster. Vigoda's analysis of the flip dynamics directly parallels the distance-decreasing path-coupling arguments we described above for this new Markov chain, albeit involved a more sophisticated coupling. Vigoda then used a "comparison theorem" due to Diaconis and Saloff-Coste [DS93] to show that polynomial-time mixing of the flip dynamics implies polynomial-time mixing of the Glauber dynamics.

Recently, Chen et al. [Che+19] proved polynomial-time mixing for $k > (\frac{11}{6} - \eta)\Delta$ for some $\eta > 0$. They analyzed the space of all colorings from a linear programming perspective and showed that Vigoda's coupling [Vig00] is optimal (in the sense that it generates the smallest threshold $k > \alpha\Delta$ with $\alpha = \frac{11}{6}$) among all couplings of the flip dynamics and analyze a decrease in the Hamming distance. They then developed two ways to get around this obstacle: One uses *variable-length couplings* which may couple multiple steps of the flip dynamics together, not just a single step at a time, and the other uses a new metric which is not the Hamming metric. This linear programming approach to couplings follows on a long line of related work; for instance, Bubley, Dyer, and Greenhill [BDG98] solved tens of thousands of linear programs to conclude that the Glauber dynamics mix rapidly when $\Delta = 3$ and $k = 5$, and in triangle-free graphs $\Delta = 4$ and $k = 7$.

2.5 Comparing coupling to spectral methods for bounding

It is interesting to compare coupling and spectral methods for bounding the convergence time of Markov chains. Last year, spectral methods related to the analysis of high-dimensional expanders were used by Anari, Liu, Gharan, and Vinzant [Ana+19] to prove polynomial-time mixing for the Glauber dynamics of spanning forests. More recently, Chen, Galanis, Štefankovič, and Vigoda [Che+20] also used methods from high-dimensional expanders to prove that the Glauber dynamics for colorings mixes rapidly in triangle-free graphs when $k > \alpha\Delta$ where $\alpha \approx 1.763 < \frac{11}{6}$ is the solution to $e^{1/\alpha} = \alpha$. A brief overview of their approach: The so-called *spectral independence* methodology of Anari, Liu, and Gharan [ALG20] for showing the Glauber dynamics mixes in polynomial time for sampling some object, it suffices to prove certain expansion bounds on "conditioned" instances of the sampling problem. In the context of colorings, these conditioned instances correspond to sampling colorings once the colors of a subset of vertices are fixed. This is an example of a *list coloring* problem, which is a generalization of the standard coloring problem wherein the goal is to sample colorings where each vertex has a designated subset of the colors it is allowed to have. Chen et al. proved bounds on the *local expansion* of a list-coloring instance, which equals the spectral graph of a certain natural

[2] Note that in this case, the coupling actually sets v's new color differently between x and y.

associated graph.[3] Good bounds on local expansion for all list-coloring problems on G imply the so-called *spectral independence* of the Glauber dynamics for coloring G, following earlier work by Anari, Liu, and Gharan [ALG20]. Using spectral techniques, the spectral gap of a random walk on this graph is algebraically related to the *largest* eigenvalue, which tracks the *influence* that selecting a color for one vertex has on the color for another vertex.[4] Finally, given these reductions, a recursive argument is used to bound (a version of) the aggregate influences of all the vertices, implying the desired theorem.

Regarding their approach, Chen et al. wrote:

> "One further feature of using the spectral independence approach to study colorings is that it avoids many of the technical complications in previous approaches caused by coupling arguments... the key improvement on the running time [(previous results had exponential dependence on Δ and k)] is based on relatively simple combinatorial arguments which are then translated into spectral bounds." — [Che+20]

With regard to coupling, quoting Moore and Mertens [MM11]:

> "The best coupling might consist of a complicated correspondence between the moves we perform in the two copies. This correspondence might not even be one-to-one — a move in one copy could be associated with a probability distribution over moves in the other." — [MM11]

In our opinion, the two approaches (spectral methods and coupling arguments) are rather orthogonal. Spectral methods often facilitate the analysis of highly structured graphs (such as those arising from derandomized squares and zig-zag products, or in the context of MCMC, high-dimensional expanders); coupling arguments have more intuitive appeal but it can be very ad hoc, and it may be difficult to prove optimal bounds using couplings without solving very complicated combinatorial optimization problems.

3 Simulations and experimental results

3.1 Experimental design

We built a C++ tool to analyze the convergence and spectral gap of the Glauber dynamics for k-colorings using the `igraph` C library (`https://igraph.org/c/`) as follows. Given (n, Δ, k), we constructed a random, undirected, connected, Δ-regular graph without self-loops. We denote this graph as G. Our code constructs the Glauber dynamics, the Markov chain whose state space is the set of all the possible graph colorings of G.

In order to measure convergence in total variation distance, after initializing at an arbitrary greedily-picked coloring, we store a distribution over colorings at each step of the chain, updating according to the transition probabilities, and measuring the total variation distance until it hits a certain stopping threshold ε. The total number of steps is then the mixing time of ε.

We also used our code to estimate the spectral gap of the Glauber dynamics using the power method, i.e., the second-smallest eigenvalue ν_2 of the Laplacian of the walk matrix, which equals 1 minus the second-larger eigenvalue of the random-walk matrix.. To do this, we sampled a random vector from $\{-1, 1\}^N$ where N is the total number of colorings of G, and then repeatedly multiplied by the Glauber dynamics' stochastic matrix, while projecting onto $1^\perp$ and renormalizing for numerical stability, and at each step measuring the new ℓ_2-norm. (See [Tre17, §9.3] for the analysis of this algorithm.

Since the Markov chain we studied may have exponentially many states in n, the number of vertices in G, we required substantial optimizations to be able to simulate a nontrivial number of vertices or colors. Each coloring was represented as an integer in $\{0, \ldots, k^n - 1\}$ using a standard place-value system: The

[3] For concreteness: This graph H has vertices (v, c) where v is a vertex in G and c is a color, and the edge-weight between (v, c) and (w, d) is the probability that in a uniform list-coloring of G, v has color c and w has color d.

[4] The $((v, c), (w, d))$-th entry of this matrix is 0 if $v = w$, and otherwise, the difference between the conditional probability that w's color is d given that v's color is c and the (marginal) probability that w's color is d.

coloring x which assigns color $x(i)$ to vertex i is represented by $\sum_{i=0}^{n-1} x(i) \cdot k^i$. Using a number of C++ data structures such as `vectors` and `unordered_maps`, we were able to represent the entire Glauber dynamics sparsely in memory (after discovering it with a DFS). This sparse representation allowed us to quickly implement multiplication by the stochastic matrix of the Glauber dynamics, which is the main subroutine in both convergence and spectral gap estimation. Our final code was over 1000 times faster than our original C implementation, which was itself much more efficient than our prototype Python code.

We ran results to calculate TV-distance and ν_2 for 3-regular graphs for $k = 5, 6, 7$ colors and $n = 4, 6, 8$ up to 12 vertices and 4-regular graphs for $k = 6, 7$, $n = 5, 6, \ldots 11$ vertices. Some cases were not computed due to slow running time.

See www.github.com/gracetian6/mcmc for the complete code implementation.

3.2 Results

3.2.1 Total variation distance and stopping time

Earlier we discussed theoretical arguments that constructed δ-distance decreasing couplings to prove $O(n \log n)$ mixing time bounds for $k > 3\Delta$ or 2Δ. Conjecture 1 posits that rapid mixing happens for $k \geq \Delta + 2$. A natural question that arises is how these theoretical bounds scale practically. If it is easier to prove theoretical bounds for larger k relative to Δ, is it the case that for larger k the mixing time converges quicker? There hasn't been known literature on experimental results for convergence of MCMC sampling coloring problem.

We first consider a 4-regular graph on $n = 5$ vertices. In Figure 1, As the number of colors increase, the steps to reach a mixing of 0.01 decays exponentially to around 30. In other words, when we add more colors, this means it takes less time for Markov chain to converge to the uniform distribution (i.e. for any possible coloring to appear with equal probability). Intuitively this makes sense because even though we have more colorings in the state space, these colorings are better connected, and thus do not require as many steps to converge.

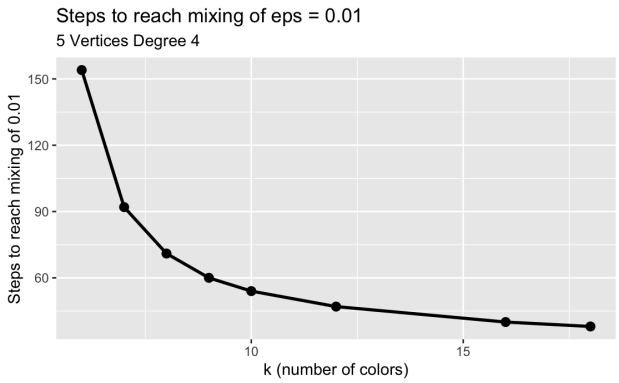

Figure 1: Steps to reach mixing of eps = 0.001 as K increases for 4-regular, 5 Vertex graph.

In Figure 2, to see whether it is always the case that fewer steps are needed to hit any threshold value, we look at plot of how TV distance changes as steps increase from $k = 6$ to 18. The TV distance decays exponentially as steps increases for all the colorings. While the curves overlap slightly at the beginning, the curves don't seem to intersect much. It seems to consistently be that for any given ε (imagine a horizontal line that intersects all the curves), the higher the number of colors, the fewer steps it takes. Experimentally, the limiting curve appears to be around $1.9e^{-0.14 \times \text{steps}}$ as $k \to \infty$ based on the exponential fits of previous curves.

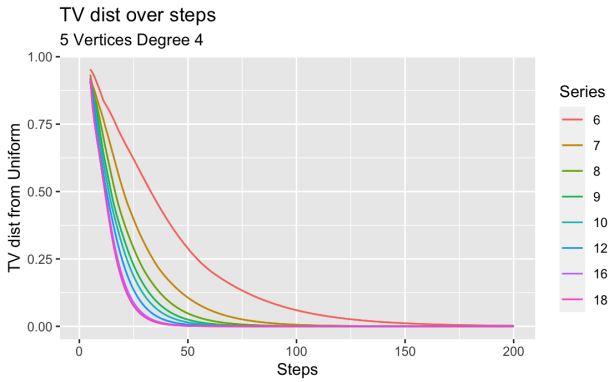

Figure 2: TV distance as steps increase for $k = 6$ to 18 for 4-regular, 5 Vertex graph.

In figure 3, we now generalize from the 5 vertex, 4-regular graph to other graphs where we vary the number of vertices, degrees, and colors. We plot the graph that measures how number of steps to hit a mixing time of 0.001 changes as the number of vertices increases. We connect points that correspond to graphs with the same number of colors and degrees (k, Δ).

Surprisingly, when we connect points that hold (k, Δ) constant, all curves with different (k, Δ) are disjoint for $n \in [4, 8]$. As n ranges from 8 to 12, the curves corresponding to $(k, \Delta) = (5, 3)$ and $(6, 4)$ intersect. All curves are upward sloping, which means that as we increase the number of vertices and hold k, Δ constant, the stopping time increases. Observe that the ordering of the curves for $n \in [4, 8]$ is determined by the ratio $\frac{k}{\Delta}$. Intuitively that makes sense because our theoretical bounds relate the ratios of $\frac{k}{\Delta}$. As this ratio gets smaller, it gets harder to prove that rapid mixing still holds. As we see for larger n, the ordering of the curves does not always hold; however, this may be due to randomly selecting harder graph instances.

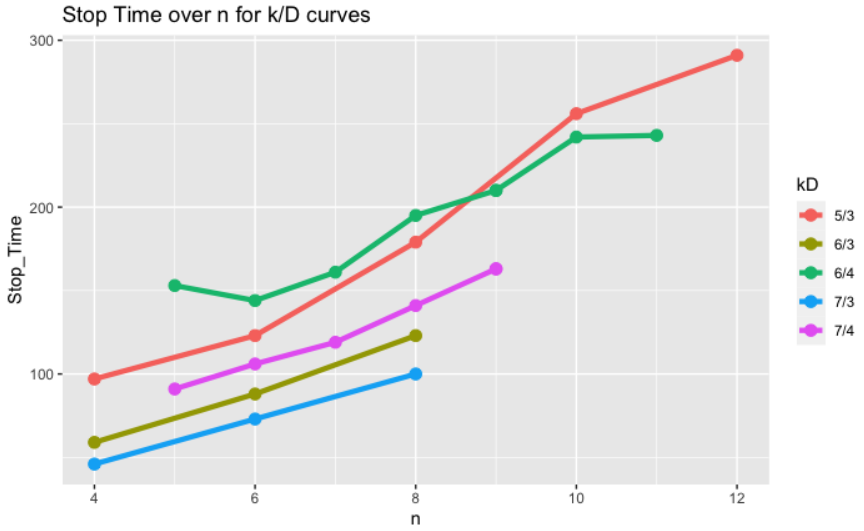

Figure 3: Stop Time as Vertices increases for graphs with different colorings and degrees

In figure 4, we fit a regression line to predict stopping time with n, Δ, k: $stopTime = 133.3985 + 7.1494n \log n + 44.1298\Delta - 40.1998k$ for $k \geq \Delta + 2$. The regression seems to be a good fit since all predictors are significant, and the R^2 is 0.9618. This is consistent with our observations above that stopping time increases with n or Δ increases and decreases with k. The positive and significant relationship between

stopping time and $n \log n$ seems to support the conjecture that the mixing time has an upper bound of $O(n \log n)$ for $k \geq \Delta + 2$, but more data points will be needed to make more certain conclusions.

```
Call:
lm(formula = Stop_Time ~ nlogn + D + k, data = mcmc)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 133.3985    26.8211   4.974 6.38e-05 ***
nlogn         7.1494     0.4508  15.861 3.65e-13 ***
D            44.1298     6.2953   7.010 6.39e-07 ***
k           -40.1998     4.2969  -9.356 6.14e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.54 on 21 degrees of freedom
Multiple R-squared:  0.9618,Adjusted R-squared:  0.9564
F-statistic: 176.4 on 3 and 21 DF,  p-value: 4.766e-15
```

Figure 4: Regression to predict stopping time wrt n, Δ, k.

Notice that the effect of changing V (coefficient of 16.735) on stopping time seems to be less pronounced than that of D or K (coefficients of 46.091 and -31.750 respectively). This is probably the case because V has a bigger range while D and K have smaller ranges so the coefficients scale appropriately as a result.

3.2.2 The spectral gap, ν_2

Another metric for convergence is $\nu_2 = 1 - \omega_2$ for undirected graphs. As is well-known, small ω_2 (i.e., big ν_2) implies fast convergence for a (sufficiently lazy) random walk (cf. Spielman [Spi19, §10.4]). This is consistent with our experimental results in Figure 5, where we see a negative relationship with ν_2 and stopping time.

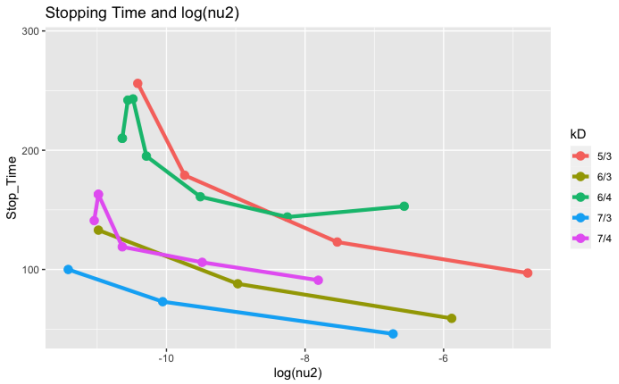

Figure 5: Negative relationship between $\log(\nu_2)$ and stopping time.

In the next figure, we also see a relationship between ν_2 and n, corresponding to the fact that the Glauber dynamics tends to mix more slowly on larger starting graphs.

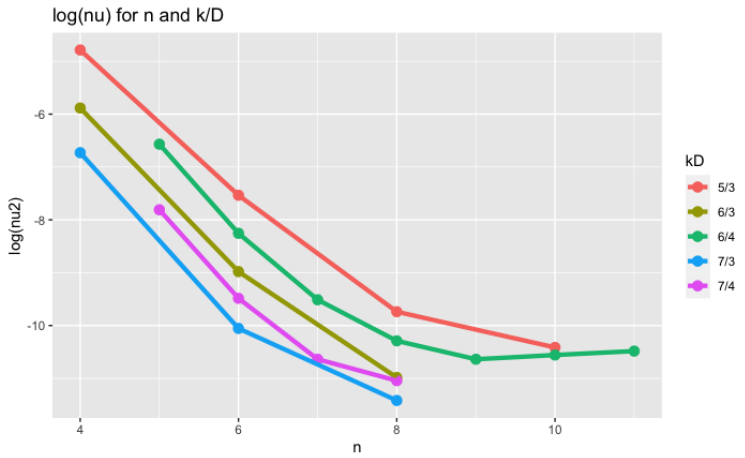

Figure 6: $\log(\nu_2)$ as n increases.

In figure 7, we experimentally investigate the theoretical bound that $t_{mix} \leq O\left(\frac{\log\left(\frac{n}{\varepsilon}\right)}{1-\omega_\pi}\right)$, where ω_π is the spectral norm of the stochastic matrix for the Glauber dynamics (cf. [Spi19, §10.4]). Since the Glauber dynamics has constant laziness for fixed k and Δ, $\omega_\pi = \max\{\omega_2, -\omega_n\} = \omega_2 = 1 - \nu_2$. Hence $t_{mix} \leq O\left(\frac{\log(n/\varepsilon)}{\nu_2}\right)$. It appears that the constant in this bound is very large, since the magnitude of $\log(n/0.01)/\nu_2$ is 10^5 while the stopping time has magnitude of 10^2. While as we discuss below (in §4), this does not mean that analyzing the spectral gap may not suffice to prove that the Glauber dynamics mixes rapidly, it may indicate e.g. that greedily picking a coloring is *not* a good proxy for the worst-case starting distribution. Note that the ordering of the curves is consistent with the ordering of stopping time over n in Figure 3, which may be because of the negative relationship of ν_2 and n shown in Figure 3.

Figure 7: Empirical plot to investigate $t_{mix} \leq O(\log(n/\varepsilon)/\nu_2$.

4 Conclusion and future work

Our experimental results for mixing time are generally consistent with theoretical bounds. As the ratio $\frac{k}{\Delta}$ gets smaller, the mixing time generally seems to increase with n (but not necessarily for $n \geq 8$). Similarly,

as this ratio gets smaller, theoretical bounds have been harder to prove that rapid mixing still holds. More generally, we found that as stopping time increases when n and Δ increases and decreases when k increases. For the 5-vertex, 4-regular graph, we found an exponential decay relationship with TV distance as steps increase.

Our experimental results for spectral gap ν_2 are generally consistent with the theory for ν_2 and convergence. While since the Glauber dynamics for colorings (in the context of proper colorings) are undirected and have constant laziness for fixed k and Δ, if the spectral gap is sufficiently large, the worst-case mixing time grows linearly with the spectral gap as a function of n, our experiments showed that the constants involved in this relationship are very large, and so there may be interesting information "hidden" in the spectral gap or mixing time that cannot easily be inferred from the other.

Given more computational time, we would run more trials for large values of n, with the aim of gleaning the distribution of hardness between different input instances. How much spread is there in the mixing time of the Glauber dynamics for different graphs of the same size? Can we identify any graph-specific features that can be used to predict how long mixing time will be? What about the same questions for the spectral gap? Also, more trials would be useful for our experiments that measured the spectral gap, since the algorithm relies on an appropriate choice of a random starting vector.

Finally in terms of the experiments, we would also like to explore whether it is possible to estimate convergence without storing the entire distribution at every step using heuristics.

There are also two interesting theoretical questions raised by our experiments which we plan to address in future work. One is, is there a provable strong form of monotonicity as k increases for mixing time or the spectral gap? Formally, we conjecture:

Conjecture 8. *Let G be an undirected graph. $t_{mix}(\epsilon)$ (resp. the spectral gap) of the Glauber dynamics with k colors is greater than (resp. less than) the time (resp. gap) for $k+1$ colors.*

Also, we observed that for a fixed starting graph, as k goes to infinity, the mixing-time curve appears to tend towards a fixed, nonzero limit. We also conjecture:

Conjecture 9. *Let G be an undirected graph. $t_{mix}(\epsilon)$ approaches $\Theta\left(n \log\left(\frac{n}{\epsilon}\right)\right)$ as $k \to \infty$.*

The intuition for this conjecture is that as the number of colors gets very large, Glauber dynamics approaches the random walk on the generalized hypercube $[k]^n$, which will mix in time $\Theta(n \log n)$ according to the coupon collector problem.

References

[ALG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *Proceedings of the 60th Annual Symposium on Foundations of Computer Science*. 2020.

[Ana+19] Nima Anari et al. "Log-Concave Polynomials II: High-Dimensional Walks and an FPRAS for Counting Bases of a Matroid". In: *STOC 2019: Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing*. Vol. 51. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12.

[BD97] Russ Bubley and Martin Dyer. "Path coupling: A technique for proving rapid mixing in Markov chains". In: *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*. 1997, pp. 223–231.

[BDG98] Russ Bubley, Martin Dyer, and Catherine Greenhill. "Beating the 2Δ Bound for Approximately Counting Colourings: A Computer-Assisted Proof of Rapid Mixing". In: *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '98. San Francisco, California, USA: Society for Industrial and Applied Mathematics, 1998, pp. 355–363.

[Che+19] Sitan Chen et al. "Improved Bounds for Randomly Sampling Colorings via Linear Programming". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '19. San Diego, California: Society for Industrial and Applied Mathematics, 2019, pp. 2216–2234.

[Che+20] Zongchen Chen et al. *Rapid Mixing for Colorings via Spectral Independence*. July 2020. arXiv: 2007.08058 [cs.DS].

[Das11] Constantinos Daskalakis. "6.896: Probability and Computation, Spring 2011". 2011. URL: http://people.csail.mit.edu/costis/6896sp11/.

[DS93] Persi Diaconis and Laurent Saloff-Coste. "Comparison Theorems for Reversible Markov Chains". In: *The Annals of Applied Probability* 3.3 (1993), pp. 696–730.

[FV05] Alan Frieze and Eric Vigoda. *A survey on the use of Markov chains to randomly sample colorings*. Oct. 2005. URL: https://www.math.cmu.edu/~af1p/Texfiles/colouringsurvey.pdf.

[Jer95] Mark Jerrum. "A Very Simple Algorithm for Estimating the Number of k-Colorings of a Low-Degree Graph". In: *Random Structures & Algorithms* 7.2 (Sept. 1995), pp. 157–165.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. "A Polynomial-Time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries". In: 51.4 (July 2004), pp. 671–697.

[JVV86] Mark Jerrum, Leslie G. Valiant, and Vijay Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical Computer Science* 43 (1986), pp. 169–188.

[LS93] Mona Lubin and Alan D. Sokal. "Comment on "Antiferromagnetic Potts models"". In: *Physics Review Letters* 71 (11 Sept. 1993), p. 1778.

[MM11] Cristopher Moore and Stephan Mertens. *The Nature of Computation*. Oxford, England, UK: Oxford University Press, 2011.

[Spi19] Daniel A. Spielman. "Spectral and Algebraic Theory". Dec. 2019. URL: http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.

[SS97] Jesús Salas and Alan Sokal. "Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem". In: *Journal of Statistical Physics* 86 (1997), pp. 551–579.

[Tre17] Luca Trevisan. "Lecture Notes on Graph Partitioning, Expanders and Spectral Methods". June 2017. URL: https://people.eecs.berkeley.edu/~luca/books/expanders-2016.pdf.

[Vig00] Eric Vigoda. "Improved bounds for sampling couplings". In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1555–1569.

A Proofs of coupling theorems

Proof of Lemma 3. We follow the standard proof as presented in Daskalakis' notes [Das11, p. 3-3]. First, we recall the standard fact that given any two distributions X,Y on Ω,

$$\|X - Y\|_{TV} = \max_{S \subseteq \Omega} \left| \Pr_{x \sim X}[x \in S] - \Pr_{y \sim Y}[y \in S] \right|;$$

in other words, the total variation distance between X and Y is equal to the best advantage any "test" S can have in distinguishing X and Y. For convenience' sake, let $X'(x)$ denote $\Pr_{x' \sim X'}[x' = x]$ and $Y'(y)$ denote $\Pr_{y' \sim Y'}[y' = y]$ Then it follows that

$$\begin{aligned}
\|X_t - Y_t\|_{TV} &= \max_{S \subseteq \Omega} \left| \Pr_{x' \sim X'_t}[x' \in S] - \Pr_{y' \sim Y'_t}[y' \in S] \right| \\
&= \sum_{\substack{z \in \Omega \\ X'_t(z) \geq Y'_t(z)}} (X'_t(z) - Y'_t(z)) \\
&= \sum_{z \in \Omega} (X'_t(z) - \min\{X'_t(z), Y'_t(z)\}) \\
&= 1 - \sum_{z \in \Omega} \min\{X'_t(z), Y'_t(z)\} \\
&\leq \Pr_{(x',y') \sim (X'_t, Y'_t)}[x' \neq y']
\end{aligned}$$

The second line holds because we can always pick S such that the quantity inside the absolute value is positive.[5] The third line holds because of the constraint that $X'_t(z) \geq Y'_t(z)$. The fourth line holds because X'_t is a probability distribution. The last line follows because

$$\sum_{z \in \Omega} \min\{X'_t(z), Y'_t(z)\} \geq \sum_{z \in \Omega} \Pr_{(x',y') \sim (X'_t, Y'_t)}[x' = z, y' = z] = \Pr_{(x',y') \sim (X'_t, Y'_t)}[x' = y']$$

□

Proof of Lemma 4. Consider any starting distribution X_0, let π be M's stationary distribution, and let $(X'_1, Y'_1), \ldots, (X'_t, Y'_t), \ldots$ evolve according to M from (X_0, π). Since $d(x, y) > 0$ iff $x > y$ and d is integer-valued, by Markov's inequality,

$$\Pr_{(x,y) \sim (X'_t, Y'_t)}[x' \neq y'] \leq \mathbb{E}[d(x', y')].$$

Since (X', Y') is δ-distance decreasing for all pairs $(x_0, y_0) \in \Omega^2$, it also decreases expected distance by a $(1 - \delta)$ factor for all distributions on Ω^2 by linearity of expectation. Hence

$$\mathbb{E}_{(x,y) \sim (X'_t, Y'_t)}[d(x', y')] \leq (1 - \delta)^t \mathbb{E}_{(x_0,y_0) \sim X_0 \times Y_0}[d(x_0, y_0)] \leq (1 - \delta)^t d_{max}.$$

Hence applying Lemma 3, we get that

$$\|X_t - \pi\|_{TV} \leq (1 - \delta)^t d_{max}.$$

So taking $t = \frac{\log(d_{max}/\epsilon)}{\delta}$ suffices to get $\|X_t - \pi\|_{TV} \leq \epsilon$, as desired. □

[5] If the quantity is negative, setting $S' := \Omega \setminus S$ will yield a positive number with the same absolute value.

Proof of Theorem 5. We will show that we can construct a δ-distance-decreasing coupling between all pairs of states in Ω. For more details on this construction and its analysis, see Daskalakis' notes [Das11, p. 6-3]. We will define this coupling for pairs of start states (x_0, y_0); it will extend by linearity to all possible start distributions. Let (x_0, y_0) be an arbitrary pair of states in Ω, and consider a shortest path between them through S; that is, let $x_0 = z^0, z^1, \ldots, z^{i-1}, z^i = y_0$ be a path of shortest length such that each pair $(z^j, z^{j+1}) \in S$. The probability distribution (X_1', Y_1') evolving from (x_0, y_0) will be the result of an iterative sampling process: First we sample $(X_1', Z_1'^1)$ evolving from (x_0, z^1), and then $(Z_1'^1, Z_1'^2)$ evolving from (z^1, z^2) conditioned on $Z_1'^1 = z^1$, etc., continuing until we sample (Z'^{i-1}, Y_1'), and we output (X_1', Y_1'). The reader should verify inductively that this gives a valid coupling (or consult Daskalakis' notes).

Let (X', Y') denote this coupling. We claim that it is δ-distance-decreasing. Indeed, consider any pair of start states (x_0, y_0), and let $z^0, \ldots, z^i$ be as above. Then if $X_1', Z_1'^1, \ldots, Z_1'^{i-1}, Y_1'$ evolve from (x_0, y_0) as in the composed coupling above, we have

$$\begin{aligned}
\mathbb{E}[d(X_1', Y_1')] &\leq \mathbb{E}\left[\sum_{j=0}^{i-1} d(Z_1'^j, Z_1'^{j+1})\right] && \text{(triangle ineq.)} \\
&\leq \sum_{j=0}^{i-1} \mathbb{E}[d(Z_1'^j, Z_1'^{j+1})] && \text{(lin. of exp.)} \\
&\leq (1-\delta) \sum_{j=0}^{i-1} d(z^j, z^{j+1}) && (\delta\text{-distance-decreasing}) \\
&(1-\delta) d(x_0, y_0) && \text{(shortest path)},
\end{aligned}$$

as desired. □