Improved Bounds for Randomly Sampling Colorings via Linear Programming

Sitan Chen[*]   Michelle Delcourt [†]   Ankur Moitra[‡]   Guillem Perarnau [§]

Luke Postle [¶]

July 12, 2018

Abstract

A well-known conjecture in computer science and statistical physics is that Glauber dynamics on the set of k-colorings of a graph G on n vertices with maximum degree Δ is rapidly mixing for $k \geq \Delta + 2$. In FOCS 1999, Vigoda [Vig99] showed that the flip dynamics (and therefore also Glauber dynamics) is rapidly mixing for any $k > \frac{11}{6}\Delta$. It turns out that there is a natural barrier at $\frac{11}{6}$, below which there is no one-step coupling that is contractive with respect to the Hamming metric, even for the flip dynamics.

We use linear programming and duality arguments to fully characterize the obstructions to going beyond $\frac{11}{6}$. These extremal configurations turn out to be quite brittle, and in this paper we use this to give two proofs that the Glauber dynamics is rapidly mixing for any $k \geq (\frac{11}{6} - \epsilon_0)\Delta$ for some absolute constant $\epsilon_0 > 0$. This is the first improvement to Vigoda's result that holds for general graphs. Our first approach analyzes a variable-length coupling in which these configurations break apart with high probability before the coupling terminates, and our other approach analyzes a one-step path coupling with a new metric that counts the extremal configurations. Additionally, our results extend to list coloring, a widely studied generalization of coloring, where the previously best known results required $k > 2\Delta$.

1 Introduction

1.1 Background

Here we study the problem of sampling random proper colorings of a bounded degree graph. More precisely, let k be the number of colors and let Δ be the maximum degree. A long-standing open question is to give an algorithm that works for any $k \geq \Delta + 2$, when the space of proper colorings is first connected. Despite a long line of intensive investigation [Jer95, SS97, DF03, DFHV13, Hay03, HV03, Mol04, HV05, FV06, FV07], the best known bounds are quite far from the conjecture.

In fact, there is a natural Markov chain called the *Glauber dynamics* that is widely believed to work: in each step, choose a random node and recolor it with a random color not appearing among its neighbors. It is easy to see that its steady state distribution is uniform on all proper k-colorings, again provided that $k \geq \Delta + 2$. It is even conjectured that on an n node graph, the mixing time is $O(n \log n)$ which would be tight [HS05]. We remark that rapidly mixing Markov chains for sampling random colorings immediately give a fully polynomial randomized approximation scheme (FPRAS) for counting the number of proper colorings. There is also interest in this question in combinatorics [BW02] and in statistical physics, where it corresponds to approximating the partition function of the zero temperature anti-ferromagnetic Potts model [Pot52].

Jerrum [Jer95] gave the first significant results and showed that when $k > 2\Delta$ the Glauber dynamics mixes in time $O(n \log n)$. The modern proof of this result is easier and proceeds through *path coupling* [BD97], whereby it is enough to couple the updates between two colorings σ and τ that differ only at a single node v and show that the expected distance between them is strictly decreasing. Then Jerrum's bound follows by comparing how often the distance between the colorings decreases (when v is selected and after the update has the same color in both) vs. how often it increases (when a neighbor of v is selected and recolored in one but not the other). This result is closely related to work in the statistical physics community by Salas and Sokal [SS97] on the *Dobrushin uniqueness condition*.

In a breakthrough work, Vigoda [Vig99] gave the first algorithm for sampling random colorings that crossed the natural barrier of 2Δ. His approach was through a different Markov chain that in addition to recoloring single nodes also swaps the colors in larger Kempe components (which are also called alternating components). His chain was a variant of the Wang-Swendsen-Kotecký (WSK) algorithm [WSK89]. The key insight is that the bottleneck in Jerrum's approach — when the neighbors of v all have distinct colors — can be circumvented by flipping larger components. More precisely, when a neighbor of v is recolored in one chain in a way that would have increased the distance, one can instead match it with the flip of a Kempe component of size two in the other chain that keeps the distance the same. But now one needs to couple the flips of larger Kempe components in some manner. Vigoda devised a coupling and a choice of flip parameters that works for any $k > \frac{11}{6}\Delta$. His Markov chain mixes in time $O(n \log n)$ and one can also connect it to Glauber dynamics and prove an $O(n^2)$ mixing time under the same conditions. This is still the best known bound for general graphs.

Subsequently, there was a flurry of work on getting better bounds for restricted families of graphs. Dyer and Frieze [DF03] considered graphs of maximum degree $\Omega(\log n)$ and girth $\Omega(\log \log n)$ and proved that the Glauber dynamics mixes rapidly whenever $k > \alpha\Delta$ where α is the solution to $\alpha = e^{1/\alpha}$ and numerically $\alpha = 1.763 \cdots$. Their approach was to show that under the uniform distribution on proper colorings, the number of colors missing from the neighborhood of v is roughly $k(1 - \frac{1}{k})^\Delta$ with high probability. Results like these were later termed *local uniformity properties*. They were improved in many directions in terms of reducing the degree and/or the girth requirements to be independent of n [Hay03, Mol04, HV05, FV06, LM06], culminating in two incomparable results.

Dyer et al. [DFHV13] showed that Glauber dynamics mixes rapidly whenever $k > \beta\Delta$ where β is the solution to $(1 - e^{-1/\beta})^2 + \beta e^{-1/\beta} = 1$ and numerically $\beta = 1.489\cdots$ for girth $g \geq 6$ and the degree Δ being a sufficiently large constant. Hayes and Vigoda [HV03] showed rapid mixing for any $k > (1 + \epsilon)\Delta$ for any $\epsilon > 0$ provided that the girth $g \geq 11$ and the degree is logarithmic, using an intriguing non-Markovian coupling.

There have been many other improvements, but all for special graph families. Through an eigenvalue generalization of the Dobrushin condition, Hayes [Hay06] showed that Glauber dynamics mixes rapidly for $k > \Delta + c\sqrt{\Delta}$ on planar graphs and graphs of constant treewidth. Berger et al. [BKMP05] showed rapid mixing on graphs of logarithmic cutwidth, which was strengthened by Vardi [Var17] to graphs of logarithmic pathwidth. Some recent papers have studied settings such as bipartite or random graphs [DFFV06, MS10, EHŠV18], where it is possible to mix with fewer colors than the maximum degree. Hayes et al. [HVV15] notably improved the abovementioned result of [Hay06] to show that Glauber dynamics in fact mixes rapidly for planar graphs when $k = \Omega(\Delta/\log\Delta)$. [TVVY12] established that the mixing time of Glauber dynamics for sampling colorings of trees undergoes a phase transition at the reconstruction threshold (up to first order). *Given that there has been progress in so many restricted graph families, it is natural to wonder why there hasn't been any further progress on the general case since Vigoda's results.*

1.2 Our Results

Our main result is the first improvement on randomly sampling colorings on general bounded degree graphs since the FOCS 1999 paper of Vigoda [Vig99]. Specifically, we prove:

Theorem 1.1. *The flip dynamics for sampling k-colorings is rapidly mixing with mixing time $O(n \log n)$, for any $k \geq (\frac{11}{6} - \epsilon_0)\Delta$ where $\epsilon_0 > 0$ is an absolute constant that is independent of Δ.*

As in Vigoda's paper [Vig99], we obtain the following as implications of our main result[1]:

Theorem 1.2. *The Glauber dynamics for sampling k-colorings is rapidly mixing with mixing time $O(n^2)$, for any $k \geq (\frac{11}{6} - \epsilon_0)\Delta$ where $\epsilon_0 > 0$ is the same constant from Theorem 1.1.*

Theorem 1.3. *The k-state zero temperature anti-ferromagnetic Potts model on $\mathbb{Z}^d$ lies in the disordered phase when $k \geq (\frac{11}{3} - 2\epsilon_0)d$ where $\epsilon_0 > 0$ is the same constant from Theorem 1.1.*

Our proof is guided by linear programming and duality arguments. The starting point is the observation that choosing the best flip parameters in the flip dynamics (i.e. the probability of flipping Kempe components of each possible size), when utilizing Vigoda's greedy coupling [Vig99], can be cast as a linear program. In this manner, Vigoda's analysis provides a feasible solution and it is natural to wonder if choosing the flip parameters differently or flipping larger size components would lead to an improvement. As it turns out, the answer is no in a strong sense. Not only are Vigoda's choice of parameters an optimal solution to the linear programming relaxation[2], but we

---

[1]In order to prove rapid mixing of Glauber dynamics, one can use the comparison technique of Diaconis and Saloff-Coste [DSC93]. Vigoda [Vig99] directly used the results in [DSC93] to show that the mixing time of the Glauber dynamics is $O(n^2 \log n)$. It has been observed (see e.g. [FV07]) that $\tau_{mix}(\epsilon) = O(n(\log n + \log \epsilon^{-1}))$ for flip dynamics implies mixing time $O(n^2)$ for Glauber dynamics. This can be shown using spectral bounds on the mixing time for $\epsilon = 1/n$ [Sin92] and observing that the spectral gaps of Glauber dynamics and flip dynamics are the same up to a constant factor. The same argument applies to our case, so Theorem 1.1 implies Theorem 1.2.

Regarding Theorem 1.3, it is known (see e.g. Lemma 7 of [Vig99]) that it holds provided that *under any boundary configuration, the flip dynamics mixes in time $O(n \log n)$* on Q_L the d-dimensional cube of $\mathbb{Z}^d$ with side length $2L+1$. Theorem 1.4 below implies this is the case for $k \geq (11/6 - \epsilon_0)\Delta$ (note that the degree of Q_L is $\Delta = 2d$).

[2]Several approximations are made along the way in reaching this linear program, such as restricting to flipping components of size at most 6 and replacing certain infinite sets of constraints with a finite set of stronger constraints.

use *extremal configurations*, which correspond to tight constraints in the linear program, to build a small family of hard instances showing that $\frac{11}{6}$ is a natural barrier for a class of analyses. Specifically, for all n and infinitely many Δ we can exhibit two graphs G_1 and G_2, together with a pair of proper colorings (σ_i, τ_i) for each G_i that differ at a single node, such that the following holds: for any choice of flip parameters and one-step coupling of the flip dynamics, (σ_1, τ_1) and (σ_2, τ_2) cannot both contract in Hamming distance under the coupling provided $k < \frac{11}{6}\Delta$ (see Construction 3.1 and Lemma 3.3).

At this juncture, there are two potential approaches for circumventing the $\frac{11}{6}$ barrier:

(1) Instead of using a one-step coupling, use a multi-step coupling, still with respect to the Hamming metric.

(2) Use a one-step coupling but change the metric.

As it turns out, *both* approaches work and we are able to give two separate proofs of Theorem 1.1. The present work is a merger of [CM18] and [DPP18], and we chose to present the results together because there are parallels between the two proofs.[3] The main idea behind both approaches is that any extremal configuration is quite *brittle*, and if there are few enough extremal configurations present, there is a choice of flip parameters for which we can go below the 11/6 barrier. To get an idea for just how brittle these configurations are, suppose v is the unique node of disagreement between two colorings σ and τ. For every color c, there is an extremal configuration which if present in σ, τ would require that in σ, the maximal connected induced subgraph colored only with c and $\sigma(v)$ and containing v is a tree of size 7 and moreover that v has degree exactly 2 in this tree. There are many possible transitions in the flip dynamics that would break this rigid pattern, for instance, flipping the color of a descendant of v in the tree.

In our first proof of Theorem 1.1, we run a multi-step coupling which terminates when the Hamming distance between the two colorings changes. We argue that by the time the coupling terminates, extremal configurations like the one above are likely to break apart. In the above example, regardless of the choice of flip parameters, at any point there is an $\Omega(1/n)$ chance of breaking it (Lemma 4.6) and a $\Theta(1/n)$ chance the coupling terminates (Lemma 4.2). What remains, and this is the most delicate part of the proof, is to upper bound the probability that configurations that are not extremal transition to ones that are, and the key point is that this is also $O(1/n)$ (Lemma 4.8). These three bounds together imply that by the end of the coupling, the number of non-extremal configurations around v will be at least some constant times the number of extremal configurations in expectation. This is enough to show that for some tuning of the flip dynamics, even if the expected change in Hamming distance after one step of coupling is still positive when k is slightly below $\frac{11}{6}\Delta$, the expected change by the time the coupling terminates will be negative.

In our second proof of Theorem 1.1, the main idea is essentially the same: we want to win on colorings like G_1 and G_2 at the expense of losing on the other possible types of colorings. However we accomplish this by changing the metric. More precisely, we choose d to be the Hamming metric d_H minus a small correction factor d_B that counts the number of extremal configurations around the node v of disagreement. We show that even when k is slightly less than $\frac{11}{6}\Delta$, there is a choice of flip parameters for which the following win-win analysis shows d contracts in expectation in one step. For pairs of colorings with few non-extremal configurations around v, d_H will stay the same in expectation in one step (Corollary 5.1). When most of the configurations around v are extremal, we show that even though d_H does not contract in expectation in one step, d_B does (Corollary 5.2). It is here that our proofs bear a sharp resemblance: The proof that d_B is contractive revolves around the exact same types of reasoning about lower bounding the probability that an extremal

---

[3]The ϵ_0 obtained in Theorem 1.1 is slightly different under the two proofs, but both roughly on the order of 10^{-5}.

configuration breaks apart, and upper bounding the probability that a non-extremal configuration becomes extremal (Lemma 5.2).

While we have chosen to present the technical pieces in both proofs separately because of the subtle differences in their structure and relevant definitions[4], we believe that presenting both of them and drawing analogies when possible gives a more complete picture about the different ways to circumvent bottlenecks in one-step coupling with respect to the Hamming metric, and what the conceptual relationship is between these different techniques.

Lastly, we are able to extend our techniques to the problem of sampling *list colorings*, a natural and well-studied generalization of coloring. Jerrum's proof for $k > 2\Delta$ carries over immediately for list-coloring, and while there have been subsequent works studying this problem in the case of *triangle-free* graphs when $k > 1.763\Delta$ [GMP05, GK07, GKM15], there had been no known improvement on 2Δ for general bounded degree graphs. Given that Vigoda's algorithm for sampling random colorings with $\frac{11}{6}\Delta$ colors has been known for many years, it is somewhat surprising that it (and our subsequent improvements) can be extended to the list-coloring setting in a fairly straightforward manner:

Theorem 1.4. *The Glauber dynamics for sampling k-list-colorings is rapidly mixing with mixing time $O(n^2)$, for any $k \geq (\frac{11}{6} - \epsilon_0)\Delta$ where $\epsilon_0 > 0$ is the same constant from Theorem 1.1.*

The key step in our proof is to introduce a notion of flip dynamics for list colorings where a Kempe component is flipped only if both colors appear in all lists of vertices of the component; we call such components *flippable*. While such a chain seems natural in hindsight, we are not aware of it appearing anywhere in the literature. Roughly speaking, introducing this distinction between flippable and un-flippable Kempe components requires introducing additional constraints to Vigoda's linear program. The point is that none of these extra configurations are extremal under the flip parameters used to prove Theorem 1.1, so both proofs of Theorem 1.1 carry over to prove Theorem 1.4.

1.3 Further Discussion

Here we survey some previous uses of multi-step coupling, alternative metrics, and linear programming in approximate sampling. In terms of multi-step coupling for sampling colorings, the first improvements to Vigoda's bound for graphs of large degree and girth by Dyer and Frieze [DF03] used a burn-in method. Hayes and Vigoda [HV03] gave a sophisticated method based on looking into the future to prevent singly blocked updates in Jerrum's maximal coupling. Non-Markovian couplings have also been used to get $O(n \log n)$ mixing time — rather than $O(n^2)$ — for Glauber dynamics with $k = (2 - \epsilon)\Delta$ colors [DGG$^+$01, HV07]. The crux of these last results is to terminate the coupling at a random stopping time corresponding to the first point at which something interesting happens in the coupled evolution, e.g. the Hamming distance changing. This *variable-length coupling* approach is also one of the approaches we take.

On the other hand, to our knowledge alternative metrics have found only one application in previous works on sampling colorings, namely in the analysis of the "scan" chain for sampling colorings of bipartite graphs in [BDK06]. That said, path coupling using alternative metrics has found success in other problems in the approximate sampling literature [LV99, BD98, BDK06]. Bordewich et al. [BDK06] gave evidence that the multi-step stopping time-based approach can be captured by one-step coupling with an appropriate metric, though the metric they use to establish this connection is itself based on stopping times. In an orthogonal direction, another take on the

---

[4]See Remark 4.2 for a discussion of some of these differences.

question of designing metrics for one-step coupling is given in [HVV15] via the spectral radius of the adjacency matrix and in [EHŠ$^+$16] via the Jacobian of the belief propagation operator.

Lastly, we remark that before Vigoda's work [Vig99], Bubley et al. [BDG98] used linear programming to show that Glauber dynamics is rapidly mixing with five colors on graphs with maximum degree three. Their approach however required solving several hundred linear programs, and was subject to "combinatorial explosion" as a function of the degree.

Organization

In Section 2 we review the basics of path coupling, briefly summarize Vigoda's Markov chain and coupling analysis, and recall some basic notions in variable-length coupling. In Section 3, we interpret Vigoda's one-step coupling analysis as implicitly solving a linear program and identify a family of worst-case neighborhoods which is tight at $k > \frac{11}{6}\Delta$ for one-step coupling of the flip dynamics with respect to the Hamming metric. In Section 4, we give a proof of Theorem 1.1 via variable-length coupling. In Section 5, we give a proof via one-step coupling with an alternative metric. In Section 6, we overview how to extend our techniques to prove Theorem 1.4.

In Appendix A, we give a self-contained exposition of the details of Vigoda's coupling analysis. In Appendix B, we provide the set of configurations in Vigoda's analysis that are extremal under his choice of flip parameters. In Appendix C and D, we supply proofs of technical lemmas that appeared in Sections 4 and 5 respectively. In Appendix E, we complete the proof of Theorem 1.4 that was sketched in Section 6.

2 Preliminaries

Let $\mathbb{N}_0$ denote the set of non-negative integers. In a graph $G = (V, E)$, for vertex $v \in V$ define $N(v)$ to be the set of neighbors of v and $\Delta(v)$ to be the degree of vertex v. Given a (proper or improper) coloring $\sigma : V \to [k]$, define $A_\sigma(v)$ to be the set of colors *available to* v, i.e. the set of colors c for which no neighbor of v is colored c. Given a Markov chain with transition probability matrix P on finite state space Ω and initial state $\sigma^{(0)}$, denote the distribution of state $\sigma^{(t)}$ at time t by $P^t(\sigma^{(0)}, \cdot)$. Denote the stationary distribution of an ergodic Markov chain by π, let

$$\tau_{mix}(\epsilon) := \max_{\sigma^{(0)} \in \Omega} \min\{t : d_{\mathrm{TV}}(P^{t'}(\sigma^{(0)}, \cdot), \pi) \le \epsilon \ \forall t' \ge t\},$$

where $d_{\mathrm{TV}}(\cdot, \cdot)$ is the total variation distance, and define the *mixing time* τ_{mix} to be $\tau_{mix}(1/2e)$.

The state spaces we will be interested in are $\Omega = [k]^V$ the set of all colorings (or *labellings*) of G, and Ω^* the set of *proper* colorings of G.

2.1 The Flip Dynamics

The Markov chain we use is a variant of the Wang-Swendsen-Kotecký (WSK) algorithm [WSK89] studied in [Vig99], which we define below. In a (proper or improper) coloring σ of a graph G, for vertex v and color c let the *Kempe component* $S_\sigma(v, c)$ denote the set of vertices w for which there exists an alternating path between v and w using only the colors c and $\sigma(v)$. Under this definition, $S_\sigma(v, \sigma(v)) = \emptyset$. The motivation for this definition is that if σ is proper, then if one *flips* $S_\sigma(v, c)$, i.e. changes the color of all $\sigma(v)$-colored vertices in $S_\sigma(v, c)$ to c and that of all c-colored vertices in $S_\sigma(v, c)$ to $\sigma(v)$, the resulting coloring is still proper.

Definition 2.1. Let $\mathbf{p} = \{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ be a collection of *flip parameters*. The *flip dynamics* is a random process generating a sequence of colorings $\sigma^{(0)}, \sigma^{(1)}, \sigma^{(2)}, \ldots$ of G where $\sigma^{(0)}$ is an arbitrary coloring in Ω and $\sigma^{(t)}$ is generated from $\sigma := \sigma^{(t-1)}$ as follows:

1. Select a random vertex $v^{(t)}$ and a random color $c^{(t)}$.

2. Let $\alpha = |S_\sigma(v^{(t)}, c^{(t)})|$ and flip $S_\sigma(v^{(t)}, c^{(t)})$ with probability p_α/α.

The reason for the p_α/α term is that we have a nice equivalent way of formulating the flip dynamics. Let $\mathcal{S}_\sigma$ denote the family of all Kempe components in G under the coloring σ, i.e. all $S \subset V$ for which there exist v, c such that $S = S_\sigma(v, c)$. Here we emphasize that $\mathcal{S}_\sigma$ is a multiset.[5]
Then, $\sigma^{(t)}$ is generated from $\sigma := \sigma^{(t-1)}$ as follows:

1. Pick any component $S^{(t)} \in \mathcal{S}_\sigma$, each with probability $1/nk$.

2. Let $\alpha = |S^{(t)}|$ and flip $S^{(t)}$ with probability p_α.

Because the flip dynamics embeds Glauber dynamics, it is ergodic on the space of proper colorings for every $k \geq \Delta + 2$. As every improper coloring has a positive probability of eventually being transformed into a proper one, the space of proper colorings is the only closed subset of the space of all colorings. The following holds as a consequence:

Lemma 2.1 ([Vig99])**.** *The stationary distribution of the flip dynamics is the uniform distribution over proper colorings of G.*

The WSK algorithm corresponds to the choice of $p_\alpha = \alpha$ for all $\alpha \in \mathbb{N}_0$. For the purposes of path coupling, it turns out one only needs to flip Kempe components of size at most some absolute constant N_{max} (in Vigoda's chain, $N_{max} = 6$), and this "local" nature of the flip dynamics will simplify the analysis. Henceforth, we will take $p_0 = 0, p_1 = 1$[6], and $0 \leq p_{\alpha+1} \leq p_\alpha \leq 1$ for all $\alpha \geq 1$.

2.2 Path Coupling

Coupling is a common way to bound the mixing time of Markov chains. A T-step coupling for a Markov chain with transition matrix P and state space Ω defines for every initial $(\sigma^{(0)}, \tau^{(0)}) \in \Omega^2$ a stochastic process $(\sigma^{(t)}, \tau^{(t)})$ such that the distribution of $\sigma^{(T)}$ (resp. $\tau^{(T)}$) is the same as $P^T(\sigma^{(0)}, \cdot)$. (resp. $P^T(\tau^{(0)}, \cdot)$). The *coupling inequality* states that for any starting point $\sigma^{(0)}$ for the Markov chain,

$$d_{\text{TV}}(\sigma^{(t)}, \pi) \leq \max_{\tau^{(0)}} \mathbb{P}[\sigma^{(T)} \neq \tau^{(T)}]. \tag{1}$$

We will think of T-step couplings as random functions $\Omega^2 \to \Omega^2$, so we will denote them by $(\sigma^{(0)}, \tau^{(0)}) \mapsto (\sigma^{(T)}, \tau^{(T)})$, or more succinctly, $(\sigma, \tau) \mapsto (\sigma', \tau')$. For an initial pair of colorings σ, τ, we say that a coupling γ-*contracts for* (σ, τ) for some $\gamma \in (0, 1)$ and metric d on Ω if it satisfies

$$\mathbb{E}[d(\sigma', \tau')] \leq \gamma\, d(\sigma, \tau). \tag{2}$$

If there exist $\alpha > 0$ and a coupling that $(1 - \alpha)$-contracts for all (σ, τ), then one can show that $\tau_{mix} = O(T \log(D)/\alpha)$, where D is the diameter of Ω under d.
For the rest of this subsection, we specialize our discussion of coupling to the setting of sampling colorings. Fix any Markov chain over Ω whose stationary distribution is the uniform distribution over Ω^*, e.g. the Glauber or flip dynamics, and denote it by $\sigma \mapsto \sigma'$.
In complicated state spaces like the space of all proper colorings of a graph, it is often tricky to construct couplings that give good bounds on mixing time. Path coupling, introduced in [BD97], is a useful tool for simplifying this process: rather than define $(\sigma, \tau) \mapsto (\sigma', \tau')$ for all $(\sigma, \tau) \in \Omega^2$, it is enough to do so for a small subset of initial pairs in Ω^2. This subset is specified by a *pre-metric*.

---

[5]For each color c not in the neighborhood of v, there exists a distinct component $S_\sigma(v, c) = \{v\}$ in $\mathcal{S}_\sigma$.
[6]Note that one must set $p_1 = 1$ because otherwise, by rescaling all flip parameters by a factor of $1/p_1$, the mixing time simply scales by a factor of $1/p_1$.

Definition 2.2. A *pre-metric* on Ω is a pair (Γ, ω) where Γ is a connected, undirected graph with vertex set Ω, and ω is a function that assigns positive, real-valued weights to edges $\sigma\tau$ of Γ such that for every edge $\sigma\tau$, $\omega(\sigma\tau)$ is the minimum weight among all paths between σ and τ.

We will often refer to a pair of adjacent colorings σ, τ in Γ as a *neighboring coloring pair*, denoted by (G, σ, τ). Where the context is clear, we omit G and refer to neighboring coloring pairs as (σ, τ).

For any $\tilde{\sigma}, \tilde{\tau} \in \Omega$, let $P_{\tilde{\sigma}, \tilde{\tau}}$ denote the set of simple paths $\phi = (\phi_0, \ldots, \phi_s)$ in Γ where $\phi_0 = \tilde{\sigma}$ and $\phi_s = \tilde{\tau}$. The metric d induced by the pre-metric (Γ, ω) is defined by $d(\tilde{\sigma}, \tilde{\tau}) := \min_{\phi \in P_{\tilde{\sigma}, \tilde{\tau}}} \sum_{i=1}^{s} \omega(\phi_{i-1}\phi_i)$.

Example 2.1. *If Γ is the graph with vertex set Ω and edges between colorings σ, τ which differ on exactly one vertex, and ω assigns weight 1 to all edges of Γ, then the metric induced by (Γ, ω) is simply the Hamming distance d_H, i.e. the total number of vertices on which two colorings differ.*

Theorem 2.1 ([BD97]). *Let (Γ, ω) be a pre-metric on Ω where ω takes on values in $(0, 1]$, and let d be the metric it induces. If the Markov chain $\sigma \mapsto \sigma'$ has a coupling $(\sigma, \tau) \mapsto (\sigma', \tau')$ defined for all neighboring coloring pairs (σ, τ) that $(1 - \alpha)$-contracts for some $\alpha > 0$, then there exists a coupling defined for* all *pairs of colorings $(\sigma, \tau) \in \Omega^2$ which $(1 - \alpha)$-contracts.*

Remark 2.1. *The reason we need to extend the state space from Ω^* to Ω is already apparent in the context of path coupling with the Hamming metric. Given two colorings σ, τ for which $d_H(\sigma, \tau) = \ell$, there does not necessarily exist a sequence of* proper *colorings $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_\ell = \tau$ of length ℓ for which (σ_i, σ_{i+1}) are neighboring colouring pairs for all $0 \leq i < \ell$. However, there certainly exist such sequences if we allow the colorings to be improper.*

2.3 Variable-Length Coupling

In this section we review the basics of variable-length coupling. This will only be relevant to Section 4 which gives the first of our two proofs of Theorem 1.1.

Jerrum's $k > 2\Delta$ bound [Jer95] and Vigoda's $k > \frac{11}{6}\Delta$ bound [Vig99] can both be proved via *one-step* path couplings. Yet there is evidence that multi-step couplings can sometimes be stronger than one-step couplings. As shown in [KR01], there exist Markov chains for some sampling problems where one-step coupling analysis is insufficient. [CKKL99] used multi-step coupling for a tighter analysis of a Markov chain for sampling random permutations, and the celebrated $k \geq (1 + \epsilon)\Delta$ result of [HV03] for $\Omega(\log n)$-degree graphs uses a multi-step coupling which is constructed by looking into future time steps.

There are also several other works that carried out a multi-step coupling analysis by looking at one-step coupling over multiple time steps [DGG+01, DGM02, HV07] and obtained slight improvements over Jerrum's $k > 2\Delta$ bound by terminating path coupling of the Glauber dynamics at a random stopping time. In the literature, this is known as *variable-length coupling*, and this is the approach we take in Section 4, but for the flip dynamics.

Definition 2.3. [Definition 1 in [HV07]] For every initial neighboring coloring pair $(\sigma^{(0)}, \tau^{(0)})$, let $(\bar{\sigma}, \bar{\tau}, T_{\text{stop}})$ be a random variable where T_{stop} is a nonnegative integer, and $\bar{\sigma}, \bar{\tau}$ are sequences of colorings $(\sigma^{(0)}, \ldots, \sigma^{(T_{\text{stop}})}) \in \Omega^{T_{\text{stop}}}$ and $(\tau^{(0)}, \ldots, \tau^{(T_{\text{stop}})}) \in \Omega^{T_{\text{stop}}}$ respectively. We say that $(\bar{\sigma}, \bar{\tau}, T_{\text{stop}})$ is a *variable-length path coupling* if $\bar{\sigma}, \bar{\tau}$ are faithful copies of the Markov chain in the following sense.

For $(\sigma^{(0)}, \tau^{(0)})$ and $t \geq 0$, define the random variables σ_t, τ_t via the following experiment: 1) sample $(\bar{\sigma}, \bar{\tau}, T_{\text{stop}})$, 2) if $t \leq T$, define $\sigma_t = \sigma^{(t)}, \tau_t = \tau^{(t)}$, 3) if $t > T_{\text{stop}}$, then sample σ_t and τ_t from $P^{t-T_{\text{stop}}}(\sigma^{(T_{\text{stop}})}, \cdot)$ and $P^{t-T_{\text{stop}}}(\tau^{(T_{\text{stop}})}, \cdot)$ respectively.

We say that $\bar{\sigma}$ (resp. $\bar{\tau}$) is a *faithful copy* if for every neighboring coloring pair $(\sigma^{(0)}, \tau^{(0)})$ and $t \geq 0$, σ_t and τ_t defined above are distributed according to $P^t(\sigma^{(0)}, \cdot)$ and $P^t(\tau^{(0)}, \cdot)$ respectively.

When T_{stop} is always equal to some fixed T, this is just the usual notion of T-step path coupling. The following extends the path coupling theorem of [BD97] to variable-length path coupling.

Theorem 2.2 (Corollary 4 of [HV07]). *Let $\epsilon > 0$. For a variable-length path coupling $(\bar{\sigma}, \bar{\tau}, T_{stop})$, let*

$$\alpha := 1 - \max_{\sigma^{(0)}, \tau^{(0)}} \mathbb{E}[d_H(\sigma^{(T_{stop})}, \tau^{(T_{stop})})], \quad W := \max_{\sigma^{(0)}, \tau^{(0)}, t \leq T_{stop}} d_H(\sigma^{(t)}, \tau^{(t)}), \quad \beta := \max_{\sigma^{(0)}, \tau^{(0)}} \mathbb{E}[T_{stop}].$$

If $\alpha > 0$, then $\tau_{mix}(\epsilon) \leq 2 \lceil 2\beta W/\alpha \rceil \cdot \lceil \ln(n/\epsilon)/\alpha \rceil$.

2.4 Vigoda's Greedy Coupling

In Appendix A, we give a self-contained review of the one-step path coupling analysis from [Vig99]. We encourage readers unfamiliar with his analysis to simply read the entirety of Appendix A in place of this subsection, as here we only introduce relevant notation and state the parts of his analysis that are essential to our proofs.

Fix a neighboring coloring pair (G, σ, τ). For $c \in [k]$, let U_c denote the set of neighbors of v (in either coloring) that are colored c, and let $\delta_c = |U_c|$. We will sometimes denote the vertices in U_c by $\{u_1^c, \ldots, u_{\delta_c}^c\}$; where c is clear from context, we will simply denote these by $\{u_1, \ldots, u_{\delta_c}\}$.

Note that the symmetric difference $\mathcal{D} = \mathcal{S}_\sigma \Delta \mathcal{S}_\tau$ is precisely the Kempe components $S_\sigma(u_i^c, \tau(v))$ and $S_\sigma(v, c)$ in σ and the Kempe components $S_\tau(u_i^c, \sigma(v))$ and $S_\tau(v, c)$ in τ, for all colors c appearing in the neighborhood of v and all $i \in [\delta_c]$. All other Kempe components are shared between σ and τ, so for those, it is enough to use the identity coupling. Note that for colors $c \neq \sigma(v), \tau(v)$ not appearing in $N(v)$, the identity coupling then matches the flip of $S_\sigma(v, c)$ to that of $S_\tau(v, c)$ so that the two colorings of G become identical.

So the main concern is how to couple the flips of components in $\mathcal{D}$. It is easy to see that for $c \neq \sigma(v), \tau(v)$,

$$S_\sigma(v, c) = \left(\bigcup_{i=1}^{\delta_c} S_\tau(u_i^c, \sigma(v)) \right) \cup \{v\} \quad S_\tau(v, c) = \left(\bigcup_{i=1}^{\delta_c} S_\sigma(u_i^c, \tau(v)) \right) \cup \{v\}, \tag{3}$$

Purely for simplicity of exposition, we will assume that the sets $S_\sigma(u_i^c, \tau(v))$ are distinct as i varies, and likewise for $S_\tau(u_i^c, \sigma(v))$, and we will only consider $c \neq \sigma(v), \tau(v)$, referring the reader respectively to Remarks A.1 and A.2 in Appendix A for the missing details. We remark that the extra cases of $c = \sigma(v), \tau(v)$ are the primary place where one needs to be careful about the fact that neighboring coloring pairs σ, τ need not be proper.

For c such that $\delta_c > 0$, define $A_c := |S_\sigma(v, c)|$, $B_c := |S_\tau(v, c)|$, $a_i^c = |S_\tau(u_i^c, \sigma(v))|$, and $b_i^c = |S_\sigma(u_i^c, \tau(v))|$. Define the vectors $\mathbf{a}^c := (a_i^c : i \in [\delta_c])$ and $\mathbf{b}^c := (b_i^c : i \in [\delta_c])$. We say that a neighboring coloring pair (G, σ, τ) has *configuration* $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ *of size* δ_c for color c. Also define $a_{max}^c = \max_i a_i^c$ and denote a maximizing i by i_{max}^c. Likewise define $b_{max}^c = \max_j b_j^c$ and denote a maximizing j by j_{max}^c. When the color c is clear from context, we refer to these as $A, B, a_i, b_i, \mathbf{a}, \mathbf{b}, a_{max}, i_{max}, b_{max}, j_{max}$.

When $c \neq \sigma(v), \tau(v)$, note that

$$A_c = 1 + \sum_i a_i^c, \qquad B_c = 1 + \sum_i b_i^c. \tag{4}$$

While $S_\sigma(v, c)$ and $S_\tau(v, c)$ can be quite different, we do know that $S_\sigma(v, c) \supset S_\tau(u_i, \sigma(v))$. The idea of Vigoda's coupling is thus to greedily couple the flips of the largest components, i.e. $S_\sigma(v, c), S_\tau(v, c)$,

8

to the flips of the next largest components, i.e. $S_\tau(u_{i_{max}}, \sigma(v)), S_\sigma(u_{j_{max}}, \tau(v))$, and then to couple together as closely as possible the flips of $S_\sigma(u_i, \tau(v))$ and $S_\tau(u_i, \sigma(v))$ for each $i \in [\delta_c]$. Henceforth, we will refer to this coupling as the *greedy coupling*.

For any configuration $(A, B; \mathbf{a}, \mathbf{b})$, define

$$H(A, B; \mathbf{a}, \mathbf{b}) := (A - a_{max} - 1)p_A + (B - b_{max} - 1)p_B + \sum_i a_i q_i + b_i q_i' - \min(q_i, q_i'), \quad (5)$$

where $q_i = p_{a_i} - p_A \cdot \mathbb{1}_{i=i_{max}}$ and $q_i' = p_{b_i} - p_B \cdot \mathbb{1}_{i=j_{max}}$. In [Vig99] it is shown that under this greedy coupling, for $c \neq \sigma(v), \tau(v)$ appearing in the neighborhood of v,

$$kn \cdot \mathbb{E}[\mathbb{1}_{X_c} \cdot (d_H(\sigma', \tau') - 1)] \leq H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c), \quad (6)$$

where X_c is the event that the coupling flips components in $\mathcal{D}_c$ in both colorings. Therefore:

Lemma 2.2 ([Vig99]). *Let $(\sigma, \tau) \mapsto (\sigma', \tau')$ be the greedy coupling. Then*

$$\mathbb{E}[d_H(\sigma', \tau') - 1] \leq \frac{1}{nk} \left(-|\{c : \delta_c = 0\}| + \sum_{c:\delta_c \neq 0} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) \right). \quad (7)$$

The function H implicitly depends on the choice of flip parameters $\{p_\alpha\}$, while $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ depends on (G, σ, τ). The remaining analysis in [Vig99] once (7) has been deduced essentially boils down to picking a good set of flip parameters.

3 Linear Programming and Choice of Flip Parameters

The key idea to choose the flip parameters is to cast the problem as an instance of linear programming with variables $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ to minimize the right-hand side of (7) over all neighboring coloring pairs (G, σ, τ) where G has maximum degree Δ and σ, τ are k-colorings. The following gives terminology for quantifying over all such (G, σ, τ).

Definition 3.1. A configuration $(A, B; \mathbf{a}, \mathbf{b})$ is *realizable* if there exists a neighboring coloring pair (G, σ, τ) and color c such that $(A, B; \mathbf{a}, \mathbf{b}) = (A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$.

Vigoda's remaining analysis can thus be interpreted as solving the following linear program.

Linear Program 1. *For variables $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ and λ, minimize λ subject to: $p_0 = 0 \leq p_\alpha \leq p_{\alpha-1} \leq p_1 = 1$ for all $\alpha \geq 2$, and $H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + \lambda \cdot m$ for all realizable $(A, B; \mathbf{a}, \mathbf{b})$ of size m.*

There are three minor issues with this linear program: (*a*) the linear program has an infinite number of variables, (*b*) it has an infinite number of constraints, and (*c*) given $\mathbf{a}, \mathbf{b}$, it is not immediately obvious how to enumerate all A, B for which $(A, B; \mathbf{a}, \mathbf{b})$ is realizable.

Vigoda handles (*a*) by restricting to flips of components of size at most N_{max}, i.e. by fixing some small constant N_{max} and insisting that

$$p_\alpha = 0 \ \forall \alpha > N_{max}. \quad (8)$$

We emphasize that this still leaves an infinite number of constraints as m can be unbounded.

He handles (*b*) by shrinking the feasible region via the following observation.

Lemma 3.1 ([Vig99]). $H(A, B; \mathbf{a}, \mathbf{b}) \leq (A - 2)p_A + (B - 2)p_B + \sum_i (a_i p_{a_i} + b_i p_{b_i} - \min(p_{a_i}, p_{b_i}))$.

9

Whereas $f(u_i)$ are linear functions of $p_{a_i}, p_{b_i}, p_A, p_B$, the summands in the upper bound of Lemma 3.1 are simply linear functions of p_{a_i}, p_{b_i}. So we can pick some m^* ([Vig99] picks $m^* = 3$) and replace the infinitely many constraints for which $m \geq m^*$ in Linear Program 1 with finitely many constraints to optimize the right-hand side of Lemma 3.1.

Finally, Vigoda implicitly handles (c) as follows. To cover all constraints corresponding to realizable $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ with $c \neq \sigma(v), \tau(v)$ include

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + \lambda \cdot m, \tag{9}$$

for all $1 \leq m < m^*$ and $(A, B; \mathbf{a}, \mathbf{b})$ for which $\mathbf{a}, \mathbf{b} \in \{0, 1, \ldots, N_{max}\}^m \setminus \{(0, 0, \ldots 0)\}$ and A, B satisfy (4). As we will discuss in Appendix A.2, to cover all constraints corresponding to $c = \sigma(v)$ and $c = \tau(v)$, it is enough to include

$$(B - b_m)p_B + \sum_{i=1}^{m-1} b_i p_{b_i} \leq -1 + \lambda \cdot m, \tag{10}$$

for all $2 \leq m < m^*$, $0 \leq b_1 \leq \cdots \leq b_m \leq N_{max}$ where $b_m > 0$, and $B = \sum_i b_i$, as well as

$$\alpha \cdot p_\alpha \leq 1, \quad \text{for all } \alpha \in \mathbb{N}_0. \tag{11}$$

Concretely, we have the following relaxation of Linear Program 1.

Linear Program 2. *Fix some $N_{max} \geq 1$ and $m^* \geq 2$. For variables $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ and λ, and dummy variables x, y, minimize λ subject to the following constraints: $p_0 = 0 \leq p_\alpha \leq p_{\alpha-1} \leq p_1 = 1$ for all $\alpha \geq 2$, constraint (8), constraint (9) for all $\mathbf{a}, \mathbf{b} \in \{0, 1, \ldots, N_{max}\}^m \setminus \{(0, \ldots, 0)\}$ with $1 \leq m < m^*$ and A, B satisfying (4), constraint (10) for all $2 \leq m < m^*$ and $0 \leq b_1 \leq \cdots \leq b_m \leq N_{max}$ where $b_m > 0$ and $B = \sum_i b_i$, constraint (11), and constraints*

$$\begin{aligned} x &\geq (A-2)p_A \\ y &\geq a \cdot p_a + b \cdot p_b - \min(p_a, p_b) \\ -1 + \lambda \cdot m^* &\geq 2x + m^* \cdot y \end{aligned} \tag{12}$$

for every A, a, b satisfying $0 \leq A \leq 1 + N_{max}$ and $0 \leq a < b \leq N_{max}$.

Remark 3.1. *Note that we only add in constraints for the upper bound of Lemma 3.1 in the case of $m = m^*$ (constraint (12)) because the constraints for $m = m^*$ implies the constraints for $m > m^*$.*

Lemma 3.2. *Let λ_2^* be the objective value of Linear Program 2. If $\lambda_2^* \geq 1$ and $k > \lambda_2^* d$, then there exist flip parameters $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ for which $\mathbb{E}[d_H(\sigma', \tau') - 1] < 0$ for all neighboring coloring pairs (G, σ, τ), where $(\sigma, \tau) \mapsto (\sigma', \tau')$ is the greedy coupling.*

Proof. Let $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ be flip parameters achieving objective value λ_2^*. By (7) and Lemma 3.1, we have that

$$nk \cdot \mathbb{E}[d_H(\sigma', \tau') - 1] \leq -|\{c : \delta_c = 0\}| + \sum_{c : \delta_c \neq 0} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) \leq -k + \lambda_2^* \cdot d < 0. \qquad \square$$

In [Vig99], Vigoda shows that for $m^* = 3$, $N_{max} = 6$, and the following flip parameters, Linear Program 2 attains a value $\lambda_2^* = 11/6$:

$$p_1 = 1, \ p_2 = 13/42, \ p_3 = 1/6, \ p_4 = 2/21, \ p_5 = 1/21, \ p_6 = 1/84 \text{ and } p_\alpha = 0 \ \forall \alpha \geq 7. \tag{13}$$

Not only is this assignment a feasible solution to Linear Program 2, but it happens to be an optimal solution of Linear Programs 1 and 2. We will be interested in which constraints are tight under such optimal solutions, as they will guide us to the configurations key to our proofs.

Definition 3.2. Given a feasible solution $\mathbf{p}$ of Linear Program 1 with objective value λ, a configuration $(A, B; \mathbf{a}, \mathbf{b})$ of size m is $\mathbf{p}$-*extremal* (or simply *extremal* if the flip parameters are clear from the context) if $H(A, B; \mathbf{a}, \mathbf{b}) = -1 + \lambda \cdot m$ under the assignation $\mathbf{p}$.

Vigoda's proof in [Vig99] already implicitly gives a collection of six extremal configuration under the assignment (13) (see Observation B.1 in Appendix B). It turns out that among these tight constraints, two of them are already enough to force the objective value of Linear Program 2 to be 11/6.

Consider the following linear program obtained by restricting to constraints (9) associated with configurations $(3, 2; (2), (1))$ and $(7, 3; (3, 3), (1, 1))$, which are both realizable.

Linear Program 3. *For variables $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ and λ, minimize λ subject to: $p_0 = 0 \le p_\alpha \le p_{\alpha-1} \le p_1 = 1$ and*

$$p_1 + p_2 - 2p_3 - \min(p_1 - p_2, p_2 - p_3) \le -1 + \lambda,$$
$$2p_1 + 5p_3 - \min(p_1 - p_3, p_3 - p_7) \le -1 + 2\lambda.$$

It is easy to check that Linear Program 3 also has objective value 11/6, and its constraints are a strict subset of those of Linear Programs 1 and 2, from which we conclude that

Corollary 3.1. *The objective values of Linear Program 1, Linear Program 2 with $N_{max} \ge 6$ and $m^* = 3$, and Linear Program 3 are all equal to 11/6.*

Corollary 3.1 allows us to exhibit a family $\mathcal{C}$ of just *two* neighboring coloring pairs (G, σ, τ) for which no one-step coupling, greedy or otherwise, simultaneously contracts with respect to the Hamming metric for $k < (11/6)\Delta$ for all $(G, \sigma, \tau) \in \mathcal{C}$. To clarify, this lemma is not used in the proofs of our main result, but provides intuition for the limitations of one-step coupling with respect to the Hamming metric and motivates our two approaches for circumventing them.
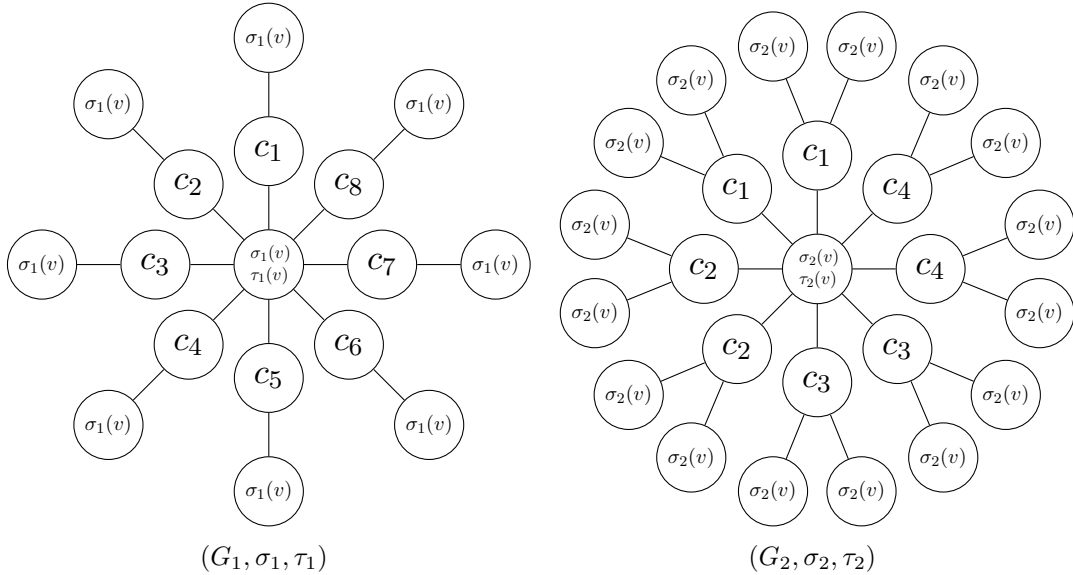

Figure 1: Neighboring coloring pairs defined in Construction 3.1 for $\Delta = 8$.

11

Construction 3.1. *Let G_1 be the tree of height 2 rooted at vertex v with Δ path graphs, each consisting of 2 other vertices, attached to v. Let $u_1, \ldots, u_\Delta$ be the neighbors of v. In colorings σ_1, τ_1, assign two different colors to v, $\sigma_1(v)$ and $\tau_1(v)$, assign each u_i the color $c_i \neq \sigma_1(v), \tau_1(v)$, and assign the child of u_i the color $\sigma_1(v)$.*

For Δ even, let G_2 be the tree of height two rooted at a vertex v with exactly Δ children $u_1, \ldots, u_\Delta$ such that each u_i has exactly two children w_1^i and w_2^i. In colorings σ_2, τ_2, assign two different colors to v, $\sigma_2(v)$ and $\tau_2(v)$, assign u_{2j-1} and u_{2j} the color $c_j \neq \sigma_2(v), \tau_2(v)$ for $j \in \{1, \ldots, \Delta/2\}$, and assign all w_ℓ^i the color $\sigma_2(v)$.

Let $\mathcal{C}^* = \{(G_1, \sigma_1, \tau_1), (G_2, \sigma_2, \tau_2)\}$ (see Figure 1).

Lemma 3.3. *If $k < (11/6)\Delta$, there exists no choice of flip parameters $\{p_\alpha\}$ and one-step coupling $(\sigma, \tau) \mapsto (\sigma', \tau')$ for which $\mathbb{E}[d_H(\sigma', \tau') - 1] < 0$ for all $(G, \sigma, \tau) \in \mathcal{C}^*$, where $\mathcal{C}^*$ is defined in Construction 3.1.*

We defer the proof of this to Appendix B. Lemma 3.3 states that it is impossible to design a one-step coupling analysis of the flip dynamics with the Hamming metric that crosses Vigoda's 11/6 barrier for general graphs (or even for the family of two trees defined by $\mathcal{C}^*$). There are two natural strategies to overcome this problem: use a *multi-step coupling* analysis with the Hamming metric, or use a one-step coupling with an alternative metric. In the remainder of the paper we present two independent proofs of Theorem 1.1. In Section 4, we present the multi-step coupling approach to the proof as devised by Chen and Moitra [CM18]. In Section 5, we use the alternative metric approach to the proof as devised by Delcourt, Perarnau and Postle [DPP18]. For the sake of clarity, we defer the proof of some technical lemmas to Appendices C and D.

4 Proof of Theorem 1.1 using Variable-Length Coupling

In this section, we prove Theorem 1.1 by arguing that a suitably defined variable-length coupling contracts with respect to the Hamming metric d_H. Recall from Definition 2.2 that under the pre-metric inducing the Hamming metric, (G, σ, τ) is a *neighboring coloring pair* if σ, τ are colorings of G that differ on exactly one vertex v.

4.1 Modifying the LP

For a color c, the condition that $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ for a neighboring coloring pair (G, σ, τ) be extremal is a very stringent condition on (G, σ, τ). The hope is that for a suitable notion of "typical," this condition holds for few colors c for a "typical" neighboring coloring pair.

Definition 4.1. Given a neighboring coloring pair (G, σ, τ) and a color c appearing in the neighborhood of v, then the pair σ, τ is in the state

1. SING_c if $\delta_c = 1$ and $c \neq \sigma(v), \tau(v)$[7]

2. BAD_c if $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ is either $(7, 3; (3, 3), (1, 1))$ or $(3, 7; (1, 1), (3, 3))$

3. GOOD_c otherwise.

---

[7]For $c \neq \sigma(v), \tau(v)$, SING_c corresponds to the extremal configuration of size 1, as well as all other configurations which satisfy $\delta_c = 1$. We include these latter configurations just for simplicity of analysis; if we did not do this, it would yield additional improvements upon our main result. Also note that we exclude the cases of $c = \sigma(v), \tau(v)$ from SING_c because in those cases, we have that $A_c = 0$ in which case $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ is not extremal.

Moreover, define $N_{sing}(\sigma, \tau)$, $N_{bad}(\sigma, \tau)$, and $N_{good}(\sigma, \tau)$ to be the number of c for which (G, σ, τ) is in state SING_c, BAD_c, GOOD_c respectively.

Observation 4.1. *Let (G, σ, τ) be any neighboring coloring pair. For $c = \sigma(v), \tau(v)$, if $\delta_c > 0$, then σ, τ are in state GOOD_c.*

Proof. If $\delta_c = 1$, then by definition σ, τ are in state GOOD_c. If $\delta_c \geq 2$, then because $A_c = 0$ for $c = \sigma(v), \tau(v)$ by Remark A.2, σ, τ must be in state GOOD_c. □

$N_{sing}(\sigma, \tau)$ can be large even for a "typical" neighboring coloring: consider any (G, σ, τ) where the neighbors of v form a clique. Indeed, this example is the reason that all existing results on sampling colorings that proceeded [Vig99] needed to at least assume triangle-freeness of G, otherwise the local uniformity properties they leverage simply do not hold. Instead of avoiding state SING_c, we want "typical" neighboring coloring pairs to avoid BAD_c for many c. Specifically, we want $N_{bad}(\sigma, \tau)$ to be at most a constant times $N_{good}(\sigma, \tau)$.

Remark 4.1. *At this point the reader may be wondering: why can we get away with only analyzing what happens to the extremal configurations of size 2 and not those of size 1? The reason is that neighboring coloring pairs where v is surrounded by many configurations of size 1 are precisely the kinds of examples on which Vigoda's analysis does particularly well: under Jerrum's maximal coupling, it is impossible to go beneath $k > 2\Delta$ for any one-step coupling of the Glauber dynamics with respect to the Hamming metric, but under Vigoda's greedy coupling of the flip dynamics, one can perfectly couple the flips of Kempe components in $\mathcal{D}_c$ for any c with $\delta_c = 1$ if N_{max} is big enough. The reason Vigoda's analysis doesn't get all the way down to $k > (1 + \epsilon)\Delta$ is simply that if N_{max} is too big and the flip parameters too tuned to configurations of size 1, one cannot closely couple flips corresponding to configurations of size at least 2. For this reason, what we really care about is actually the fraction of configurations of size at least 2 around v that are extremal.*

Consider the following thought experiment. Let $\mathcal{C}$ consist of all neighboring coloring pairs such that for every $(G, \sigma, \tau) \in \mathcal{C}$,

$$N_{bad}(\sigma, \tau) \leq \gamma \cdot N_{good}(\sigma, \tau) \tag{14}$$

for some absolute constant $\gamma > 0$. Suppose $k = (11/6 - \epsilon)\Delta$ for some small absolute constant $\epsilon > 0$, and our goal is just to pick flip parameters so every pair in $\mathcal{C}$ contracts. Observation B.1 and complementary slackness intuitively suggest that this should be possible for small enough ϵ depending only on γ. To get an effective estimate for ϵ, we encode (14) into Linear Program 2:

Linear Program 4. *Introduce into Linear Program 2 the additional variables $\lambda_{sing}, \lambda_{bad}, \lambda_{good}$. In constraints (12) and (10), replace λ with λ_{good}. In the constraints (9) corresponding to configuration $(A, B; \mathbf{a}, \mathbf{b})$, replace λ with λ_{sing} if $m = 1$, λ_{bad} if $(A, B; \mathbf{a}, \mathbf{b})$ is either $(7, 3; (3, 3), (1, 1))$ or $(3, 7; (1, 1), (3, 3))$, or λ_{good} otherwise. Lastly, introduce the constraints*

$$\lambda \geq \lambda_{sing}, \quad \lambda \geq \lambda_{good}, \quad \lambda \geq \frac{\gamma}{\gamma + 1} \cdot \lambda_{bad} + \frac{1}{\gamma + 1} \cdot \lambda_{good}.$$

Call this the γ-mixed coupling LP and denote its objective value by λ_γ^*.

The following is straightforward to prove; see Appendix C.1 for a formal proof.

Lemma 4.1. *If $k > \lambda_\gamma^* \Delta$, then under the greedy coupling $(\sigma, \tau) \mapsto (\sigma', \tau')$, $\mathbb{E}[d_H(\sigma', \tau') - 1] < 0$ for any $(G, \sigma, \tau) \in \mathcal{C}$.*

Next, we go from the intuition of this thought experiment to a rigorous notion of "typical" neighboring coloring pairs avoiding the state BAD_c. Having already reduced finding a coupling for all of $\mathcal{C}$ to analyzing the γ-mixed coupling LP, in the sequel we will reduce finding a coupling for *all* neighboring coloring pairs to analyzing the γ-mixed coupling LP.

4.2 Variable-Length Coupling

The key idea is that regardless of what neighboring coloring pair (G, σ, τ) one starts with, the probability that σ', τ' derived from one step of greedy coupling has changed in distance is $\Theta(1/n)$ (see Lemma 4.2 below). So in expectation, one can run $\Theta(n)$ steps of greedy coupling before the two colorings either coalesce to the same coloring or have Hamming distance greater than 1, but by that time the set of colors around v will have changed substantially. This is the main insight of [DGG$^+$01, HV07], who leverage it to analyze the Glauber dynamics and slightly improve upon Jerrum's $k \geq 2d$ bound under extra girth and degree assumptions. We leverage this insight to analyze the flip dynamics under no extra assumptions.

Our variable-length coupling simply runs greedy coupling until the distance between the colorings changes: start with neighboring colorings $\sigma^{(0)}, \tau^{(0)}$, initialize $t = 1$, and repeat the following.

1. Run the greedy one-step coupling of Section 2.4 to flip components S_t in $\sigma^{(t-1)}$ and S_t' in $\tau^{(t-1)}$, producing $\sigma^{(t)}, \tau^{(t)}$ (note that S_t or S_t' might be empty, e.g with probability $1 - p_\alpha$, a component of size α that is chosen to be flipped is not actually flipped).

2. If $d_H(\sigma^{(t)}, \tau^{(t)}) \neq d_H(\sigma^{(t-1)}, \tau^{(t-1)})$, terminate and define $T_{\text{stop}} = t$. Else, increment t.

We call any subsequence of pairs of flips $(S_i, S_i'), \ldots, (S_j, S_j')$ a *coupling schedule* starting from the neighboring coloring pair $(G, \sigma^{(i-1)}, \tau^{(i-1)})$.

It is easy to see that this satisfies the conditions of being a variable-length coupling as in Definition 2.3. Indeed it is the same coupling as in [HV07], except for the flip dynamics instead of the Glauber dynamics. Note that we can characterize which pairs of flips (S_t, S_t') terminate the coupling: at least one of them must belong to the symmetric difference $\mathcal{D}$ defined in Section 2.4.

Definition 4.2. Given a neighboring coloring pair (G, σ, τ), a pair of components S in σ and S' in τ is *terminating* if $S = S_\sigma(v, c)$ or $S' = S_\tau(v, c)$, or there exists $u \in N(v)$ for which $S = S_\sigma(u, \tau(v))$ or $S' = S_\tau(u, \sigma(v))$.

Note that for any t, S_t and/or S_t' may be the empty set. Moreover, because flips of components outside of $\mathcal{D}$ are matched via the identity coupling, if (S_t, S_t') is not terminating, then $S_t = S_t'$.

Lemma 4.2. *Let components S in σ and S' in τ be chosen according to the greedy coupling. Then*

$$\frac{k - \Delta - 2}{nk} \leq \mathbb{P}[(S, S') \text{ terminating}] \leq \frac{k + 2p_2\Delta}{nk},$$

where p_2 is the flip parameter for components of size 2.

Proof. For the lower bound, note that the pair $(S_\sigma(v, c), S_\tau(v, c))$ is terminating for any c. In particular, for $c \neq \{\sigma(v), \tau(v)\}$ such that $\delta_c = 0$, $S_\sigma(v, c) = S_\tau(v, c) = \{v\}$ — note that while the vertex sets for these components are all $\{v\}$, the flips are all distinct as they vary with c. Each such pair of flips has probability mass $(1/nk) \cdot p_1 = (1/nk)$, and there are at least $k - \Delta - 2$ such choices of c, giving the lower bound. We defer the proof of the upper bound to Appendix C.2. □

Corollary 4.1. $\max_{\sigma^{(0)}, \tau^{(0)}} \mathbb{E}[T_{stop}] \leq \frac{nk}{k - \Delta - 2}$.

We now give a reduction from analyzing the expected change in distance under our variable-length coupling to proving that the relation (14) from our thought experiment holds in expectation by the end of the coupling.

14

Lemma 4.3. *Suppose there exists a constant $\gamma > 0$ for which*

$$\mathbb{E}[N_{bad}(\sigma^{(T_{stop}-1)}, \tau^{(T_{stop}-1)})] \leq \gamma \cdot \mathbb{E}[N_{good}(\sigma^{(T_{stop}-1)}, \tau^{(T_{stop}-1)})] \tag{15}$$

for any initial neighboring coloring pair $(G, \sigma^{(0)}, \tau^{(0)})$.

Let $\lambda^*_{C\gamma}$ be the objective value of the $C\gamma$-mixed coupling LP, where $C := \frac{k+2p_2\Delta}{k-\Delta-2}$. Then

$$\mathbb{E}[d_H(\sigma^{(T_{stop})}, \tau^{(T_{stop})}) - 1] \leq \frac{-k + \lambda^*_{C\gamma}\Delta}{k - \Delta - 2}.$$

This mainly just follows by linearity of expectation and the calculation done in the proof of Lemma 4.1. The only complication is that the probability that the coupling terminates at any given point is not fixed, but this is fine because it is still the same *up to constant factors*, which only leads to the loss of a factor of C as defined in the lemma. We defer the full proof to Appendix C.3.

Remark 4.2. *(15) is one way to say that a typical coloring has few extremal configurations around v. This is slightly different from the analogous notion in the second proof of Theorem 1.1 in Section 5. There, the goal is to show that the number of extremal configurations around v of size 1 or 2 goes down in one step of Vigoda's greedy coupling. In contrast, we only choose to upper bound the fraction of configurations of size at least 2 at the end of our variable-length coupling which are extremal. That we don't attempt to analyze how extremal components of size 1 break apart is purely out of technical convenience, and as we discuss in Remark 4.1, still enough to break the $\frac{11}{6}$ barrier.*

4.3 Few Extremal Configurations When Coupling Terminates

We are left with proving the main technical lemma that (15) holds. Throughout this subsection, assume that $k \geq 1.833\Delta$.

Lemma 4.4. *Suppose flip parameters $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ satisfy $p_0 = 0 \leq p_\alpha \leq p_{\alpha-1} \leq p_1 = 1$ for all $\alpha \geq 2$, constraint (11), and additionally $\alpha p_{\alpha-2} \leq 3$ for all $\alpha \geq 3$. Then for $\gamma := \frac{(6k-\Delta-2)(k+2p_2\Delta)}{4(k-\Delta-2)(k-\Delta-1)}$, we have that (15) holds for any initial neighboring coloring pair $(G, \sigma^{(0)}, \tau^{(0)})$.*

Remark 4.3. *The additional constraint that $\alpha p_{\alpha-2} \leq 3$ already holds for the solutions to the γ-mixed coupling LP, so we assume it just to obtain better constant factors in our analysis.*

We first make a simple reduction. Fix any initial neighboring coloring pair $(G, \sigma^{(0)}, \tau^{(0)})$, and for every color c denote by $p_{bad}(c)$ and $p_{good}(c)$ the probability that $\sigma^{(T_{stop})}, \tau^{(T_{stop})}$ is in state BAD_c and $GOOD_c$, respectively.

By linearity of expectation we have that

$$\mathbb{E}[N_{bad}(\sigma^{(T_{stop})}, \tau^{(T_{stop})})] = \sum_c p_{bad}(c), \quad \mathbb{E}[N_{good}(\sigma^{(T_{stop})}, \tau^{(T_{stop})})] = \sum_c p_{good}(c).$$

Therefore to show Lemma 4.4, it is enough to show the following.

Lemma 4.5. *$p_{bad}(c) \leq \gamma \cdot p_{good}(c)$ for every color c.*

This is certainly true for $c = \sigma^{(0)}(v), \tau^{(0)}(v)$, in which case $p_{bad}(c) = 0$ and $p_{good}(c) = 1$ by Observation 4.1. We point out that while the case of $c = \sigma^{(0)}(v), \tau^{(0)}(v)$ is the one for which the fact that our state space includes all colorings, improper and proper, introduces complications in the definition of the greedy coupling (see Remark A.2), it happens to be the easiest case of Lemma 4.5.

So henceforth assume $c \neq \sigma^{(0)}(v), \tau^{(0)}(v)$. We proceed via a fractional matching argument. Take any coupling schedule $\Sigma_{pre} = (S_1, S_1), (S_2, S_2), \ldots, (S_{T-1}, S_{T-1})$ consisting of pairs of identical flips, and define $\mathcal{W}$ to be the set of all coupling schedules of the form $(S_1, S_1), (S_2, S_2), \ldots, (S_{T-1}, S_{T-1})$, (S_T, S_T') for (S_T, S_T') terminating. In other words, $\mathcal{W}$ consists of all T-step coupling schedules whose first $T-1$ steps are fixed to Σ_{pre} and which only changes the distance between the colorings in the last step (S_T, S_T'). We will match to the collection of schedules $\mathcal{W}$ an (infinite) collection of schedules of the following form.

Definition 4.3. Fix $S_1, \ldots, S_{T-1}$. A coupling schedule Σ_* starting from the neighboring coloring pair $(G, \sigma^{(0)}, \tau^{(0)})$ is *satisfying* if it is of the form

$$\Sigma_* = (S_1, S_1), \ldots, (S_{T-1}, S_{T-1}), (S_T^*, S_T^*), \ldots, (S_{T^*-1}^*, S_{T^*-1}^*), (S_{T^*}^*, S_{T^*}'^*) \tag{16}$$

for $(S_{T^*}^*, S_{T^*}'^*)$ terminating, and gives rise to a sequence of colorings

$$(\sigma^{(0)}, \tau^{(0)}), (\sigma^{(1)}, \tau^{(1)}), \ldots, (\sigma^{(T-1)}, \tau^{(T-1)}), (\sigma_*^{(T)}, \tau_*^{(T)}), \ldots, (\sigma_*^{(T^*)}, \tau_*^{(T^*)})$$

for which

1. $\sigma_*^{(T^*-1)}, \tau_*^{(T^*-1)}$ are in state GOOD_c

2. $\sigma_*^{(t)}, \tau_*^{(t)}$ is not in state BAD_c for any $T \leq t < T^*$.

In Definition 4.3, Property 2) ensures that from any satisfying Σ_*, we can uniquely decode the collection $\mathcal{W}$ to which it is being fractionally matched: in Σ_*, take the last pair of colorings in state BAD_c, and Σ_{pre} is the subsequence of Σ_* starting from $(\sigma^{(1)}, \tau^{(1)})$ and ending at that pair.

If we can show for any Σ_{pre} that

$$\sum_{\Sigma_* \text{ satisfying}} \mathbb{P}[(S_T^*, S_T^*), \ldots, (S_{T^*-1}^*, S_{T^*-1}^*), (S_{T^*}^*, S_{T^*}'^*) | \Sigma_{pre}] \geq \frac{1}{\gamma} \cdot \frac{k + 2p_2\Delta}{nk},$$

then this will imply that $p_{bad}(c) \leq \gamma \cdot p_{good}(c)$ because the upper bound of Lemma 4.2 tells us that $\mathbb{P}[(S_T, S_T') \text{ terminating}|\Sigma_{pre}] \leq \frac{k+2p_2\Delta}{nk}$.

To exhibit such a collection of satisfying coupling schedules Σ_*, we first define a coarsening of the state space as follows. Starting from the neighboring coloring pair $(G, \sigma^{(T-1)}, \tau^{(T-1)})$ which is in state BAD_c, take any subsequent coupling schedule

$$(S_T^*, S_T^*), \ldots, (S_{T^*}^*, S_{T^*}'^*) \tag{17}$$

with $(S_{T^*}^*, S_{T^*}'^*)$ terminating which gives rise to a sequence of pairs of colorings

$$(\sigma_*^{(T)}, \tau_*^{(T)}), \ldots, (\sigma_*^{(T^*)}, \tau_*^{(T^*)}), \tag{18}$$

where $(G, \sigma_*^{(t)}, \tau_*^{(t)})$ is a neighboring coloring pair for all t except $t = T^*$. Define the following auxiliary states. To avoid confusion with the states in Definition 4.1, we will refer to the auxiliary states defined below as *stages*.

16

Definition 4.4. Let c be any color, not necessarily one appearing in the neighborhood of v. We say that $\sigma_*^{(t)}, \tau_*^{(t)}$ is in *stage* GOODEND_c if $\sigma_*^{(t-1)}, \tau_*^{(t-1)}$ is in state GOOD_c and the pair of flips (S, S') giving rise to $\sigma_*^{(t)}, \sigma_*^{(t)}$ from $\sigma_*^{(t-1)}, \tau_*^{(t-1)}$ is terminating.

We say $\sigma_*^{(t)}, \tau_*^{(t)}$ is in *stage* BADEND_c if, intuitively, we choose to quit looking for satisfying coupling schedules among those of which $(\sigma_*^{(0)}, \tau_*^{(0)}), \ldots, (\sigma_*^{(t)}, \tau_*^{(t)})$ is a prefix. Formally, $\sigma_*^{(t)}, \tau_*^{(t)}$ is in stage BADEND_c if least one of the following conditions holds (note that these conditions aren't necessarily mutually exclusive):

(i) $t = T$ and the pair of flips (S, S') giving rise to $\sigma_*^{(T)}, \sigma_*^{(T)}$ from the initial pair $\sigma^{(T-1)}, \tau^{(T-1)}$ is terminating (i.e. if $(S_1, S_1), \ldots, (S_{T-1}, S_{T-1}), (S, S') \in \mathcal{W}$).

(ii) $t = T$ and $\sigma_*^{(t)}, \tau_*^{(t)}$ is not in state GOOD_c.

(iii) $\sigma_*^{(t-1)}, \tau_*^{(t-1)}$ is in state GOOD_c but $\sigma_*^{(t)}, \tau_*^{(t)}$ is not in state GOOD_c or stage GOODEND_c (this includes the case that c does not appear in the neighborhood of v in $\sigma_*^{(t)}, \tau_*^{(t)}$).

(iv) $t > T$ and $\sigma_*^{(t-1)}, \tau_*^{(t-1)}$ is in stage BADEND_c.

If $\sigma_*^{(t)}, \tau_*^{(t)}$ is not in stage BADEND_c or GOODEND_c and is in state BAD_c (resp. GOOD_c), then we say it is also in *stage* BAD_c (resp. *stage* GOOD_c).

Note that if a sequence of the form (18) contains a pair of colorings in stage GOODEND_c, that pair must be $\sigma_*^{(T^*)}, \tau_*^{(T^*)}$. Furthermore, given any sequence (18) for which $\sigma_*^{(T^*)}, \tau_*^{(T^*)}$ is in stage GOODEND_c with associated coupling schedule (17), note that the corresponding coupling schedule Σ_* defined in (16) is satisfying, by definition of stage BADEND_c.

So it is enough to show that if we start from a neighboring coloring pair $(G, \sigma^{(T-1)}, \tau^{(T-1)})$ which is in state BAD_c and evolve a sequence of pairs of colorings (18) according to the greedy coupling at each step, then

$$\mathbb{P}[\sigma_*^{(T^*)}, \tau_*^{(T^*)} \text{ are in stage } \text{GOODEND}_c | \sigma^{(T-1)}, \tau^{(T-1)}] \geq \frac{1}{\gamma} \cdot \frac{k + 2p_2\Delta}{nk}. \tag{19}$$

It remains to bound the probabilities of the transitions between the different stages of Definition 4.4 under the flip dynamics and the greedy coupling (see Figure 2 for a depiction of the transitions that can occur). A key point is that these bounds will be independent of the specific colorings or structure of G.
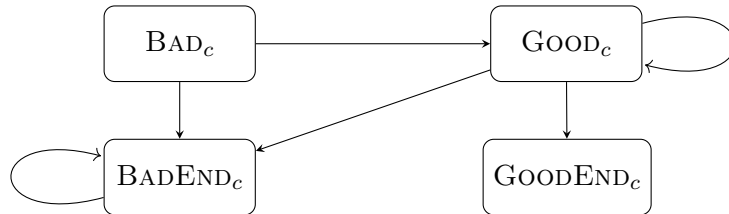

Figure 2: Possible transitions among stages of Definition 4.4

For $\sigma_*^{(T^*)}, \tau_*^{(T^*)}$ to be in stage GOODEND_c, $\sigma_*^{(t)}, \tau_*^{(t)}$ cannot be in stage BAD_c for any $t \geq T$. In other words, because $\sigma_*^{(T-1)}, \tau_*^{(T-1)}$ is in state BAD_c, the pair of colorings must escape from BAD_c in the very first step of (17) and never return. The following says that this probability of escape is comparable to the total probability mass of $\mathcal{W}$. We defer its proof to Appendix C.4.

17

Lemma 4.6. *Let σ, τ be any neighboring coloring pair in state BAD_c, and let σ', τ' be derived from one step of greedy coupling. Then $P[\sigma', \tau'$ in state $\mathrm{GOOD}_c] \geq \frac{4(k-\Delta-1)}{nk}$.*

Once a pair of colorings has escaped from state BAD_c into state GOOD_c, at every step it can only stay at GOOD_c, end at stage $\mathrm{GOODEND}_c$, or get absorbed into stage BADEND_c. The next two lemmas say that the last two events have probability $\Omega(1/n)$ and $O(1/n)$ respectively.

Lemma 4.7. *Let σ, τ be any neighboring coloring pair in stage GOOD_c, and let σ', τ' be derived from one step of greedy coupling. Then $P[\sigma', \tau'$ in stage $\mathrm{GOODEND}_c] \geq \frac{k-\Delta-2}{nk}$.*

Lemma 4.8. *Let σ, τ be any neighboring coloring pair in stage GOOD_c, and let σ', τ' be derived from one step of greedy coupling. Then $P[\sigma', \tau'$ in stage $\mathrm{BADEND}_c] \leq \frac{5}{n}$.*

Lemma 4.7 follows immediately from Lemma 4.2. Lemma 4.8 is the most technically involved step, and we defer its proof to Appendix C.4.

We can now complete the proofs of Lemma 4.4 and Theorem 1.1.

Proof of Lemma 4.4. Starting from $\sigma^{(T-1)}, \tau^{(T-1)}$ in stage BAD_c, by Lemma 4.6, the probability of transitioning to stage GOOD_c in the very next step is at least $\frac{4(k-\Delta-1)}{nk}$. As shown in Figure 2, once we leave stage BAD_c we never return. From stage GOOD_c, it is at most $\frac{5}{n} \cdot \frac{nk}{k-\Delta-2} = \frac{5k}{k-\Delta-2}$ times as likely to eventually end up at stage BADEND_c as it is to end up at stage $\mathrm{GOODEND}_c$, by Lemmas 4.7 and Lemma 4.8. So the probability of ending in stage $\mathrm{GOODEND}_c$ is at least $\frac{k-\Delta-2}{5k+(k-\Delta-2)} \cdot \frac{4(k-\Delta-1)}{nk}$, and we conclude that (19) and consequently Lemma 4.4 hold for γ as defined. $\qquad\square$

Proof of Theorem 1.1. Note that for $k > 1.833\Delta$ and $p_2 < 0.3$, $\gamma = \frac{(6k-\Delta-2)(k+2p_2\Delta)}{4(k-\Delta-2)(k-\Delta-1)} < 7.683410$, while $C = \frac{k+2p_2\Delta}{k-\Delta-2} < 2.920764$, so $C\gamma < 25.597784$ as defined in Lemma 4.3. Thus, substituting 25.597784 into the γ parameter for Linear Program 4 and solving numerically[8], we find that for

$$\hat{p}_1 = 1, \hat{p}_2 \approx 0.296706, \hat{p}_3 \approx 0.166762, \hat{p}_4 \approx 0.101790, \hat{p}_5 \approx 0.058475, \hat{p}_6 \approx 0.025989, \hat{p}_\alpha = 0 \ \forall \alpha \geq 7, \tag{20}$$

Linear Program 4 attains value $\lambda^* < 1.833239$. So provided $k \geq 1.833239\Delta$, Lemma 4.3 implies that the variable-length coupling is $(1-\alpha)$-contractive for absolute constant $\alpha := \frac{k-\lambda^*\Delta}{k-\Delta-2}$.

For $k \geq 1.833239\Delta$, Corollary 4.1 implies that β in the definition of Theorem 2.2 is at most $\frac{nk}{k-\Delta-2} \leq 2.21n$, so applying Theorem 2.2 with $\beta = 2.21n$, and $W = 2N_{max} + 1 = 13$ gives that $\tau_{mix}(\epsilon) = O(n\log(n/\epsilon))$. In particular, $\tau_{mix} = O(n\log n)$. $\qquad\square$

5 Proof of Theorem 1.1 Using an Alternative Metric

This section contains the proof of Theorem 1.1 using an alternative metric as presented in [DPP18]. Lemma 3.3 shows that a one-step coupling analysis using the Hamming metric cannot yield any improvement over 11/6. The first step in our proof is to find a set of flip parameters that only has two extremal configurations, up to symmetry, namely the ones used to define Linear Program 3. We then introduce a new metric that depends on the number of colors in non-extremal configurations. Analyzing the one-step coupling defined in Appendix A (greedy coupling) with this metric, we obtain a constant improvement over the 11/6 bound.

---

[8]Code for solving Linear Program 4 can be found at https://github.com/sitanc/mixedlp.

5.1   Choice of Flip Parameters and Expected Change in Hamming metric

By Corollary 3.1, the objective value of Linear Program 3 is $11/6$. It is straightforward to check that $\lambda = 11/6$ only if the solution satisfies $p_3 = 1/6$ and $p_\alpha = 0$ for all $\alpha \geq 7$. Since for any such assignment, the constraints corresponding to $(3, 2; (1), (1))$ and $(7, 3; (3, 3), (1, 1))$ are tight, we introduce the following variant of Linear Program 1:

Linear Program 5. *For variables $\{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ and λ, minimize λ subject to the following constraints: $p_0 = 0 \leq p_\alpha \leq p_{\alpha-1} \leq p_1 = 1$ for all $\alpha \geq 2$, $p_3 = 1/6$, $p_\alpha = 0$ for $\alpha \geq 7$ and for all realizable configurations $(A, B; \mathbf{a}, \mathbf{b})$ of size m different from $(3, 2; (2), (1))$, $(2, 3; (1), (2))$, $(7, 3; (3, 3), (1, 1))$ and $(3, 7; (1, 1), (3, 3))$, define a constraint*

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + \lambda m.$$

Consider the following reduced linear program with a finite set of variables and constraints.

Linear Program 6. *For variables $\{p_1, \ldots, p_6\}$ and λ, minimize λ subject to the following constraints: $0 \leq p_\alpha \leq p_{\alpha-1} \leq p_1 = 1$ for all $\alpha \geq 2$, $p_3 = 1/6$, constraints*

$$i(p_i - p_{i+1}) + j(p_j - p_{j+1}) - \min\{p_i - p_{i+1}, p_j - p_{j+1}\} \leq -1 + \lambda, \tag{21}$$

for $i \in \{1, \ldots, 6\}$, $j \in \{2, \ldots, 6\}$ with $(i, j) \neq (1, 2)$, and

$$2p_1 + 3p_2 - \min\{p_2 - p_5, p_3 - p_1\} \leq -1 + 2\lambda.$$

One can solve the program using a computer.

Observation 5.1. *The Linear Program 6 has objective value $\hat{\lambda} = \frac{161}{88} = 1.8295\ldots$ and an optimal solution is given by*

$$\hat{p}_1 = 1, \ \hat{p}_2 = \frac{185}{616}, \ \hat{p}_3 = \frac{1}{6}, \ \hat{p}_4 = \frac{47}{462}, \ \hat{p}_5 = \frac{9}{154}, \ \hat{p}_6 = \frac{2}{77}.$$

Lemma 5.1. *The assignment $\hat{\mathbf{p}} = \{\hat{p}_\alpha\}_{\alpha \in \mathbb{N}_0}$ where $\hat{p}_\alpha$ is given by Observation 5.1 for $\alpha \in [6]$ and $\hat{p}_\alpha = 0$ otherwise, forms a feasible solution of Linear Program 5 with objective value $\hat{\lambda}$.*

We defer the proof of the lemma to Appendix D.1.

Observation 5.2. *Consider the solution $\hat{\mathbf{p}}$ of Linear Program 5 given in Observation 5.1. The constraints in Linear Program 1 that are not contained in Linear Program 5 are implied by the conditions $p_1 = 1$, $p_3 = 1/6$ and $p_7 = 0$. Thus, $\hat{\mathbf{p}}$ is a feasible solution of Linear Program 1 with objective value $11/6$. Since the objective value of $\hat{\mathbf{p}}$ in Linear Program 5 is strictly smaller than $11/6$, up to symmetry, there are only two $\hat{\mathbf{p}}$-extremal configurations: $(3, 2; (1), (1))$ and $(7, 3; (3, 3), (1, 1))$.*

We conclude that

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq \begin{cases} \frac{11}{6} & \text{for every } \hat{\mathbf{p}}\text{-extremal configuration } (A, B; \mathbf{a}, \mathbf{b}), \\ \hat{\lambda} = \frac{161}{88} & \text{otherwise.} \end{cases} \tag{22}$$

5.2 Definition of the Alternative Metric

In this section we introduce the alternative metric we will use for the analysis of the one-step coupling, defined using a pre-metric (Γ, ω). Let Γ be the graph with vertex set Ω where two colorings are adjacent if and only if they differ at exactly one vertex. We proceed to define ω.

Fix the assignation of flip parameters $\hat{\mathbf{p}}$ given in Observation 5.1. Our definition of the pre-metric is driven by the fact that extremal configurations are an obstacle to show contraction of the metric when $k < 11\Delta/6$. Define the following sets of colors,

$$C_{\sigma,\tau}^1(v) := \{c : (A_c, B_c, \mathbf{a}^c, \mathbf{b}^c) \text{ is an extremal configuration for } (\sigma, \tau) \text{ of size } 1\} ,$$
$$C_{\sigma,\tau}^2(v) := \{c : (A_c, B_c, \mathbf{a}^c, \mathbf{b}^c) \text{ is an extremal configuration for } (\sigma, \tau) \text{ of size } 2\} .$$

Let $C_{\sigma,\tau}(v) = C_{\sigma,\tau}^1(v) \cup C_{\sigma,\tau}^2(v)$ be the set of colors c such that (σ, τ) has an extremal configuration for c. Note that for each color $c \in C_{\sigma,\tau}^2(v)$, there are two neighbors of v with color c. Let $\gamma_{\sigma,\tau}(v) = (|C_{\sigma,\tau}^1(v)| + 2|C_{\sigma,\tau}^2(v)|)/\Delta$, that is, the number of neighbors of v that participate in extremal configurations of (σ, τ) normalised by a factor Δ; thus $\gamma_{\sigma,\tau}(v) \leq 1$.

Let $\eta \in \left(0, \frac{1}{2}\right)$ be a sufficiently small constant to be fixed later. The *weight function* that we will use for our pre-metric is defined as

$$\omega(\sigma, \tau) := 1 - \eta(1 - \gamma_{\sigma,\tau}(v)) . \tag{23}$$

Note that $\omega(\sigma, \tau) \in [1 - \eta, 1]$. Since $\eta < \frac{1}{2}$ and $\gamma_{\sigma,\tau}(v) \leq 1$, every path containing at least two edges has weight greater than one. So every edge is a minimum weight path, implying that (Γ, ω) is a pre-metric. Let d be the metric on Ω obtained from (Γ, ω) using minimum weight paths in Γ. By Remark 2.1, for every $(\tilde{\sigma}, \tilde{\tau}) \in \Omega^2$ there exists a path between $\tilde{\sigma}$ and $\tilde{\tau}$ in Γ with $d_H(\tilde{\sigma}, \tilde{\tau})$ edges in which every edge has weight at most 1. It follows that $d(\tilde{\sigma}, \tilde{\tau}) \leq d_H(\tilde{\sigma}, \tilde{\tau})$.

Define

$$d_B(\tilde{\sigma}, \tilde{\tau}) := d_H(\tilde{\sigma}, \tilde{\tau}) - d(\tilde{\sigma}, \tilde{\tau}) . \tag{24}$$

In general, d_B is not a metric, here we will only use that it is non-negative. The contribution of d_B will be crucial for the constant improvement over $\frac{11}{6}$ in this approach.

Given the greedy coupling $(\sigma, \tau) \to (\sigma', \tau')$ for neighboring coloring pairs (σ, τ), define

$$\nabla(\sigma, \tau) := nk \, \mathbb{E}\left[d(\sigma', \tau') - d(\sigma, \tau)\right] . \tag{25}$$

Define the rescaled contributions to the expected change of d_H and d_B as

$$\nabla_H(\sigma, \tau) := nk \, \mathbb{E}\left[d_H(\sigma', \tau') - 1\right] ,$$
$$\nabla_B(\sigma, \tau) := -nk \, \mathbb{E}\left[d_B(\sigma', \tau') - d_B(\sigma, \tau)\right] ,$$

and note that $\nabla(\sigma, \tau) = \nabla_H(\sigma, \tau) + \nabla_B(\sigma, \tau)$.

We first bound ∇_H. Recall that X_c is the event that the coupling flips Kempe components in $\mathcal{D}_c$ in both chains. Denote by $\overline{X}$ the complement of the event $\cup_{c:\delta_c>0} X_c$. Define

$$\nabla_H(\sigma, \tau, c) := nk \, \mathbb{E}\left[\mathbb{1}_{X_c} \cdot (d_H(\sigma', \tau') - 1)\right] . \tag{26}$$

Note that $\mathbb{E}\left[\mathbb{1}_{\overline{X}} \cdot (d_H(\sigma', \tau') - 1)\right] = 0$ as we use the identity coupling if $\overline{X}$ holds, so σ' and τ' only differ at v. Using (6), Lemma 2.2, and (22), we obtain a bound on the expected change of the Hamming part of the metric in terms of the number on non-extremal configurations.

Corollary 5.1. *Let $\delta = \frac{11}{6} - \frac{161}{88}$. For every neighboring coloring pair (σ, τ), we have*

$$\nabla_H(\sigma, \tau) \leq \left(\frac{11}{6} - \delta\left(1 - \gamma_{\sigma,\tau}(v)\right)\right)\Delta - k .$$

5.3 Contribution of the Extremal Part of the Metric

In this section we bound $\nabla_B(\sigma, \tau)$ from above. Define the contributions

$$\nabla_B(\sigma, \tau, c) := -nk\,\mathbb{E}[\mathbb{1}_{X_c} \cdot (d_B(\sigma', \tau') - d_B(\sigma, \tau))]\,,$$
$$\overline{\nabla_B}(\sigma, \tau) := -nk\,\mathbb{E}[\mathbb{1}_{\overline{X}} \cdot (d_B(\sigma', \tau') - d_B(\sigma, \tau))]\,.$$

By equations (23) and (24), since $d_H(\sigma, \tau) = 1$, we have $d_B(\sigma, \tau) = 1 - \omega(\sigma, \tau) = \eta(1 - \gamma_{\sigma,\tau}(v))$. Moreover, $d_B(\sigma', \tau') \geq 0$. By the properties of the greedy coupling,

$$\nabla_B(\sigma, \tau, c) \leq \eta(1 - \gamma_{\sigma,\tau}(v))nk\mathbb{P}[X_c = 1] \leq 2\eta(1 - \gamma_{\sigma,\tau}(v))(\delta_c + 1)\,,$$

where we have used that the probability of flipping a given Kempe component is at most $1/nk$.

We can bound the expected change of ∇_B as follows

$$\nabla_B(\sigma, \tau) = \overline{\nabla_B}(\sigma, \tau) + \sum_{c \in [k]} \nabla_B(\sigma, \tau, c) \leq \overline{\nabla_B}(\sigma, \tau) + 2\eta(k + \Delta)(1 - \gamma_{\sigma,\tau}(v))\,. \tag{27}$$

Let $\overline{\mathcal{D}} = (\mathcal{S}_\sigma \cup \mathcal{S}_\tau) \setminus \mathcal{D}$ denote the set of Kempe components of σ and τ that do not involve vertex v. Note that each component in $\overline{\mathcal{D}}$ is a Kempe component of both σ and τ. An important difference here as opposed to the analysis of the contribution of ∇_H, is that the components in $\overline{\mathcal{D}}$ have an effect on the expected change of ∇_B. For a coloring σ and $S \in \mathcal{S}_\sigma$, let σ_S denote the coloring obtained by flipping S in σ. For $S \in \overline{\mathcal{D}}$, since $d_H(\sigma_S, \tau_S) = 1$, we have $d_B(\sigma_S, \tau_S) = \eta(1 - \gamma_{\sigma_S,\tau_S}(v))$. It follows that, $\overline{\nabla_B}(\sigma, \tau) = \eta\sum_{S \in \overline{\mathcal{D}}} p_{|S|}(\gamma_{\sigma_S,\tau_S}(v) - \gamma_{\sigma,\tau}(v))$.

For each $c \in [k]$ and $i \in \{1, 2\}$ and $S \in \overline{\mathcal{D}}$, let

$$\xi_{\sigma,\tau}(v, c, S) := \begin{cases} -i & \text{if } c \in C^i_{\sigma,\tau}(v) \text{ and } c \notin C_{\sigma_S,\tau_S}(v), \\ i & \text{if } c \notin C_{\sigma,\tau}(v) \text{ and } c \in C^i_{\sigma_S,\tau_S}(v), \\ -1 & \text{if } c \in C^2_{\sigma,\tau}(v) \text{ and } c \in C^1_{\sigma_S,\tau_S}(v), \\ 1 & \text{if } c \in C^1_{\sigma,\tau}(v) \text{ and } c \in C^2_{\sigma_S,\tau_S}(v), \\ 0 & \text{otherwise.} \end{cases}$$

The variable $\xi_{\sigma,\tau}(v, c, S)$ can be understood as the contribution of color c to $\gamma_{\sigma_S,\tau_S}(v) - \gamma_{\sigma,\tau}(v)$. For every $\mathcal{S}' \subseteq \overline{\mathcal{D}}$, we define

$$\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}') := \frac{\eta}{\Delta} \sum_{S \in \mathcal{S}'} p_{|S|}\xi_{\sigma,\tau}(v, c, S)\,,$$

and note that $\overline{\nabla_B}(\sigma, \tau) = \sum_{c \in [k]} \overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}})$.

Next lemma bounds from above the contribution $\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}})$ for each c. This is the most technical part of our approach and we defer its proof to Appendix D.2.

Lemma 5.2. *For every neighboring coloring pair (σ, τ) and color c, we have:*

   i) For $i \in \{1, 2\}$, if $c \in C^i_{\sigma,\tau}(v)$, then $\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \leq -i\eta\left(\frac{k}{\Delta} - \frac{3}{2}\right)$;

   ii) If $c \notin C_{\sigma,\tau}(v)$, then $\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \leq 2\eta\left(9 + \frac{15k}{\Delta}\right)$.

The following bound on ∇_B follows directly from (27) and Lemma 5.2.

Corollary 5.2. *For every neighboring coloring pair (σ, τ), we have*

$$\nabla_B(\sigma, \tau) \leq -\eta\left(\frac{k}{\Delta} - \frac{3}{2}\right)\gamma_{\sigma,\tau}(v) + 2\eta\left(10 + \frac{16k}{\Delta}\right)(1 - \gamma_{\sigma,\tau}(v))\,.$$

5.4 Contraction of the Metric and Proof of Theorems 1.1 and 1.2

We now can show that the metric d contracts in expectation.

Theorem 5.1. *For the flip parameters $\hat{\mathbf{p}}$ given in Observation 5.1 there exists $\epsilon_0, \mu > 0$ such that for every $k \geq \left(\frac{11}{6} - \epsilon_0\right) \Delta$ and every neighboring coloring pair (σ, τ), the greedy coupling satisfies*

$$\nabla(\sigma, \tau) \leq -\mu k .$$

Proof. Recall that $\delta = \frac{11}{6} - \frac{161}{88} = \frac{1}{264}$, and set $\eta = \frac{\delta \Delta}{53k}$. Fix $\epsilon_0 = \frac{1}{84000}$ and note that $\left(\frac{11}{6} - \frac{\delta}{318}\right) \Delta \leq k - \mu k$ for some small constant $\mu > 0$. Note that $\frac{k}{\Delta} \geq \frac{9}{5}$. Using (25) and Corollaries 5.1 and 5.2, it follows that

$$\nabla(\sigma, \tau) \leq \left(\frac{11}{6} - \left(\delta - 2\eta\left(10 + \frac{16k}{\Delta}\right)\right)\right)(1 - \gamma_{\sigma,\tau}(v)) - \eta\left(\frac{k}{\Delta} - \frac{3}{2}\right)\gamma_{\sigma,\tau}(v)\right)\Delta - k$$

$$\leq \left(\frac{11}{6} - \frac{\delta}{318}\right)\Delta - k \leq -\mu k . \qquad \square$$

We now proceed with the proof of our main result.

Proof of Theorem 1.1. Consider the metric d on Ω defined in Section 5.2. By Theorem 5.1, for the flip probabilities $\hat{\mathbf{p}}$ there exist $\epsilon_0, \mu > 0$ such that if $k \geq \left(\frac{11}{6} - \epsilon_0\right)\Delta$, then the greedy coupling $(\sigma, \tau) \to (\sigma', \tau')$ defined on neighbouring colouring pairs (σ, τ) satisfies

$$\mathbb{E}\left[d(\sigma', \tau')\right] \leq d(\sigma, \tau) - \frac{\mu}{n} \leq \left(1 - \frac{\mu}{n}\right) d(\sigma, \tau) . \tag{28}$$

By Theorem 2.1 with $\alpha = \mu/n$, we can extend the coupling over all $(\sigma, \tau) \in \Omega^2$ so (28) is still satisfied. As $\eta < 1/2$, for $\sigma \neq \tau$ one has $d(\sigma, \tau) \in (1/2, n]$. We use the coupling bound in (1) together with Markov's inequality, to obtain for $\sigma^{(0)} \in \Omega$

$$d_{\mathrm{TV}}(P^t(\sigma^{(0)}, \cdot), \pi) \leq \max_{\tau^{(0)} \in \Omega} \mathbb{P}[\sigma^{(t)} \neq \tau^{(t)}] = \max_{\tau^{(0)} \in \Omega} \mathbb{P}[d(\sigma^{(t)}, \tau^{(t)}) \geq 1/2] \leq \max_{\tau^{(0)} \in \Omega} 2\,\mathbb{E}[d(\sigma^{(t)}, \tau^{(t)})]$$

$$\leq 2(1 - \mu/n)^t n .$$

It follows that $\tau_{\mathrm{mix}}(\epsilon) \leq Cn(\log n + \log \epsilon^{-1})$, for some absolute constant $C > 0$. $\qquad \square$

6 List Colorings

In this section we introduce the notation for list-colorings and give an overview of the proof of Theorem 1.4, deferring the details to Appendix E.

A *list assignment* of G is a function $L : V(G) \to 2^{\mathbb{N}}$. An *L-coloring* is a function $\sigma : V(G) \to \mathbb{N}$ such that $\sigma(u) \in L(u)$ for all $u \in V(G)$. Usually in the literature list-colorings are assumed to be proper, here we will not require this but distinguish between proper and not necessarily proper list-colorings. We denote by Ω^L the set of all L-colorings of G. If $|L(u)| = k$ for all $u \in V(G)$, then we say that L is a *k-list-assignment* and that an L-coloring is a *k-list-coloring*.

The *Glauber dynamics for L-colorings* is a discrete-time Markov chain $(\sigma^{(t)})$ with state space Ω^L where $\sigma^{(t)}$ is generated from $\sigma := \sigma^{(t-1)}$ as follows:

1. Choose $v^{(t)}$ uniformly at random from $V(G)$.

2. For all vertices $v \neq v^{(t)}$, let $\sigma^{(t)}(v) = \sigma(v)$.

22

3. Choose $c^{(t)}$ uniformly at random from $L(v^{(t)})$, if c does not appear among the colors in the neighborhood of $v^{(t)}$ then let $\sigma^{(t)}(v^{(t)}) = c$, otherwise let $\sigma^{(t)}(v^{(t)}) = \sigma(v^{(t)})$.

Although we define the chain over Ω^L, $\sigma^{(t)}$ will converge to the uniform distribution on proper L-colorings, as in the case of non-list-colorings.

Given $\sigma \in \Omega^L$, one can define Kempe components of σ as for colorings and we denote by $\mathcal{S}_\sigma^L$ the multiset of Kempe components $S = S_\sigma(u, c)$ with $u \in V(G)$ and $c \in L(u)$. Let σ_S be the coloring obtained from σ by swapping the colors in S and note that σ_S is not necessarily an L-coloring as the new color of a vertex might not be in its list. Given $S = S_\sigma(v, c) \in \mathcal{S}_\sigma^L$, we say that S is *flippable* if for every $u \in S$ we have $\{\sigma(v), c\} \subseteq L(u)$. If S is flippable, then $\sigma_S \in \Omega^L$.

Let $\mathbf{p} = \{p_\alpha\}_{\alpha \in \mathbb{N}_0}$ be a collection of *flip parameters*. The *flip dynamics for L-colorings* is a random process generating a sequence of colorings $\sigma^{(0)}, \sigma^{(1)}, \sigma^{(2)}, \cdots$ of G where $\sigma^{(0)}$ is an arbitrary coloring in Ω^L and $\sigma^{(t)}$ is generated from $\sigma := \sigma^{(t-1)}$ as follows:

1. Choose $v^{(t)}$ uniformly at random from $V(G)$.

2. Choose $c^{(t)}$ uniformly at random from $L(v^{(t)})$.

3. Let $S = S_\sigma(v^{(t)}, c^{(t)})$ and $\alpha = |S|$. If S is flippable, with probability p_α/α let $\sigma^{(t)} = \sigma_S$, otherwise let $\sigma^{(t)} = \sigma$.

We prove the analogue of Theorem 1.1 for list-colorings.

Theorem 6.1. *The flip dynamics for sampling k-list-colorings is rapidly mixing with mixing time $O(n \log n)$, for any $k \geq (\frac{11}{6} - \epsilon_0)\Delta$ where $\epsilon_0 > 0$ is the same constant from Theorem 1.1.*

The proof of this theorem follows the same strategy as the proof of Theorem 1.1. The main difference in the analysis of the flip dynamics for list-coloring is that some of the moves that are valid in the non-list case, are not legal here. An important observation is that (4) no longer holds for list-colorings. For instance, it might be the case that $A_c = 0$ since $S_\sigma(v, c)$ is not flippable, while some of the $a_i^c \neq 0$. This leads to a weaker definition of realizable configuration and produces a linear program whose set of contraints is a superset of the constraints in Linear Program 1. An analysis similar to the one given for $c \in \{\sigma(v), \tau(v)\}$ implies that the new constraints have positive slack for the optimal solutions of the linear program we use, so no new extremal configuration arises. Thus, the analysis of the expected change of the one-step greedy coupling with respect to the Hamming metric is the same as for non-list-colorings in Section 3.

The other key observation common to extending both approaches of this work to list colorings is that if $S = S_\sigma(v^{(t)}, c^{(t)})$ has size 1, then S is always flippable since $c^{(t)} \in L(v^{(t)})$. This is enough to show that the same flip parameters used to prove Theorem 1.1 in either approach also work in the list coloring setting.

In Section 4, only size 1 flips are used to lower bound the probabilities of breaking apart extremal configurations and terminating the coupling, so Lemma 4.6, Lemma 4.7, and the lower bound in Lemma 4.2 still hold. It is also obvious that upper bounds on these events (Lemma 4.8 and the upper bound in Lemma 4.2) still hold, so Section 4 carries over to the list setting.

In Section 5, the list version of Corollary 5.1 holds for the flip parameters $\hat{\mathbf{p}}$ given by Observation 5.1. As components of size one are the ones giving the negative contribution in Lemma 5.2, the list version of Corollary 5.2 also holds. These two corollaries imply Theorem 6.1 in a similar way as in Section 5.4.

We provide a full proof of Theorem 6.1 in Appendix E using the approach of Section 5. This implies Theorem 1.4 in the same way that Theorem 1.1 implies Theorem 1.2.

References

[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 (FOCS), 1997.*, pages 223–231. IEEE, 1997.

[BD98]   Russ Bubley and Martin E Dyer. Faster random generation of linear extensions. In *SODA*, pages 350–354, 1998.

[BDG98]   Russ Bubley, Martin Dyer, and Catherine Greenhill. Beating the 2Δ bound for approximately counting colourings: a computer-assisted proof of rapid mixing. In *9th Annual Symposium on Discrete Algorithms, ACM–SIAM, New York–Philadelphia*, pages 355–363, 1998.

[BDK06]   Magnus Bordewich, Martin Dyer, and Marek Karpinski. Stopping times, metrics and approximate counting. In *International Colloquium on Automata, Languages, and Programming*, pages 108–119. Springer, 2006.

[BKMP05]   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.

[BW02]   Graham R Brightwell and Peter Winkler. Random colorings of a Cayley tree. *Contemporary Combinatorics*, 10:247–276, 2002.

[CKKL99]   Artur Czumaj, Przemka Kanarek, Mirosław Kutyłowski, and Krzyztof Loryś. Delayed path coupling and generating random permutations via distributed stochastic processes. In *Proceedings of the tenth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 271–280, 1999.

[CM18]   Sitan Chen and Ankur Moitra. Linear programming bounds for randomly sampling colorings. *arXiv preprint arXiv:1804.03156*, 2018.

[DF03]   Martin Dyer and Alan Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures & Algorithms*, 23(2):167–179, 2003.

[DFFV06]   Martin Dyer, Abraham D Flaxman, Alan M Frieze, and Eric Vigoda. Randomly coloring sparse random graphs with fewer colors than the maximum degree. *Random Structures & Algorithms*, 29(4):450–465, 2006.

[DFHV13]   Martin Dyer, Alan Frieze, Thomas P Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. *Random Structures & Algorithms*, 43(2):181–200, 2013.

[DGG+01]   Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, Mark Jerrum, and Michael Mitzenmacher. An extension of path coupling and its application to the Glauber dynamics for graph colorings. *SIAM Journal on Computing*, 30(6):1962–1975, 2001.

[DGM02]   Martin Dyer, Catherine Greenhill, and Mike Molloy. Very rapid mixing of the Glauber dynamics for proper colorings on bounded-degree graphs. *Random Structures & Algorithms*, 20(1):98–114, 2002.

[DPP18]   Michelle Delcourt, Guillem Perarnau, and Luke Postle. Rapid mixing of Glauber dynamics for colorings below Vigoda's 11/6 threshold. *arXiv preprint arXiv:1804.04025*, 2018.

[DSC93]   Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, pages 696–730, 1993.

[EHŠ+16]   Charilaos Efthymiou, Thomas P Hayes, Daniel Štefankovic, Eric Vigoda, and Yitong Yin. Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 704–713. IEEE, 2016.

[EHŠV18]   Charilaos Efthymiou, Thomas P Hayes, Daniel Štefankovič, and Eric Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1759–1771. SIAM, 2018.

[FV06]   Alan Frieze and Juan Vera. On randomly colouring locally sparse graphs. *Discrete Mathematics and Theoretical Computer Science*, 8, 2006.

[FV07]   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.

[GK07]   David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *Proceedings of the eighteenth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1245–1254. Society for Industrial and Applied Mathematics, 2007.

[GKM15]   David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 46(4):599–613, 2015.

[GMP05]   Leslie Ann Goldberg, Russell Martin, and Mike Paterson. Strong spatial mixing with fewer colors for lattice graphs. *SIAM Journal on Computing*, 35(2):486–517, 2005.

[Hay03]   Thomas P Hayes. Randomly coloring graphs of girth at least five. In *Proceedings of the thirty-fifth annual ACM Symposium on Theory of Computing*, pages 269–278. ACM, 2003.

[Hay06]   Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2006.*, pages 39–46. IEEE, 2006.

[HS05]   Thomas P Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2005.*, pages 511–520. IEEE, 2005.

[HV03]   Thomas P Hayes and Eric Vigoda. A non-Markovian coupling for randomly sampling colorings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2003.*, pages 618–627. IEEE, 2003.

[HV05]   Tom Hayes and Eric Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. In *Proceedings of the sixteenth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 971–979. Society for Industrial and Applied Mathematics, 2005.

[HV07]   Thomas P Hayes and Eric Vigoda. Variable length path coupling. *Random Structures & Algorithms*, 31(3):251–272, 2007.

[HVV15]   Thomas P Hayes, Juan C Vera, and Eric Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. *Random Structures & Algorithms*, 47(4):731–759, 2015.

[Jer95]   Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.

[KR01]   VS Anil Kumar and H Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.

[LM06]   Lap Chi Lau and Michael Molloy. Randomly colouring graphs with girth five and large maximum degree. In *Latin American Symposium on Theoretical Informatics*, pages 665–676. Springer, 2006.

[LV99]   Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures & Algorithms*, 15(3-4):229–241, 1999.

[Mol04]   Michael Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM Journal on Computing*, 33(3):721–737, 2004.

[MS10]   Elchanan Mossel and Allan Sly. Gibbs rapidly samples colorings of G(n,d/n). *Probability Theory and Related Fields*, 148(1-2):37–69, 2010.

[Pot52]   Renfrey Burnard Potts. Some generalized order-disorder transformations. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pages 106–109. Cambridge University Press, 1952.

[Sin92]   Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

[SS97]   Jesús Salas and Alan D Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86(3-4):551–579, 1997.

[TVVY12]   Prasad Tetali, Juan C Vera, Eric Vigoda, and Linji Yang. Phase transition for the mixing time of the Glauber dynamics for coloring regular trees. *The Annals of Applied Probability*, 22(6):2210–2239, 2012.

[Var17]   Shai Vardi. Randomly coloring graphs of logarithmically bounded pathwidth. *arXiv:1708.02677*, 2017.

[Vig99]   Eric Vigoda. Improved bounds for sampling colorings. In *40th Annual Symposium on Foundations of Computer Science (FOCS), 1999.*, pages 51–59. IEEE, 1999.

[WSK89]   Jian-Sheng Wang, Robert H Swendsen, and Roman Koteckỳ. Antiferromagnetic Potts models. *Physical Review Letters*, 63(2):109, 1989.

A Review of Vigoda's Greedy Coupling

A.1 The Coupling

In this section we give a self-contained overview of Vigoda's coupling analysis. For the reader's convenience, we restate some notation that was already introduced in Section 2.4.

Fix a neighboring coloring pair (G, σ, τ). For $c \in [k]$, let U_c denote the set of neighbors of v (in either coloring) that are colored c, and let $\delta_c = |U_c|$. We will sometimes denote the vertices in U_c by $\{u_1^c, ..., u_{\delta_c}^c\}$; where c is clear from context, we will simply denote these by $\{u_1, ..., u_{\delta_c}\}$.

Note that the symmetric difference $\mathcal{D} = \mathcal{S}_\sigma \Delta \mathcal{S}_\tau$ is precisely the Kempe components $S_\sigma(u_i^c, \tau(v))$ and $S_\sigma(v, c)$ in σ and the Kempe components $S_\tau(u_i^c, \sigma(v))$ and $S_\tau(v, c)$ in τ, for all colors c appearing in the neighborhood of v and all $i \in [\delta_c]$. All other Kempe components are shared between σ and τ, so for those, it is enough to use the identity coupling. Note that for colors $c \neq \sigma(v), \tau(v)$ not appearing in $N(v)$, the identity coupling then matches the flip of $S_\sigma(v, c)$ to that of $S_\tau(v, c)$ so that the two colorings of G become identical.

So the main concern is how to couple the flips of components in $\mathcal{D}$. We can decompose $\mathcal{D}$ into $\cup_{c:\delta_c>0} \mathcal{D}_c$, where the sets $\mathcal{D}_c$ are defined as follows:

Definition A.1. Let $\mathcal{D}_c$ be the set of Kempe components consisting of $S_\sigma(v, c), S_\tau(v, c)$, and all $S_\sigma(u, \tau(v))$ and $S_\tau(u, \sigma(v))$ for all $u \in U_c$.

Informally, $\mathcal{D}_c$ is the subset of $\mathcal{D}$ that involves the color c. It is easy to see that for $c \neq \sigma(v)$,

$$S_\sigma(v, c) = \left(\bigcup_{i=1}^{\delta_c} S_\tau(u_i^c, \sigma(v)) \right) \cup \{v\}, \tag{29}$$

and when $c = \sigma(v)$, $S_\sigma(v, c), S_\tau(u, \sigma(v)) = \emptyset$ for any $u \in U_c$. Likewise we have that for $c \neq \tau(v)$,

$$S_\tau(v, c) = \left(\bigcup_{i=1}^{\delta_c} S_\sigma(u_i^c, \tau(v)) \right) \cup \{v\}, \tag{30}$$

and when $c = \tau(v)$, $S_\tau(v, c), S_\sigma(u, \tau(v)) = \emptyset$ for $u \in U_c$.

The sets $\mathcal{D}_c$ are disjoint except possibly the pair $\mathcal{D}_{\sigma(v)}, D_{\tau(v)}$, as these both contain $(\sigma(v), \tau(v))$-colored Kempe components, though we defer this point to later.

Remark A.1. *One subtlety is that there may exist multiple neighbors $u_1', \ldots, u_m' \in N(v)$ which are colored c but which satisfy $S_\tau(u_1', \sigma(v)) = \cdots = S_\tau(u_m', \sigma(v))$; to guarantee that the flip of each component is considered exactly once, redefine $S_\tau(u_i', \sigma(v)) = \emptyset$ for all $1 < i \leq m$. Handle the components $S_\sigma(u_i', \tau(v))$ analogously.*

In [Vig99], Vigoda couples flips of Kempe components within $\mathcal{D}_c$ as follows. First we require some notation. For c such that $\delta_c > 0$, define $A_c := |S_\sigma(v, c)|$, $B_c := |S_\tau(v, c)|$, $a_i^c = |S_\tau(u_i^c, \sigma(v))|$, and $b_i^c = |S_\sigma(u_i^c, \tau(v))|$. Define the vectors $\mathbf{a}^c := (a_i^c : i \in [\delta_c])$ and $\mathbf{b}^c := (b_i^c : i \in [\delta_c])$. Also define $a_{max}^c = \max_i a_i^c$ and denote the maximizing i by i_{max}^c. Likewise define $b_{max}^c = \max_j b_j^c$ and denote the maximizing j by j_{max}^c. When it is clear from context that we are just focusing on a generic color c, we will refer to these as $A, B, a_i, b_i, \mathbf{a}, \mathbf{b}, a_{max}, i_{max}, b_{max}, j_{max}$. We say that neighboring coloring pair (G, σ, τ) has a *configuration* $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ of size δ_c.

Naively we have the bounds

$$1 + a_{max} \leq A \leq 1 + \sum_i a_i, \qquad 1 + b_{max} \leq B \leq 1 + \sum_i b_i, \tag{31}$$

and moreover the upper bounds in (4) are equalities when $c \neq \sigma(v), \tau(v)$.

Note that $S_\sigma(v, c)$ and $S_\tau(v, c)$ can be quite different but $S_\sigma(v, c) \supset S_\tau(u_i, \sigma(v))$ so it is easier to understand the overlap between these two components. Among all choices of i, this overlap is maximized for $i = i_{max}$, and the idea of Vigoda's coupling is thus to greedily couple the flips of the biggest components, i.e. $S_\sigma(v, c), S_\tau(v, c)$, to the flips of the next biggest components, i.e. $S_\tau(u_{i_{max}}, \sigma(v)), S_\sigma(u_{j_{max}}, \tau(v))$, and then to couple together as closely as possible the flips of $S_\sigma(u_i, \tau(v))$ and $S_\tau(u_i, \sigma(v))$ for each $i \in [\delta_c]$. Formally, assuming $p_1 \geq p_2 \geq \cdots$ we have:

1. Flip $S_\sigma(v, c)$ and $S_\tau(u_{i_{max}}, \sigma(v))$ together with probability p_A.

2. Flip $S_\tau(v, c)$ and $S_\sigma(u_{j_{max}}, \tau(v))$ together with probability p_B.

3. For $i \in [\delta_c]$, define

$$q_i = \begin{cases} p_{a_i} - p_A & \text{if } i = i_{max} \\ p_{a_i} & \text{otherwise} \end{cases} \tag{32}$$

$$q_i' = \begin{cases} p_{b_i} - p_B & \text{if } i = j_{max} \\ p_{b_i} & \text{otherwise} \end{cases} \tag{33}$$

Note that q_i and q_i' are the remaining probability associated to flips $S_\tau(u_i, \sigma(v))$ and $S_\sigma(u_i, \tau(v))$ respectively.

   (a) Flip $S_\tau(u_i, \sigma(v))$ and $S_\sigma(u_i, \tau(v))$ together with probability $\min(q_i, q_i')$
   (b) Flip only $S_\tau(u_i, \sigma(v))$ together with probability $q_i - \min(q_i, q_i')$
   (c) Flip only $S_\sigma(u_i, \tau(v))$ together with probability $q_i' - \min(q_i, q_i')$

Coupled moves 1) and 2) change the Hamming distance by at most $A - a_{max} - 1$ and $B - b_{max} - 1$ respectively (with equality, for example, if G is a tree rooted at v). For any given $i \in [\delta_c]$, coupled move 3a) changes the Hamming distance by $a_i + b_i - 1$, where the extra -1 term comes from the fact that $S_\tau(u_i, \sigma(v))$ and $S_\sigma(u_i, \tau(v))$ are of size a_i and b_i respectively but share vertex u_i. On the other hand, coupled moves 3b) and 3c) obviously change the Hamming distance by a_i and b_i respectively. For a configuration $(A, B; \mathbf{a}, \mathbf{b})$, define

$$H(A, B; \mathbf{a}, \mathbf{b}) = (A - a_{max} - 1)p_A + (B - b_{max} - 1)p_B + \sum_i f(u_i), \tag{34}$$

where

$$f(u_i) = a_i q_i + b_i q_i' - \min(q_i, q_i') \tag{35}$$

The above discussion implies that for $c \neq \sigma(v), \tau(v)$ appearing in the neighborhood of v,

$$kn \cdot \mathbb{E}[\mathbb{1}_{X_c} \cdot (d(\sigma', \tau') - 1)] \leq H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c), \tag{36}$$

where X_c is the random event that the coupling flips components in $\mathcal{D}_c$ in both colorings. For c not appearing in the neighborhood of v, the Hamming distance will not change if Kempe components containing the color c are flipped in both colorings, as the coupling is the identity on these components, except if v is flipped to c in both colorings, in which case the Hamming distance decreases by 1.

Lastly, we review how the case of $c = \sigma(v), \tau(v)$ and $\mathcal{D}_{\sigma(v)} \cup D_{\tau(v)} \neq \emptyset$ is handled in [Vig99]. This is the main place where one needs to be careful about the fact that neighboring coloring pairs σ, τ need not be proper.

Remark A.2. *When $c = \sigma(v), \tau(v)$, we must make sure not to double count flips, as it is possible that $\mathcal{D}_{\sigma(v)}$ and $\mathcal{D}_{\tau(v)}$ share Kempe components. In this remark, suppose $\mathcal{D}_{\sigma(v)} \cap D_{\tau(v)} \neq \emptyset$. This can only happen if there exist $x_i, y_j \in N(v)$ colored $\sigma(v), \tau(v)$ respectively for which $S_\sigma(v, \tau(v)) = S_\sigma(x_i, \tau(v))$ and $S_\tau(v, \sigma(v)) = S_\tau(x_i, \sigma(v))$. To avoid double counting, Vigoda sets $S_\sigma(v, \tau(v)) = S_\tau(y_j, \sigma(v)) = \emptyset$ in this case. The bound (6) then holds for both $c = \sigma(v), \tau(v)$. The only difference is that some values among A_c, B_c and the entries of $\mathbf{a}^c, \mathbf{b}^c$ will be zero, in which case we take $p_0 = 0$.*

Specifically, for $c = \tau(v)$, we have $A_c = 0$, $B_c = b_{max}^c = 0$, and at least one a_j^c is zero, namely the one corresponding to the component $S_\tau(y_j, \sigma(v)) = S_\tau(v, \sigma(v))$. In this case one can check that $H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) = \sum a_j^c p_{a_j^c}$, and provided $\alpha p_\alpha \leq 1$ for all α, this is at most $\delta_c - 1$.

For $c = \sigma(v)$, we have $A_c = 0$, $a_i^c = 0$ for all i, and $B_c = \sum_j b_j^c$. Let j^* be the index of the unique neighbor u_{j^*} of v for which $S_\tau(v, \sigma(v)) = S_\sigma(u_{j^*}, \sigma(v))$. Then because $S_\sigma(u_{j^*}, \sigma(v))$ contains v, we need to modify the definition of b_{max}^c. Let $b_{max}^c = \max_j(b_j^c - \mathbb{I}_{j=j^*})$ and denote the maximizing j by j_{max}^c. Then the lower bound on B_c in (4) still holds, and (6) still holds. Moreover, if $\delta_c = 1$, then $\mathbb{E}[d(\sigma', \tau') - 1 | X_{\sigma(v)}] = H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) = 0$.

Henceforth, we will refer to the coupling defined above as the *greedy coupling*. We can conclude the following, implicit in [Vig99]:

Lemma A.1. *Let $(\sigma, \tau) \mapsto (\sigma', \tau')$ be the greedy coupling. Then*

$$\mathbb{E}[d(\sigma', \tau') - 1] \leq \frac{1}{nk} \left(-|\{c : \delta_c = 0\}| + \sum_{c : \delta_c \neq 0} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) \right). \tag{37}$$

The function H implicitly depends on the choice of flip parameters $\{p_\alpha\}$, while $A_c, B_c, \mathbf{a}^c, \mathbf{b}^c$ depend on (G, σ, τ). The remaining analysis in [Vig99] once (7) has been deduced essentially boils down to picking a good set of flip parameters.

A.2 Enumerating Realizable Configurations

In Section 3 we raised the issue of enumerating all realizable configurations in Linear Program 1. In particular, while it was easy to enumerate realizable configurations $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ for which $c \neq \sigma(v), \tau(v)$, we provided without proof two types of constraints ((10) and (11)) that we claimed would handle realizable configurations for which $c = \sigma(v), \tau(v)$. In this subsection we fill in the details for why these two constraints suffice for configurations with $c = \sigma(v), \tau(v)$.

For $c = \sigma(v)$, by Remark A.2, any realizable $(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$ satisfies $H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) = 0$ if $\delta_c = 1$, and satisfies $A_c = 0$, $\mathbf{a}^c = (0, ..., 0)$, $B_c = \sum_i b_i^c$, and

$$H(A_c, B_c, \mathbf{a}^c, \mathbf{b}^c) = (B_c - b_{max}^c - 1)p_{B_c} + \sum_{i \neq j_{max}} b_i^c p_{b_i^c} + b_{j_{max}}(p_{b_{j_{max}}^c} - p_{B_c}) \leq (B_c - b_m^c)p_{B_c} + \sum b_i^c p_{b_i^c}$$

if $\delta_c > 1$. So the relaxed constraint (10) covers all constraints corresponding to $c = \sigma(v)$.

For $c = \tau(v)$, we know by Remark A.2 that when $c = \tau(v)$, (11) ensures that $H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) \leq \delta_c - 1$, and for any $\lambda > 1$ (corresponding to $k > d$, which is the regime we are interested in to begin with), we automatically have that $\delta_c - 1 < -1 + \lambda \cdot \delta_c$.

B Extremal Configurations for Vigoda's Choice of Flip Parameters and Missing Proofs from Section 3

Observation B.1. *Consider the assignment (13) in Linear Program 2 for $N_{max} = 6$ and $m^* = 3$. Constraint (11) is tight under the assignment (13) only for $\alpha = 1$. Among the constraints of the*

form (9) *associated to a realizable configuration* $(A, B; \mathbf{a}, \mathbf{b})$, *up to symmetry, there are six tight constraints:*

    i) $m = 1$, $A - 1 = a_1 \in \{2, 3, 4, 5\}$ *and* $B - 1 = b_1 = 1$;

    ii) $m = 2$, $A = a_1 + a_2 + 1$, $a_1 = a_2 \in \{2, 3\}$, $B = 1$ *and* $b_1 = b_2 = 1$.

Any other constraints of the form (9) *that do not meet these conditions, and all constraints of the form* (10) *and* (12), *are not tight under the assignment* (13). *This can be verified numerically. It follows that Vigoda's solution has six extremal realizable configurations, up to symmetries.*

Proof. The tightness of (11) only for $\alpha = 1$ is obvious. That the other constraints mentioned in the observation have zero slack can be checked by hand. We verify that all other constraints have nonzero slack.

Case 1. *Constraint* (9) *for* $m = 1$

We first consider realizable $(A, B; \mathbf{a}^c, \mathbf{b}^c)$. It is easy to see that $(i - 1)(p_i - p_{i+1}) \leq 1/7$ with equality if and only if $i \in \{2, 3, 4, 5\}$, and that $i(p_i - p_{i+1}) \leq 29/42$ with equality if and only if $i = 1$. Note that for $m = 1$,

$$H(A, B; \mathbf{a}, \mathbf{b}) = \max\left(a_1(p_{a_1} - p_{a_1+1}) + (b_1 - 1)(p_{b_1} - p_{b_1+1}), (a_1 - 1)(p_{a_1} - p_{a_1+1}) + b_1(p_{b_1} - p_{b_1+1})\right)$$
$$\leq \frac{29}{42} + \frac{1}{7} = \frac{5}{6},$$

with equality if and only if $a_1 = 1$ and $b_1 \in \{2, 3, 4, 5\}$ or $b_1 = 1$ and $a_1 \in \{2, 3, 4, 5\}$.

Case 2. *Constraint* (9) *for* $m = 2$

We analyze this case in the same way that Claim 6 of [Vig99] is proved. Assume without loss of generality that $p_{a_{max}} - p_A \leq p_{b_{max}} - p_B$ and $a_1 \geq a_2$. In [Vig99] it is noted that one may assume that $b_2 \geq b_1$ so that

$$H(A, B; \mathbf{a}, \mathbf{b}) = (A - 2a_1)p_A + (B - 2b_2 - 1) + (a_1 - 1)p_{a_1} + a_2 p_{a_2} + b_1 p_{b_1} + b_2 p_{b_2} - \min(p_{a_2}, p_{b_2} - p_B).$$

Now we proceed by casework on $\min(p_{a_2}, p_{b_2} - p_B)$:

- $p_{a_2} \leq p_{b_2} - p_B$: In this case we have

$$H(A, B; \mathbf{a}, \mathbf{b}) = (a_1 - 1)p_{a_1} + (a_2 - 1)p_{a_2} + (A - 2a_1)p_A + b_1 p_{b_1} + b_2 p_{b_2} + (B - 2b_2 - 1)p_B. \tag{38}$$

One can check that $(a - 1)p_a \leq 1/3$ with equality if and only if $a = 3$. If $a_1 = 3$, $(A - 2a_1)p_A \leq 0$ with equality if and only if $a_2 = 3$ and $6 \leq A \leq 7$. However $(6, 3; (3, 3); (b_1, b_2))$ is not realizable. If $a_1 \neq 3$, $(A - 2a_1)p_A > 0$ if and only if $a_1 = a_2$ and $A = a_1 + a_2 + 1$. It turns out that $p_3 < 2p_2 + p_5 = 4p_3 + p_7 = 2/3$ and thus (38) is only maximized when $a_1 = a_2 \in \{2, 3\}$ and $A = a_1 + a_2 + 1$. In a similar manner, we can verify that for any fixed A, a_1, a_2, (38) is only maximized when $b_1 = b_2 = 1$ and $B = 3$ and, in such a case, the contribution to the part involving B, b_1, b_2 is 2. So, with the given assumptions, the only two configurations with $H(A, B; \mathbf{a}, \mathbf{b}) = 8/3$ are $(7, 3; (3, 3); (1, 1))$ and $(5, 3; (2, 2); (1, 1))$, up to .

- $p_{a_2} > p_{b_2} - p_B$: In this case we have

$$H(A, B; \mathbf{a}, \mathbf{b}) = (a_1 - 1)p_{a_1} + a_2 p_{a_2} + (A - 2a_1)p_A + b_1 p_{b_1} + (b_2 - 1)p_{b_2} + (B - 2b_2)p_B.$$

This is symmetric with respect to flipping the roles of (a_1, a_2) and (b_2, b_1), and it can be verified that $(a_1 - 1)p_{a_1} + a_2 p_{a_2} + (A - 2a_1)p_A < 4/3$, so $H(A, B; \mathbf{a}, \mathbf{b}) < 8/3$.

Case 3. *Constraint* (10)

For $m = 2$, the left-hand side of (10) is $b_1 p_{b_1+b_2} + b_1 p_{b_1} + b_2 p_{b_2}$, which attains its maximum value of $p_2 + 2 < -1 + 2\lambda$ at $b_1 = b_2 = 1$. For $m > 2$, note that (11) implies that the left-hand side of (10) is at most $m + 1 < -1 + \lambda \cdot m$ provided $\lambda > 5/3$, which is certainly the case.

Case 4. *Constraint* (12)

One can check that $(A - 2)p_A \le 4/21$. And if $p_a \le p_b$, then $a \cdot p_a + b \cdot p_b - \min(p_a, p_b) = (a - 1)p_a + b \cdot p_b$. But $(a - 1)p_a \le 1/3$ and $b \cdot p_b \le 1$. So $x^* = 4/21$ and $y^* = 4/3$, and it is clear that $-1 + (11/6) \cdot 3 > 2 \cdot x^* + m^* \cdot y^*$ for $m^* = 3$, so (12) has nonzero slack. $\qquad\square$

Proof of Lemma 3.3. We first show there exists no choice of flip parameters for which greedy coupling contracts for all of $\mathcal{C}^*$. Let $\lambda = k/d$, and suppose to the contrary that $1 \le \lambda < 11/6$ and yet there exists a set of flip parameters $\{p_\alpha\}$ for which all pairs of colorings in $\mathcal{C}^*$ contracted in distance. The expected change in distance for G_1 is

$$\frac{d}{nk} \cdot H(3, 2; (2), (1)) = d\,(p_1 + p_2 - 2p_3 - \min(p_1 - p_2, p_2 - p_3)) < 0 \le \frac{d(-1 + \lambda)}{nk}.$$

The expected change in distance for G_2 is

$$\frac{d}{2nk} \cdot H(7, 3; (3, 3), (1, 1)) = (d/2)\,(2p_1 + 5p_3 - \min(p_1 - p_3, p_3 - p_7)) < 0 \le \frac{d(-1 + 2\lambda)}{2nk}.$$

As this implies $p_1 - p_3 \le -1 + \lambda$ and $2p_1 + 4p_3 + p_7 \le -1 + 2\lambda$, using $p_1 = 1$ and $p_7 \ge 0$ it yields $\lambda \ge 11/6$, a contradiction.

Finally, it is straightforward to check that no one-step coupling can do better than the greedy coupling. This is clear for G_1. Indeed, certainly for any component not in $\mathcal{D}$, the coupling should just be the identity. Now for any neighbor u of v with color c, suppose a nonzero amount of probability mass p for the flip of $S_\tau(u, \sigma(v))$ is matched in the optimal one-step coupling to the flip of a component other than $S_\sigma(v, c)$. The expected change in distance conditioned on this pair of components being chosen in the coupling is strictly greater than the expected change if that mass p were instead reallocated to the empty flip in σ, contradicting optimality. By symmetry we can show that the flip of $S_\sigma(u_i, \tau(v))$ is coupled only to the empty flip in τ and the flip of $S_\tau(u_i, c)$. Finally, if not all of the probability mass for the flip of $S_\sigma(v, c)$ is matched to the flip of $S_\tau(u, \sigma(v))$, then we can strictly improve the coupling by reallocating that mass to $S_\tau(u, \sigma(v))$.

A similar argument shows that the optimal one-step coupling for G_2 is the greedy coupling. $\quad\square$

C Missing Proofs from Section 4

C.1 Proof of Lemma 4.1

Proof. Denote a minimizing choice of $\{p_\alpha\}$ and $\lambda_{sing}, \lambda_{bad}, \lambda_{good}$ for the γ-mixed coupling LP by $\{p_\alpha^*\}$ and $\lambda_{sing}^*, \lambda_{bad}^*, \lambda_{good}^*$. Then for any $(G, \sigma, \tau) \in \mathcal{C}$ and the greedy coupling $(\sigma, \tau) \mapsto (\sigma', \tau')$, observe that

$$\mathbb{E}[d_H(\sigma', \tau') - 1] \le -|\{c : \delta_c = 1\}| + \sum_{\substack{c:\sigma,\tau \\ \text{SING}_c}} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) + \sum_{\substack{c:\sigma,\tau \\ \text{BAD}_c}} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c) + \sum_{\substack{c:\sigma,\tau \\ \text{GOOD}_c}} H(A_c, B_c; \mathbf{a}^c, \mathbf{b}^c)$$

$$\le -|\{c : \delta_c = 1\}| + \sum_{\substack{c:\sigma,\tau \\ \text{SING}_c}} (-1 + \lambda_{sing}) + \sum_{\substack{c:\sigma,\tau \\ \text{BAD}_c}} (-1 + 2\lambda_{bad}) + \sum_{\substack{c:\sigma,\tau \\ \text{GOOD}_c}} (-1 + \delta_c \lambda_{good})$$

$$= -k + \lambda_{sing} \cdot N_{sing}(\sigma, \tau) + 2\lambda_{bad} \cdot N_{bad}(\sigma, \tau) + \sum_{\substack{c:\sigma,\tau \\ \text{GOOD}_c}} \delta_c \cdot \lambda_{good}. \tag{39}$$

But because $\delta_c \geq 2$ for any $c \neq \sigma(v), \tau(v)$ for which σ, τ are in state GOOD_c, because σ, τ are always in state $\text{GOOD}_{\sigma(v)}$, $\text{GOOD}_{\tau(v)}$, and because

$$N_{sing}(\sigma, \tau) + 2N_{bad}(\sigma, \tau) + \sum_{c:\sigma,\tau \ \text{GOOD}_c} \delta_c = \Delta(v),$$

we conclude that (39) is a convex combination of the terms

$$-k + \lambda_{sing}^* \cdot \Delta(v), \quad -k + \lambda_{good}^* \cdot \Delta(v), \quad -k + \left(\frac{\gamma}{\gamma+1} \cdot \lambda_{bad}^* + \frac{1}{\gamma+1} \cdot \lambda_{good}^*\right) \Delta(v).$$

So we conclude that $\mathbb{E}[d_H(\sigma', \tau') - 1] \leq -k + \lambda_\gamma^* \Delta(v) < 0$ as long as $k > \lambda_\gamma^* \Delta$. $\qquad \square$

C.2 Proof of Upper Bound in Lemma 4.2

Proof. Fix a color c for which $\delta_c \neq 0$ and some $i \in [\delta_c]$. For $i \neq i_{max}, j_{max}$, the pairs of flips $(S_\sigma(u_i, \tau(v)), S_\tau(u_i, \sigma(v)))$, $(S_\sigma(u_i, \tau(v)), \emptyset)$, and $(\emptyset, S_\tau(u_i, \sigma(v)))$ have probability mass $\min(p_{b_i}, p_{a_i})$, $\max(0, p_{b_i} - p_{a_i})$, and $\max(p_{a_i} - p_{b_i}, 0)$, for a total of $\max(p_{a_i}, p_{b_i})$. The remaining pairs of flips have probability masses which depend on whether $i_{max} = j_{max}$, as shown in Table 1.

Table 1: Probability masses for some coupled flips

Flip in σ	Flip in τ	$i_{max} = j_{max}$	$i_{max} \neq j_{max}$
$S_\sigma(v, c)$	$S_\tau(u_{i_{max}}, \sigma(v))$	p_A	p_A
$S_\tau(v, c)$	$S_\sigma(u_{j_{max}}, \tau(v))$	p_B	p_B
$S_\sigma(u_{i_{max}}, \tau(v))$	$S_\tau(u_{i_{max}}, \sigma(v))$	$\min(p_{a_{i_{max}}} - p_A, p_{b_{i_{max}}} - p_B)$	$\min(p_{a_{i_{max}}} - p_A, p_{b_{i_{max}}})$
$(S_\sigma(u_{j_{max}}, \tau(v))$	$S_\tau(u_{j_{max}}, \sigma(v))$	N/A	$\min(p_{b_{j_{max}}} - p_B, p_{a_{j_{max}}})$
$(S_\sigma(u_{i_{max}}, \tau(v))$	$\emptyset$	$\max(0, p_{b_{i_{max}}} - p_B - p_{a_{i_{max}}} + p_A)$	$\max(0, p_{b_{i_{max}}} - p_{a_{i_{max}}} + p_A)$
$\emptyset$	$S_\tau(u_{i_{max}}, \sigma(v))$	$\max(0, p_{a_{i_{max}}} - p_A - p_{b_{i_{max}}} + p_B)$	$\max(0, p_{a_{i_{max}}} - p_A - p_{b_{i_{max}}})$
$S_\sigma(u_{j_{max}}, \tau(v))$	$\emptyset$	N/A	$\max(0, p_{a_{j_{max}}} - p_{b_{j_{max}}} + p_B)$
$\emptyset$	$S_\tau(u_{j_{max}}, \sigma(v))$	N/A	$\max(0, p_{b_{j_{max}}} - p_B - p_{a_{j_{max}}})$
Total		$\max(p_{a_{i_{max}}} + p_A, p_{b_{i_{max}}} + p_B)$	$\max(p_{a_{j_{max}}} + p_B, p_{b_{j_{max}}}) + \max(p_{b_{i_{max}}} + p_A, p_{a_{i_{max}}})$

From these we can conclude that

$$nk \cdot \mathbb{P}[(S, S') \text{ terminating}] \leq \left(\sum_{c:\delta_c>0} p_{A_c} + p_{B_c}\right) + \left(\sum_{c:\delta_c>0, i\in[\delta_c]} \max(p_{a_i^c}, p_{b_i^c})\right) + \left(\sum_{c:\delta_c=0} p_1\right).$$

The sum of the second and third summands is at most k. For the first summand, note that when A_c, B_c are nonzero and $\delta_c > 0$, $A_c, B_c \geq 2$, so the first summand is at most $2p_2\Delta$. The desired upper bound follows. $\qquad \square$

C.3 Proof of Lemma 4.3

Proof. Let $\lambda_{sing}^*, \lambda_{tree}^*, \lambda_{good}^*$ be the values for $\lambda_{sing}, \lambda_{tree}, \lambda_{good}$ of the minimizer of the $C\gamma$-mixed coupling LP from Definition 4. For Kempe components S, S' in σ, τ respectively, define $E_{\sigma,\tau}^{S,S'} = d_H(\sigma', \tau') - 1$, where σ', τ' is the pair of colorings obtained by flipping S in σ and S' in τ, and let $p_{\sigma^{(T-1)},\tau^{(T-1)}}^{S,S'}$ be the probability that S, S' are flipped in one step of greedy coupling starting from $\sigma^{(T-1)}, \tau^{(T-1)}$. Then we have that

$$\mathbb{E}[d_H(\sigma^{(T_{\text{stop}})}, \tau^{(T_{\text{stop}})}) - 1] = \sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot Z(\sigma^{(T-1)}, \tau^{(T-1)}), \qquad (40)$$

where

$$Z(\sigma^{(T-1)}, \tau^{(T-1)}) := \sum_{S,S'} \mathbb{I}[(S, S') \text{ terminating}] \cdot p_{\sigma^{(T-1)},\tau^{(T-1)}}^{S,S'} \cdot E_{\sigma^{(T-1)},\tau^{(T-1)}}^{S,S'}$$

But note that for $v^{(T)}, c^{(T)}$ not terminating, $E_{\sigma^{(T-1)},\tau^{(T-1)}}^{S,S'} = 0$ because the one-step coupling is just the identity coupling, so $Z(\sigma^{(T-1)}, \tau^{(T-1)})$ is just the expected change in distance under one step of greedy coupling on the neighboring coloring pair $(G, \sigma^{(T-1)}, \tau^{(T-1)})$. Therefore, for any $T \leq T_{\text{stop}}$,

$$Z(\sigma^{(T-1)}, \tau^{(T-1)}) = \mathbb{E}[d_H(\sigma^{(T)}, \tau^{(T)}) - 1]$$

$$\leq \frac{1}{nk}\left((-1 + \lambda_{sing}^*) \cdot N_{sing}(\sigma^{(T-1)}, \tau^{(T-1)}) + (-1 + 2\lambda_{bad}^*) \cdot N_{bad}(\sigma^{(T-1)}, \tau^{(T-1)}) \right.$$

$$\left. + \sum_{\substack{c:\sigma^{(T-1)},\tau^{(T-1)} \\ \text{Good}_c}} (-1 + \delta_c \cdot \lambda_{good}^*) - |\{c : \delta_c = 0\}|\right)$$

$$= \frac{1}{nk} \cdot \left(-k + \lambda_{sing}^* N_{sing}(\sigma^{(T-1)}, \tau^{(T-1)}) + 2\lambda_{bad}^* N_{bad}(\sigma^{(T-1)}, \tau^{(T-1)}) \right.$$

$$\left. + \sum_{\substack{c:\sigma^{(T-1)},\tau^{(T-1)} \\ \text{Good}_c}} \delta_c \cdot \lambda_{good}^*\right). \qquad (41)$$

Because

$$N_{sing}(\sigma, \tau) + 2N_{bad}(\sigma, \tau) + \sum_{\substack{c:\sigma,\tau \\ \text{Good}_c}} \delta_c = \Delta(v)$$

for all neighboring coloring pairs (G, σ, τ), we conclude from (40) and (41) that

$$\mathbb{E}[d_H(\sigma^{(T_{\text{stop}})}, \tau^{(T_{\text{stop}})}) - 1] = \frac{1}{nk}\left(\sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau]\right)\left(-k + \lambda_{sing}^* \Delta_{sing} + \lambda_{bad}^* \Delta_{bad} + \lambda_{good}^* \Delta_{good}\right)$$

$$(42)$$

for some $\Delta_{sing}, \Delta_{bad}, \Delta_{good} \geq 0$ satisfying

$$\Delta_{sing} + \Delta_{bad} + \Delta_{good} = \Delta(v). \qquad (43)$$

But note that

$$\frac{\Delta_{bad}}{\Delta_{good}} = \frac{\sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot N_{bad}(\sigma, \tau)}{\sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot N_{good}(\sigma, \tau)}$$

$$\leq C \cdot \frac{\mathbb{E}[N_{bad}(\sigma^{(T_{\text{stop}}-1)}, \tau^{(T_{\text{stop}}-1)})]}{\mathbb{E}[N_{good}(\sigma^{(T_{\text{stop}}-1)}, \tau^{(T_{\text{stop}}-1)})]}$$

$$\leq C\gamma \qquad (44)$$

for $C := \frac{k+2p_2\Delta}{k-\Delta-2}$, where the second inequality follows by hypothesis and the first inequality follows by the fact that for $s \in \{\text{bad}, \text{good}\}$,

$$\mathbb{E}[N_s(\sigma^{(T_{\text{stop}}-1)}, \tau^{(T_{\text{stop}}-1)})] = \sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot \mathbb{P}[(S_T, S_T') \text{ terminating}] \cdot N_s(\sigma, \tau)$$

$$\in \left[\frac{k-\Delta-2}{nk}, \frac{k+2p_2\Delta}{nk}\right] \cdot \sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot N_s(\sigma, \tau),$$

where we use the notation $x \in [a, b] \cdot y$ to denote the fact that $a \cdot y \leq x \leq b \cdot y$. The first step above follows by definition and the second step follows by Lemma 4.2.

Finally, observe that (43) and (44) imply that $-k + \lambda_{sing}^* \Delta_{sing} + \lambda_{bad}^* \Delta_{bad} + \lambda_{good}^* \Delta_{good}$ is a convex combination of $-k + \lambda_{sing}^* \Delta(v)$, $-k + \lambda_{good}^* \Delta(v)$, and $-k + \left(\frac{C\gamma}{C\gamma+1} \lambda_{bad}^* + \frac{1}{C\gamma+1} \lambda_{good}^*\right) \Delta(v)$, so in particular from (42) we get that

$$\mathbb{E}[d_H(\sigma^{(T_{\text{stop}})}, \tau^{(T_{\text{stop}})}) - 1] \leq \frac{1}{nk} \left(\sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau]\right)(-k + \lambda_{C\gamma}^* \Delta(v)) \leq \frac{-k + \lambda_{C\gamma}^* \Delta(v)}{k - \Delta - 2},$$

where the final step follows from the fact that

$$\sum_{T,\sigma,\tau} \mathbb{P}[\sigma^{(T-1)} = \sigma, \tau^{(T-1)} = \tau] \cdot \mathbb{P}[(S_T, S_T') \text{ terminating}] = 1$$

and the lower bound of Lemma 4.2. □

C.4 Proof of Lemmas from Section 4.3

Proof of Lemma 4.6. Without loss of generality suppose that $(A_c, B_c, \mathbf{a}^c, \mathbf{b}^c) = (7, 3, (3, 3), (1, 1))$. Let u_1, u_2 be the two c-colored neighbors of v, and denote the elements of $S_\tau(u_1, \sigma(v))$ and $S_\tau(u_2, \sigma(v))$ by $\{u_1, w_1^1, w_1^2\}$ and $\{u_2, w_2^1, w_2^2\}$ respectively. We know that the vertices $\{w_1^1, w_1^2, w_2^1, w_2^2\}$ are all distinct. With probability $\frac{4}{n} \cdot \frac{k-\Delta-1}{k}$, the pair of flips (S, S') chosen under the greedy coupling satisfies $S = S' = S_\sigma(w_j^i, c')$ for some $i, j \in \{1, 2\}$ and $c' \in A_\sigma(w_j^i) \backslash \{\sigma(w_j^i)\}$ (note that $A_\sigma(w_j^i) \backslash \{\sigma(w_j^i)\}$ contains neither $\sigma(v)$ nor c). In this case, the flips are just of vertex w_j^i from color $\sigma(w_j^i)$ to a different color not already present in its neighborhood, so the neighboring coloring pair σ', τ' resulting from the flips is in state GOOD_c. □

Proof of Lemma 4.8. First note that in order for σ', τ' to be in stage BADEND_c given that σ, τ were in stage GOOD_c, it must be that condition (iii) of Definition 4.4 holds. Furthermore, the pair of components (S, S') flipped to get from σ, τ to σ', τ' cannot be terminal, so $S = S'$.

Suppose that $\delta_c > 2$. In this case, the probability that σ, τ leave stage GOOD_c for stage BADEND_c is at most the probability that enough c-colored neighbors of v are flipped so that δ_c becomes at most 2. A Kempe component S outside of D_c and containing at least $(\delta_c - 2)$ c-colored neighbors of v must be flipped in both colorings to achieve this, and the probability the greedy coupling chooses any particular such (S, S) is $p_{\delta_c-2}/(nk)$. The number of such Kempe components is at most $\delta_c \cdot (k-2)$, so by a union bound the probability that δ_c becomes 2 is at most $\frac{\delta_c p_{\delta_c-2} \cdot (k-2)}{nk} < 3/n$.

On the other hand, if $\delta_c < 2$, then σ, τ are not in stage GOOD_c to begin with. So for the rest of the proof, we consider the case of $\delta_c = 2$. We will proceed by casework on $(A_c, B_c, \mathbf{a}^c, \mathbf{b}^c)$, which we will denote as $(A, B, (a_1, a_2), (b_1, b_2))$ for simplicity.

34

Let $\mathcal{E}$ denote the event that σ, τ transition to stage BADEND$_c$. Denote the two c-colored neighbors of v by u_1, u_2. We have that $\mathcal{E} \subseteq \mathcal{E}_1 \cup \mathcal{E}_2$, where $\mathcal{E}_1$ is the event that u_1 or u_2 is flipped in both colorings to a new color, and $\mathcal{E}_2$ is the event that u_1 or u_2 are not flipped but σ, τ nevertheless transition to stage BADEND$_c$. Obviously $\mathbb{P}[\mathcal{E}_1] \leq 2/n$. We now proceed to bound $\mathbb{P}[\mathcal{E}_2]$.

Case 1. *If $a_i > 3$ or $b_i > 3$ for some $i = 1, 2$, then $\mathbb{P}[\mathcal{E}_2] \leq \frac{3}{n}$.*

Proof. Without loss of generality, say that $a_1 > 3$. From the vertices of $S_\tau(u_1, \sigma(v))$ pick out $w, w' \neq u_1$ such that w, w', u_1 form a Kempe component. We have that event $\mathcal{E}_2 \subseteq \mathcal{A} \cup \mathcal{B}$, where $\mathcal{A}$ is the event that all vertices in $S_\tau(u_1, \sigma(v)) \setminus \{u_1, w, w'\}$ are flipped so that $S_{\tau'}(u_1, \sigma(v)) \subseteq \{u_1, w, w'\}$, and $\mathcal{B}$ is the event that w or w' is flipped and no longer belongs to $S_{\tau'}(u_1, \sigma(v))$. Obviously $\mathbb{P}[\mathcal{B}] \leq 2/n$. For $\mathcal{A}$, the $(a_1 - 3)$ neighbors of u_1, w, w' in $S_\tau(u_1, \sigma(v))$ must be flipped at once, which by a union bound occurs with probability at most $\frac{1}{n} \cdot (a_1 - 3) \cdot p_{a_1 - 3} \leq \frac{1}{n}$, where the inequality follows by (11). So $\mathbb{P}[\mathcal{E}_2] \leq \mathbb{P}[\mathcal{A}] + \mathbb{P}[\mathcal{B}] \leq 3/n$. $\square$

Case 2. *If $a_i = 0$ for some i and $b_1, b_2 \leq 3$, or if $b_i = 0$ for some i and $a_1, a_2 \leq 3$, then $P[\mathcal{E}_2] \leq \frac{1}{n}$.*

Proof. Suppose without loss of generality that $a_1 = 0$ and $b_1, b_2 \leq 3$. By the definition of the greedy coupling and the fact that $c \neq \sigma(v), \tau(v)$, $a_1 = 0$ if and only if $S_\tau(u_2, \sigma(v))$ consists of u_1, u_2, w for some $\sigma(v)$-colored $w \in N(u_1) \cup N(u_2)$. So $\mathcal{E}_2$ is a subset of the event that w is flipped to any other color. Thus, $\mathbb{P}[\mathcal{E}_2] \leq \frac{1}{n}$. $\square$

Case 3. *If $1 \leq a_1, a_2, b_1, b_2 \leq 3$, and if (a_1, a_2) and (b_1, b_2) are both not among $\{(1, 1), (3, 3)\}$, then $P[\mathcal{E}_2] \leq \frac{48}{nk}$.*

Proof. Suppose (a_1, a_2) and (b_1, b_2) are both not among $\{(1, 1), (3, 3)\}$. Then $\mathcal{E}_2$ is a subset of the event that the pair of flips (S, S) chosen increases or decreases at least one of a_1, a_2 and decreases or increases at least one of b_1, b_2, respectively. But for a flip S to decrease some a_i for $i \in \{1, 2\}$, it must contain a member of $S_\tau(u_i, \sigma(v))$, and for a flip S to increase some b_j for $j \in \{1, 2\}$, it must contain the color c or $\tau(v)$. There are at most 3 members of $S_\tau(u_i, \sigma(v))$, so the probability of (S, S) both increasing a_i and decreasing b_j is at most $\frac{3}{n} \cdot \frac{2}{k} = \frac{6}{nk}$, and by a union bound over the eight different choices of i, j, and increasing/decreasing, we conclude that $\mathbb{P}[\mathcal{E}_2] \leq \frac{48}{nk}$. $\square$

Case 4. *If $1 \leq a_1, a_2, b_1, b_2 \leq 3$ and exactly one of the tuples (a_1, a_2) and (b_1, b_2) is among $\{(1, 1), (3, 3)\}$, then $P[\mathcal{E}_2] \leq \frac{4\Delta + 48}{nk}$.*

Proof. Suppose $(b_1, b_2) = (1, 1)$. $\mathcal{E}_2 \subseteq \mathcal{X} \cup \mathcal{Y}$, where $\mathcal{X}$ is the event that the pair of flips (S, S) chosen increases or decreases some a_i and decreases or increases some b_i, respectively, and $\mathcal{Y}$ is the event that (a_1, a_2) becomes $(3, 3)$. We already know by Case 3 that $\mathbb{P}[\mathcal{X}] \leq \frac{48}{nk}$. Supposing without loss of generality that $a_1 < 3$, the event $\mathcal{Y}$ is a subset of the event that a neighbor of a vertex in $S_\tau(u_1, \sigma(v))$ is flipped to the color c or $\sigma(v)$. There are at most 2Δ such neighbors, so $\mathbb{P}[\mathcal{Y}] \leq \frac{2\Delta}{n} \cdot \frac{2}{k} = \frac{4\Delta}{nk}$, and thus $\mathbb{P}[\mathcal{E}_2] \leq \frac{4\Delta + 48}{nk}$.

Now suppose $(b_1, b_2) = (3, 3)$. $\mathcal{E}_2 \subseteq \mathcal{X} \cup \mathcal{Z}$ where $\mathcal{X}$ is the event defined above and $\mathcal{Z}$ is the event that (a_1, a_2) becomes $(1, 1)$. Supposing without loss of generality that $a_1 > 1$, $\mathcal{Z}$ is a subset of the event that one of the members of $S_\tau(u_1, \sigma(v))$ other than u_1 is flipped. There are at most two such vertices, so $\mathbb{P}[\mathcal{Z}] \leq 2/n$ and $\mathbb{P}[\mathcal{E}_2] \leq \frac{2k + 48}{nk}$. $\square$

Case 5. *If $1 \leq a_1, a_2, b_1, b_2 \leq 3$ and $(a_1, a_2, b_1, b_2) = (1, 1, 1, 1)$, then $P[\mathcal{E}_2] \leq \frac{4\Delta}{nk}$.*

Proof. $\mathcal{E}_2$ is a subset of the event that one of the neighbors of u_1 or u_2 is flipped to the color $\sigma(v)$ or $\tau(v)$, so $\mathbb{P}[\mathcal{E}_2] \leq \frac{2\Delta}{n} \cdot \frac{2}{k} = \frac{4\Delta}{nk}$. $\square$

Case 6. *If $1 \leq a_1, a_2, b_1, b_2 \leq 3$ and $(a_1, a_2, b_1, b_2) = (3, 3, 3, 3)$, then $P[\mathcal{E}_2] \leq \frac{2}{n}$.*

Proof. $\mathcal{E}_2 \subseteq \mathcal{S} \cup \mathcal{T}$, where $\mathcal{S}$ (resp. $\mathcal{T}$) is the event that all $\sigma(v)$-colored (resp. $\tau(v)$-colored) neighbors in $N(u_1) \cup N(u_2)$ in τ (resp. σ) are flipped to a different color. Consider an arbitrary σ-colored neighbor w of u_1. $\mathcal{S}$ is a subset of the event that w is flipped, so $\mathbb{P}[\mathcal{S}] \leq 1/n$. We can bound $\mathbb{P}[\mathcal{T}]$ similarly, so $\mathbb{P}[\mathcal{E}_2] \leq 2/n$. $\qquad \square$

Of the upper bounds on $\mathbb{P}[\mathcal{E}_2]$ in all of the above cases, the bound of $3/n$ from Case 1 is the greatest when $k \geq 1.833\Delta$, completing the proof of Lemma 4.8. $\qquad \square$

D   Missing Proofs from Section 5

D.1   Proof of Lemma 5.1

Proof. We need to prove that for every m and every realizable configuration $(A, B; \mathbf{a}, \mathbf{b})$ of size m different from the four excluded ones, we have

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + m\hat{\lambda}. \tag{45}$$

It is straightforward to check that for every $\alpha \in \{1, \ldots, 6\}$ the given assignment satisfies $\alpha \hat{p}_\alpha \leq 1$, $(\alpha - 1)\hat{p}_\alpha \leq 1/3$ and $(\alpha - 2)\hat{p}_\alpha \leq (3\hat{\lambda} - 5)/2$; we will use some of these inequalities in the proof.

Let us first assume that $c \neq \sigma(v), \tau(v)$, so $A = 1 + \sum_i a_i$ and $B = 1 + \sum_i b_i$. If $m = 1$, then any realizable configuration has the form $(i + 1, j + 1, (i), (j))$. Observe that (21) correspond to the constraint $H(i + 1, j + 1, (i), (j)) \leq -1 + \lambda$. We may assume that $i, j \leq 6$ and that $j \neq 1$ as otherwise we obtain weaker constraints, since $\hat{p}_\alpha = 0$ for $\alpha \geq 7$. Thus (45) follows for every configuration of size 1 in Linear Program 5, from the constraints in Linear Program 6.

If $m = 2$, as $\hat{p}_\alpha = 0$ for every $\alpha \geq 7$, there is a finite amount of non-trivial realizable non-extremal configurations $(A, B; \mathbf{a}, \mathbf{b})$ of size 2. One can check by computer that for the given values of $\hat{\mathbf{p}}$ and $\hat{\lambda}$, any configuration of size 2 in Linear Program 5 satisfies $H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + 2\hat{\lambda}$, with equality if and only if the configuration is $(5, 3, (2, 2), (1, 1))$ or $(3, 5, (1, 1), (2, 2))$.

Suppose that $m = 3$. By Lemma 3.1, recall that

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq (A - 2)p_A + (B - 2)p_B + \sum_i (a_i p_{a_i} + b_i p_{b_i} - \min\{p_{a_i}, p_{b_i}\}).$$

Using the properties of $\hat{\mathbf{p}}$, we have $a_i p_{a_i} + b_i p_{b_i} - \min\{p_{a_i}, p_{b_i}\} \leq 4/3$. Thus,

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq (A - 2)p_A + (B - 2)p_B + \sum_i (a_i p_{a_i} + b_i p_{b_i} - \min\{p_{a_i}, p_{b_i}\}) \leq -1 + 3\hat{\lambda}.$$

If $m \geq 4$ and since $\hat{\lambda} > 4/3$, we have

$$H(A, B; \mathbf{a}, \mathbf{b}) \leq 3\hat{\lambda} - 5 + 4m/3 \leq -1 + m\hat{\lambda}.$$

We finally deal with the case $c \in \{\sigma(v), \tau(v)\}$. If $c = \tau(v)$, Remark A.2 and $\alpha p_\alpha \leq 1$ implies $H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + m \leq -1 + m\hat{\lambda}$. If $c = \sigma(v)$ and $m = 1$ then Remark 2.2 implies $H(A, B; \mathbf{a}, \mathbf{b}) = 0 \leq -1 + \hat{\lambda}$. Finally, if $c = \sigma(v)$ and $m \geq 2$, then (10) holds for $\hat{\lambda}$ as $\alpha p_\alpha \leq 1$, $(\alpha - 1)p_\alpha \leq 1/3$, so $H(A, B; \mathbf{a}, \mathbf{b}) \leq -1 + m\hat{\lambda}$.

We conclude that $\hat{\mathbf{p}}$ is a feasible solution to Linear Program 5 with objective value $\hat{\lambda}$. $\qquad \square$

D.2 Proof of Lemma 5.2

Proof. Recall that the extremal configurations for our choice of flip parameters are $(3, 2, (2), (1))$ and $(7, 3, (3, 3), (1, 1))$, up to symmetries, and $U_c = \{u_1, \ldots, u_m\}$, where $m = \delta_c$.

Assume first that $c \in C^i_{\sigma, \tau}(v)$ for some $i \in \{1, 2\}$. Consider the sets of components

$$\mathcal{S}_0 := \{S \in \overline{\mathcal{D}} : c \notin C_{\sigma_S, \tau_S}(v)\},$$

$$\mathcal{S}_2 := \{S \in \overline{\mathcal{D}} : c \in C^2_{\sigma_S, \tau_S}(v)\}.$$

Note that when $i = 1$, then for every $S \in \overline{\mathcal{D}} \setminus (\mathcal{S}_0 \cup \mathcal{S}_2)$ we have $\xi_{\sigma, \tau}(v, c, S) \le 0$; therefore,

$$\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \le \overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_0) + \overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_2).$$

Note that when $i = 2$, then for every $S \in \overline{\mathcal{D}} \setminus \mathcal{S}_0$ we have $\xi_{\sigma, \tau}(v, c, S) \le 0$; therefore,

$$\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \le \overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_0).$$

We proceed to bound $\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_0)$ for $i \in \{1, 2\}$. Without loss of generality, assume that $a_1 > b_1$. Let $w \in S_\tau(u_1, \sigma(v))$ with $\tau(w) = \sigma(v)$; we note that $w \notin U_c \cup \{v\}$ and that such a vertex always exists as $a_1 \ge 2$. Choose a color $c' \in [k]$ with $c' \notin \sigma(N(w)) \cup \{\sigma(v), \tau(v)\}$. Let $S = S_\sigma(w, c') \in \overline{\mathcal{D}}$. As $S = \{w\}$, (σ_S, τ_S) has either a $(2, 2, (1), (1))$ or a $(j + 4, 3, (j, 3), (1, 1))$ (with $j \in \{1, 2\}$) configuration for c, i.e. $c \notin C_{\sigma_S, \tau_S}(v)$. As there are at least $k - \Delta - 2$ choices for c' and as $p_{|S|} = p_1 = 1$, we have

$$\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_0) \le -\frac{\eta(k - \Delta - 2)}{\Delta} \cdot i.$$

Now we bound $\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_2)$, provided that $i = 1$. Let $S \in \mathcal{S}_2$, then $|S \cap (N(v) \setminus \{u_1\})| \ge 1$ and if $u \in S \cap (N(v) \setminus \{u_1\})$, then $\sigma_S(u) = c$. Thus, S can be described as $S = S_\sigma(u, c)$ for $u \in N(v)$, implying that $|\mathcal{S}_2| \le \Delta$. Moreover, $|S| \ge 2$ as at least two vertices need to change their color to transform an extremal 1-configuration into an extremal 2-configuration. Since $p_{|S|} \le p_2 \le \frac{1}{3}$ and $\xi_{\sigma, \tau}(v, c, S) = 1$, we have

$$\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_2) \le \frac{\eta}{3}.$$

From the bounds on $\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_0)$ and $\overline{\nabla_B}(\sigma, \tau, c, \mathcal{S}_2)$ derived above, we obtain that for $i \in \{1, 2\}$ and $c \in C^i_{\sigma, \tau}(v)$

$$\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \le -\frac{\eta \left(k - \frac{4\Delta}{3} - 2\right)}{\Delta} \cdot i \le -i\eta \left(\frac{k}{\Delta} - \frac{3}{2}\right),$$

and this proves the first statement.

To prove the second statement, assume that $c \notin C_{\sigma, \tau}(v)$ and let $\mathcal{T} := \{S \in \overline{\mathcal{D}} : c \in C_{\sigma_S, \tau_S}(v)\}$. Again, for every $S \in \overline{\mathcal{D}} \setminus \mathcal{T}$, we have $\xi_{\sigma, \tau}(v, c, S) \le 0$. Therefore,

$$\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \le \overline{\nabla_B}(\sigma, \tau, c, \mathcal{T}).$$

Define $U^S_c := N(v) \cap (\sigma_S)^{-1}(c)$ with $m^S := |U^S_c|$ and note that $m^S \le 2$. Consider the partition $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ with

$$\mathcal{T}_1 := \{S \in \overline{\mathcal{D}} : U_c \setminus U^S_c \ne \emptyset\},$$

$$\mathcal{T}_2 := \{S \in \overline{\mathcal{D}} : U^S_c \setminus U_c \ne \emptyset\} \setminus \mathcal{T}_1,$$

$$\mathcal{T}_3 := \{S \in \overline{\mathcal{D}} : U^S_c = U_c\}.$$

For every $S \in \mathcal{T}_3$, if $c \in C^1_{\sigma_S, \tau_S}(v)$, let $(A^S, B^S, (a_1^S), (b_1^S))$ be the extremal 1-configuration for c in (σ_S, τ_S) and if $c \in C^2_{\sigma_S, \tau_S}(v)$, let $(A^S, B^S, (a_1^S, a_2^S), (b_1^S, b_2^S))$ be the extremal 2-configuration for c in (σ_S, τ_S). Recall that $(A, B; (a_1, \ldots, a_m), (b_1, \ldots, b_m))$ denotes the configuration for c in (σ, τ). As it is non-extremal, there exists $x \in \{a, b\}$ and $j \in [m^S]$, such that $x_j \neq x_j^S$. Note that $x_j^S \leq 3$.

Consider the partition $\mathcal{T}_3 = \mathcal{T}_3^+ \cup \mathcal{T}_3^-$ with

$$\mathcal{T}_3^+ := \{S \in \mathcal{T}_3 : x_j > x_j^S\} \,,$$
$$\mathcal{T}_3^- := \{S \in \mathcal{T}_3 : x_j < x_j^S\} \,.$$

To bound the size of $\mathcal{S}' \in \{\mathcal{T}_1, \mathcal{T}_3^+\}$ we will proceed as follows. For every $S \in \mathcal{S}'$, there is a vertex in a Kempe component of either σ or τ that does not belong to the corresponding component in either σ_S or τ_S. If there exists $R(\mathcal{S}') \subseteq S_\sigma(v, c) \cup S_\tau(v, c)$ such that $S \cap R(\mathcal{S}') \neq \emptyset$ for every $S \in \mathcal{S}'$, then, any $S \in \mathcal{S}'$ can be described as $S = S_\sigma(w, c')$ for $w \in R(\mathcal{S}')$ and $c' \in [k]$, and $|\mathcal{S}'| \leq |R(\mathcal{S}')| k$.

If $\mathcal{S}' = \mathcal{T}_1$ and $S \in \mathcal{S}'$, then observe that $|S \cap U_c| = |U_c \setminus U_c^S| \geq \max\{m - m^S, 1\}$. Let $\ell = \min\{m^S + 1, m\}$. If $R(\mathcal{T}_1) = R_1 = \{u_1, \ldots, u_\ell\}$, it follows that $|S \cap R_1| \geq |S \cap U_c| - (m - (m^S + 1)) \geq 1$ and $|\mathcal{T}_1| \leq (m^S + 1) k \leq 3k$.

If $\mathcal{S}' = \mathcal{T}_3^+$ and $S \in \mathcal{S}'$, recall that $x_j > x_j^S$ and set $\varphi = \sigma$ if $x = b$ and $\varphi = \tau$ if $x = a$, and let $\pi \in \{\sigma, \tau\} \setminus \{\varphi\}$. Let $R(\mathcal{T}_3^+) = R_3$ be an arbitrary set of x_j^S vertices in $S_\varphi(u_j, \pi(v)) \setminus \{u_j\}$. As $u_j \notin R_3$, we have $S \cap R_3 \neq \emptyset$. Since there are 4 choices for the extremal configuration, we have $|\mathcal{T}_3^+| \leq 4 x_j^S k \leq 12k$.

To bound the size of $\mathcal{S}' \in \{\mathcal{T}_2, \mathcal{T}_3^-\}$ we will proceed as follows. For every $S \in \mathcal{S}'$, there is a vertex in the neighborhood of a Kempe component of either σ or τ, that belongs to the corresponding component in either σ_S or τ_S. If there exists a set $N(\mathcal{S}')$ of neighbors of $S_\varphi(v, c)$ such that $S \cap N(\mathcal{S}') \neq \emptyset$ for every $S \in \mathcal{S}'$, then, any $S \in \mathcal{S}'$ can be described as $S = S_\varphi(w, c')$ for $w \in N(\mathcal{S}')$ and a unique $c' \in \{c, \pi(v)\}$, and $|\mathcal{S}'| \leq |N(\mathcal{S}')|$.

If $\mathcal{S}' = \mathcal{T}_2$ and $S \in \mathcal{S}'$, then let $N(\mathcal{T}_2) = N_2 = N(v) \setminus U_c$. Clearly $S \cap N_2 \neq \emptyset$ and $|\mathcal{T}_2| \leq \Delta$.

If $\mathcal{S}' = \mathcal{T}_3^-$ and $S \in \mathcal{S}'$, recall that $x_j < x_j^S$ and set $\varphi = \sigma$ if $x = b$ and $\varphi = \tau$ if $x = a$, and let $\pi \in \{\sigma, \tau\} \setminus \{\varphi\}$. Let $N(\mathcal{T}_3^-) = N_3$ be the set of neighbors of $S_\varphi(u_j, \pi(v))$, which satisfies $S \cap N_3 \neq \emptyset$. As $S \in \mathcal{T}_3^-$, $|S| \leq x_j \Delta \leq (x_j^S - 1) \Delta \leq 2\Delta$. Since there are 4 choices for the extremal configuration, we have $|\mathcal{T}_3^-| \leq 8\Delta$.

Since $p_{|S|} \leq 1$ and $\xi_{\sigma, \tau}(v, c, S) \leq 2$, we conclude the second statement of the lemma,

$$\overline{\nabla_B}(\sigma, \tau, c, \overline{\mathcal{D}}) \leq \overline{\nabla_B}(\sigma, \tau, c, \mathcal{T}) \leq \frac{2\eta}{\Delta}(3k + 12k + \Delta + 8\Delta) = 2\eta\left(9 + \frac{15k}{\Delta}\right) \,. \qquad \square$$

E Proof of Theorem 6.1

Proof. The proof follows the same lines as the proof of Theorem 1.1 presented in Section 5, although a similar analysis can be done using the multi-step coupling approach presented in Section 4. We will describe the proof strategy, stressing the parts where the argument is different for list-coloring and omitting the ones that are straightforward adaptations of the coloring case.

Let $\sigma, \tau \in \Omega^L$ that differ only at a vertex v. For the rest of the proof, we will assume $c \neq \sigma(v), \tau(v)$; the case $c \in \{\sigma(v), \tau(v)\}$ produces weaker constraints and can be dealt similarly as in the non-list-colouring case (see Remark A.2). For $\varphi \in \{\sigma, \tau\} \subseteq \Omega^L$, $\pi \in \{\sigma, \tau\} \setminus \{\varphi\}$, $c \in L(v)$ and $U_c = \{u_1, \ldots, u_m\}$ the set of neighbors of v with color c with $m = \delta_c$, we define the configurations $(A^L, B^L, (a_1^L, \ldots, a_m^L); (b_1^L, \ldots, b_m^L))$ for c in (σ, τ) as before, with the sole difference that we also set $A^L = 0$ if $S_\sigma(v, c)$ is un-flippable, $B^L = 0$ if $S_\tau(v, c)$ is un-flippable, $a_i^L = 0$ if $S_\tau(u_i, \sigma(v))$ is

un-flippable and $b_i^L = 0$ if $S_\sigma(u_i, \tau(v))$ is un-flippable. As $c \neq \sigma(v), \tau(v)$, if $A^L \neq$, then $A^L = 1 + a_1^L + \cdots + a_m^L$, and similarly for B^L. We define $i_{\max}^L$, $j_{\max}^L$, $a_{\max}^L$ and $b_{\max}^L$ analogously as before, and note that the latter two can be zero. Define $q_i(L)$ and $q_i'(L)$ as in (32) and (33) for the list version of the parameters.

According to this, we use the same definition of extremal configurations, metric d on Ω^L, d_H and d_B. Again, for any pair $\sigma', \tau' \in \Omega^L$, we have $d(\sigma', \tau') \leq d_H(\sigma', \tau')$, which implies that $d_B(\sigma', \tau') \geq 0$. We use the same coupling as the one defined in Appendix A and define ∇^L, ∇_H^L and ∇_B^L analogously as for colorings. Fix the flip parameters $\hat{\mathbf{p}}$ provided in Observation 5.1.

We will prove an analogue of Corollary 5.1 to bound ∇_H^L for list-colorings. As in Section 5.2, we have

$$\nabla_H^L(\sigma, \tau) = \sum_{c \in \mathbb{N}} \nabla_H^L(\sigma, \tau, c) .$$

Suppose first that $m = 0$. Then $c \in L(v)$ and $\nabla_H^L(\sigma, \tau, c) = -1$.

If $m \geq 1$, the list analogues of equations (5) and (6) hold, so

$$\nabla_H^L(\sigma, \tau, c) \leq (A^L - a_{\max}^L - 1)p_{a^L} + (B^L - b_{\max}^L - 1)p_{b^L}$$
$$+ \sum_{i \in [m]} (a_i^L q_i(L) + b_i^L q_i'(L) - \min\{q_i(L), q_i'(L)\}) . \tag{46}$$

We will bound each term $\nabla_H^L(\sigma, \tau, c)$ depending on whether $c \in L(v)$ or $c \notin L(v)$.

If $c \notin L(v)$, then it suffices to show that $\nabla_H^L(\sigma, \tau, c) \leq m\hat{\lambda}$. Note that $A^L = B^L = 0$, $q_i(L) = p_{a_i^L}$ and $q_i'(L) = p_{b_i^L}$ for every $i \in [m]$. Let $\{c_i^L, d_i^L\} = \{a_i^L, b_i^L\}$ with $p_{c_i^L} \geq p_{d_i^L}$. Using (46), we obtain

$$\nabla_H^L(\sigma, \tau, c) \leq \sum_{i \in [m]} (a_i^L p_{a_i^L} + b_i^L p_{b_i^L} - \min\{p_{a_i^L}, p_{b_i^L}\})$$
$$= \sum_{i \in [m]} c_i^L p_{c_i^L} + (d_i^L - 1)p_{d_i^L} \leq \frac{4}{3}m < m\hat{\lambda} ,$$

where we have used that $\alpha p_\alpha \leq 1$ and $(\alpha - 1)p_\alpha \leq \frac{1}{3}$.

Now assume that $c \in L(v)$. We will compare these bounds with the ones we obtained in Section 5.1 by plugging the values of the configuration $(A^L, B^L, (a_1^L, \ldots, a_m^L), (b_1^L, \ldots, b_m^L))$. Observe that there are only two differences with respect to non-list-colorings; first, A^L and B^L can be zero, and second, $a_{\max}^L$ and $b_{\max}^L$ can be zero. Recall that $p_0 = 0$. It is important to stress that, since $c \in L(v)$, $a_{\max}^L = 0$ implies $A^L = 0$, and similarly for B^L. Therefore, the only difference between (46) and the bound obtained from (5), are the cases where either $A^L = 0$ or $B^L = 0$. If $A^L = a_{\max}^L = 0$, then the total contribution of this part is zero and analogously for B^L. Therefore, the only interesting case is when $A^L = 0$ and $a_{\max}^L \neq 0$; in this case $m \geq 2$. Since $A^L = 0$ and $c \in L(v)$, there exists $j \in [m]$ such that $a_j^L = 0$. Consider the configuration of size $m - 1$

$$(A, B; (a_1^L, \ldots, a_{j-1}^L, a_{j+1}^L, \ldots, a_m^L), (b_1^L, \ldots, b_{j-1}^L, b_{j+1}^L, \ldots, b_m^L)) . \tag{47}$$

where $A = 1 + \sum_{i \neq j} a_i^L$ and $B = 1 + \sum_{i \neq j} b_i^L$. Let $b_{\max}$ be the maximum of the b_i^L with $i \neq j$ and note that $b_{\max} \leq b_{\max}^L$. Recall that $\hat{\mathbf{p}}$ is a feasible solution for Linear Program 1 with objective value $\frac{11}{6}$ and a feasible solution of Linear Program 5 with objective value $\frac{161}{88}$. If $m \geq 4$, then the configuration of size $m - 1$ in (47) for c is non-extremal and

$$\nabla_H(\sigma, \tau, c) \leq (A - a_{\max}^L - 1)p_A + (B - b_{\max} - 1)p_B + \sum_{i \neq j} a_i^L q_i + b_i^L q_i' - \min\{q_i, q_i'\}$$

$$\leq \frac{161}{88}(m - 1) - 1 .$$

If $1 \le m \le 3$, then (47) can be extremal and

$$\nabla_H(\sigma, \tau, c) \le (A - a_{\max}^L - 1)p_A + (B - b_{\max} - 1)p_B + \sum_{i \ne j} a_i^L q_i + b_i^L q_i' - \min\{q_i, q_i'\}$$

$$\le \frac{11}{6}(m-1) - 1 \le \frac{161}{88}(m-1) - \frac{131}{132} .$$

For $i = i_{\max}^L$ we have $q_i = p_{a_{\max}} - p_A$ and $q_i(L) = p_{a_{\max}}$. Moreover, we have $q_j' = 0$ and $q_j'(L) \le p_{b_j^L}$. We may assume that $b_{\max}^L \ne 0$, as otherwise we have $b = b_{\max} = 0$ and the contribution of this part is zero, as before. Using these bounds and (46), we obtain that for any such $c \in [k]$

$$\nabla_H^L(\sigma, \tau, c) \le (B^L - b_{\max}^L - 1)p_{B^L} + \sum_{i \in [m]} (a_i^L q_i(L) + b_i^L q_i'(L) - \min\{q_i(L), q_i'(L)\})$$

$$\le \nabla_H(\sigma, \tau, c) - (A - 2a_{\max}^L - 1)p_A + (B^L - b_{\max}^L - 1)p_{B^L} + b_j^L p_{b_j^L}$$

$$\le \nabla_H(\sigma, \tau, c) + (A - 1)p_A + (B^L - 2)p_{B^L} + b_j^L p_{b_j^L}$$

$$\le \nabla_H(\sigma, \tau, c) + \frac{19}{12}$$

$$\le \frac{161}{88} \cdot m - 1 ,$$

where we have used that $b_{\max}^L \ge 1$, $A \ge a_{\max}^L + 1$, $\alpha p_\alpha \le 1$, $(\alpha - 1)p_\alpha \le \frac{1}{3}$ and $(\alpha - 2)p_\alpha \le \frac{1}{4}$. Thus, Corollary 5.1 also holds for ∇_H^L.

Corollary 5.2 also holds for ∇_B^L as well, since all the negative contributions on the bound are given by Kempe components $S = S_\sigma(u, c)$ of size 1, which are always flippable as $c \in L(u)$. The positive contributions of the Kempe components is still bounded by the same quantity since, in the worst case, they are all flippable.

Using the same flip parameters and reasoning as in the proof of Theorem 5.1, it follows that for every k-list assignment L with $k \ge (\frac{11}{6} - \eta)\Delta$, flip dynamics for L-colorings satisfies $\tau_{\mathrm{mix}}(\epsilon) = O(n(\log n + \log \epsilon^{-1}))$, concluding the proof of Theorem 6.1.

$\square$