6.S891 Lecture 19: Disagreement Percolation

Kuikui Liu

November 16, 2023

In this lecture, we return to coupling-based methods for establishing spectral independence. Previously, we saw how to bound the transportation distance between two distributions μ, ν assuming there is a local Markov chain for μ which admits a path coupling proof of rapid mixing. Here, we will directly build couplings recursively, and highlight the flexibility of this technique. For a distribution μ on $[q]^n$, recall that we say μ is *η-coupling independent* if for all $i \in [n]$ and $\mathfrak{b}, \mathfrak{c} \in [q]$,

$$\mathscr{W}_1\left(\mu^{i\leftarrow\mathfrak{b}}, \mu^{i\leftarrow\mathfrak{c}}\right) \leq 1 + \eta.$$

Throughout, all instantiations of Wasserstein distance are w.r.t. the Hamming metric. We previously proved that η-coupling independence implies η-spectral independence. Since we'll mainly consider iteratively constructed couplings in graphical settings, and since we are aiming to minimize Hamming distance, i.e. the number of *disagreements* between two samples, we broadly refer to the class of methods in this lecture as "disagreement percolation": Our goal is to prevent the components of disagreeing vertices from growing unboundedly.

1 Warm-Up: Monomer-Dimer Model

To illustrate the basic idea, let us return to the classic monomer-dimer model. Given a graph $G = (V, E)$ and an activity parameter $\lambda \geq 0$, the monomer-dimer Gibbs distribution is given by

$$\mu(M) \propto \lambda^{|M|}, \qquad \forall \text{ matchings } M \subseteq E.$$

Lemma 1.1. *Let μ be the monomer-dimer Gibbs distribution with activity $\lambda \geq 0$ on a graph $G = (V, E)$ of maximum degree Δ. Then for every $e \in E$,*

$$\mathscr{W}_1\left(\mu^{e\leftarrow\mathsf{in}}, \mu^{e\leftarrow\mathsf{out}}\right) \leq O\left(\lambda^2\Delta^2\right),$$

and μ is $O\left(\lambda^2\Delta^2\right)$-spectrally independent.

There are many other ways of proving the spectral independence claim (e.g. via correlation decay, zero-freeness, etc.), although these don't yield coupling independence. Combined with the local-to-global theorems from the previous lectures, this immediately implies $O(n \log n)$ mixing for Glauber dynamics on matchings in bounded-degree graphs for arbitrary $\lambda \geq 0$. Path coupling notably fails beyond a certain threshold of λ, and so the previous approach for establishing coupling independence using contractive Markov chains does not work. We build the coupling directly.

Proof of Lemma 1.1. Fix an edge $e \in E$. Our goal is to design a process/algorithm, not necessarily efficient, which jointly samples a pair $\tau, \sigma : E \to \{\mathsf{in}, \mathsf{out}\}$ such that marginally $\tau \sim \mu^{e\leftarrow\mathsf{in}}, \sigma \sim \mu^{e\leftarrow\mathsf{out}}$, and $d_H(\tau, \sigma)$ is small in expectation. To achieve this, we sample τ, σ iteratively. Let us first give a high-level skeleton of the construction. Let $e_1, \ldots, e_m$ be some (possibly random) ordering of the edges of E with $e_1 = e$, which we have yet to specify. We build a random *sequence* $\tau_k, \sigma_k : \{e_1, \ldots, e_k\} \to \{\mathsf{in}, \mathsf{out}\}$, with the invariant that $\tau_k(e_k) \sim \mu_{e_k}^{\tau_{k-1}}$ and $\sigma_k(e_k) \sim \mu_{e_k}^{\sigma_{k-1}}$ for every k; in particular, $\tau_1(e_1) = \mathsf{in}$ and $\sigma_1(e_1) = \mathsf{out}$ with probability 1. This enforces $\tau \sim \mu^{e\leftarrow\mathsf{in}}, \sigma \sim \mu^{e\leftarrow\mathsf{out}}$, while allowing us to locally argue about how the set of *disagreements* evolves in each step.

Now let us implement it in full, with some minor deviations from the above outline. We actually traverse the vertices $v_1, \ldots, v_n$ in some order, which will be intertwined with our sequence

of sampled partial configurations. For each vertex $v \in V$, write $E(v) = \{e \in E : e \sim v\}$ for its set of incident edges. We build this ordering inductively as follows: Assume we have visited vertices $v_1, \dots, v_\ell$ already, and built up two partial configurations $\tau_\ell, \sigma_\ell : E_\ell \to \{\mathsf{in}, \mathsf{out}\}$ where $E_\ell = \bigcup_{i=1}^{\ell} E(v_i)$ is the set of visited edges. Let $D(\ell) \stackrel{\mathsf{def}}{=} \{f \in E_\ell : \tau_\ell(f) \neq \sigma_\ell(f)\}$ denote the current set of disagreements.

- **Phase 1:** Suppose there is an unvisited vertex v which is incident to some edge of disagreement $e \in D(\ell)$. We optimally couple $\mu^{\tau_\ell}_{E(v)\setminus E_\ell}$ and $\mu^{\sigma_\ell}_{E(v)\setminus E_\ell}$. This determines the assignments of all edges incident to v, thus extending $E_\ell, \tau_\ell, \sigma_\ell, D(\ell)$ by (at most) $|E(v) \setminus E_\ell|$ many edges. Set $v_{\ell+1} = v$ and mark v as visited.
- **Phase 2:** Otherwise, there is no such v. Then either all vertices and edges have been visited, in which case the coupling terminates and have our samples τ, σ, or there are no longer any vertices incident to edges of disagreement. In the latter case, we optimally couple $\mu^{\tau_\ell}, \mu^{\sigma_\ell}$ and again terminate the coupling.

Note that we are traversing G and ensuring that each time a new vertex is visited, all of its incident edges receive assignments from $\{\mathsf{in}, \mathsf{out}\}$ before moving on to the next vertex. This traversal is almost but not exactly the same as depth-first traversal. Obviously, the set of visited vertices and edges is always connected.

Claim 1.2 (Structure of the Disagreements)**.** *Under this coupling, for any new vertex v visited during Phase 1, at most one of its incident unvisited edges is added to the disagreement set. Furthermore, this occurs with probability at most $\frac{\lambda\Delta}{1+\lambda\Delta}$. Note this implies that $D(\ell)$ is a single path or cycle, for every ℓ.*

Proof. Let v be such a vertex and let ℓ be the current step at which we discover v. Let e be any edge of disagreement in $D(\ell)$ which is incident to v. Since $\tau_\ell(e) \neq \sigma_\ell(e)$, one of them must be in; without loss of generality, assume $\tau_\ell(e) = \mathsf{in}, \sigma_\ell(e) = \mathsf{out}$. Then under μ^{τ_ℓ}, all edges incident to v must be out with probability 1 by the matching constraint. Similarly, under μ^{σ_ℓ}, at most one edge $f^* \in E(v) \setminus E_\ell$ can be assigned in, giving our new disagreeing edge; all other edges must be assigned out w.r.t. both τ_ℓ and σ_ℓ. This establishes the first claim.

For the second claim, let H denote the graph obtained by deleting $\sigma_\ell^{-1}(\mathsf{out})$ and all vertices incident to an edge in $\sigma_\ell^{-1}(\mathsf{in})$. Note that the probability of there being such a new disagreeing edge f^* is given by

$$\Pr_{M\sim\mu^{\sigma_\ell}}[f \in M \text{ for some } f \sim v, f \notin E_\ell] = \frac{\lambda \sum_{u\in H: u\sim v} Z_{H-u-v}(\lambda)}{Z_{H-v}(\lambda) + \lambda \sum_{u\in H: u\sim v} Z_{H-u-v}(\lambda)}.$$

Every matching which contains some $f \sim v$ also yields a matching not saturating v, since we can just remove f. Hence, $Z_{H-v}(\lambda) \geq Z_{H-u-v}(\lambda)$ for all $u \sim v$ in H, and we see that this marginal probability is at most $\frac{\lambda\cdot|N_H(v)|}{1+\lambda\cdot|N_H(v)|} \leq \frac{\lambda\Delta}{1+\lambda\Delta}$. The final claim that $D(\ell)$ is always either a path or cycle follows immediately from the first claim. □

Claim 1.3 (Isolating the Disagreements)**.** *Suppose we have entered Phase 2 at some step ℓ. Then we can sample $\tau \sim \mu^{\tau_\ell}, \sigma \sim \mu^{\sigma_\ell}$ such that $\tau(e) = \sigma(e)$ for all $e \notin E_\ell$ with probability 1. In other words, we can perfectly couple all remaining unvisited edges once we have entered Phase 2, which must occur entirely after Phase 1.*

Proof. We claim that the set of disagreeing edges $D(\ell)$ must be completely surrounded by edges which are pinned to out in both τ_ℓ, σ_ℓ. To see this, note that if we are in Phase 2, then for any visited vertex, we have completely determined the status of its incident edges in both τ_k, σ_k. In particular, if e is any edge not in $D(\ell)$ but incident to some disagreeing edge f, then $\tau_\ell(e), \sigma_\ell(e)$ have been determined. Furthermore, we claim that $\tau_\ell(e) = \sigma_\ell(e) = \mathsf{out}$. Indeed, since $\tau_\ell(f) \neq \sigma_\ell(f)$, either $\tau_\ell(f) = \mathsf{in}$ or $\sigma_\ell(f) = \mathsf{in}$, which forces $\tau_\ell(e) = \mathsf{out}$ or $\sigma_\ell(e) = \mathsf{out}$, respectively. But because e itself is not disagreeing, we obtain $\tau_\ell(e) = \sigma_\ell(e) = \mathsf{out}$.

Now let H denote the graph obtained by deleting all vertices participating in a disagreeing edge. The above argument implies that the conditional marginals $\mu^{\tau_\ell}_{E\setminus E_\ell}, \mu^{\sigma_\ell}_{E\setminus E_\ell}$ are both just the monomer-dimer Gibbs distribution μ_H of H. In particular, they are identical distributions and can be coupled perfectly. □

In light of Claim 1.2 and Claim 1.3, let A_ℓ denote the event that Phase 1 terminates after completely processing the ℓth visited vertex. Then

$$\begin{aligned}
\mathbb{E}\left[d_H(\tau,\sigma)\right] &= \sum_{\ell=1}^{n} \mathbb{E}\left[d_H\left(\tau,\sigma\right) \mid A_\ell\right] \cdot \Pr[A_\ell] && \text{(Claim 1.3)} \\
&\lesssim \sum_{\ell=1}^{n} \ell \cdot \Pr[A_\ell] && \text{(Claim 1.2: First part)} \\
&\lesssim \sum_{\ell=1}^{\infty} \ell \cdot \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^{\ell} && \text{(Claim 1.2: Second part)} \\
&\leq O\left(\lambda^2\Delta^2\right).
\end{aligned}$$

□

This coupling process, and modifications of it, are the basis of [CZ23; CG23]. The main application in [CZ23] is to the even subgraphs model, which recall is intimately related to the ferromagnetic Ising model. [CG23] studied an extension of matchings and edge covers to *b-matchings* and *b-edge covers*, where every vertex is incident to at most (resp. at least) b edges. They generalize this considerably to "Holant-type" problems, which were previously studied through the lens of zero-freeness. These ideas will also be important when we study the sampling Lovász Local Lemma.

2 Spin Systems on Graphs with Large Girth

In the remainder of this lecture, we study multi-spin systems (e.g. proper colorings) in graphs. We use a similar coupling method to give a general reduction from spectral independence on graphs of sufficiently large (but constant) girth,[1] to correlation decay on trees. When the number of spins is 2, this was previously achieved using the beautiful self-avoiding walk tree/computation tree gadget, even without a girth assumption [Wei06]. Unfortunately, even though there is an analogous computation tree for multi-spin systems [GK12; LY13; GKM15], we do not know how to combine this with contraction of the tree recursion to deduce spectral independence. Very roughly speaking, the main obstacle is that the computation tree for q-spin systems recursively branches into Δq many distinct problem subinstances, rather than Δ. While this remains an outstanding open problem (as of this writing), we show here how to circumvent this barrier if we impose additional girth lower bounds.

Theorem 2.1 (Very Informal; [Che+23])**.** *Suppose a spin system (e.g. proper colorings) has "strong exponential decay of correlations" on all trees of maximum degree Δ. Then there exists a constant $g = g(\Delta) > 0$ such that the spin system has $O(1)$-spectral independence for all graphs of maximum degree Δ and girth at least g.*

Remark 1. We will see that $g(\Delta)$ is approximately of order roughly $\log \Delta$, but for now, let us leave it as a parameter to be determined later.

For concreteness, we specialize our discussion to the case of proper q-colorings of graphs. For $\Delta, q \in \mathbb{N}$ fixed, which will be clear from context, we write μ to denote the uniform measure over proper q-colorings of a graph $G = (V, E)$ of maximum degree $\leq \Delta$. One significant corollary of Theorem 2.1 is fast mixing for proper colorings almost down to the conjectured phase transition threshold $\Delta + 2$,[2] at least when the graph has sufficiently large girth.

Corollary 2.2 ([Che+23])**.** *Fix arbitrary constants $\Delta \in \mathbb{N}$ and $\epsilon > 0$. Then for every graph $G = (V, E)$ of maximum degree Δ and girth $\widetilde{\Omega}\left(\frac{1}{\epsilon^2}\log\Delta\right)$, and every integer $q \geq (1+\epsilon)\Delta$, the uniform distribution μ over proper q-colorings of G is $O(1)$-spectral independent. Furthermore, Glauber dynamics mixes in $O(n \log n)$-steps.*

Our main goal in this section is to establish Theorem 2.1. To make things more precise, "strong exponential decay of correlations" will mean the following:

[1] Recall that the girth of a graph is the length of its shortest cycle.

[2] For Glauber dynamics, the conjectured threshold for fast mixing is $\Delta + 2$, while for efficient sampling, it is $\Delta + 1$.

- **Strong Spatial Mixing:** There exists constants $0 < \delta < 1$ and $C_{\mathsf{SSM}} > 0$ such that for every color $\mathfrak{c} \in [q]$, every vertex $r \in V$, and every pair of boundary conditions $\tau, \sigma : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, we have that

$$\left| \frac{\mu_r^\tau(\mathfrak{c})}{\mu_r^\sigma(\mathfrak{c})} - 1 \right| \leq C_{\mathsf{SSM}} \cdot (1-\delta)^{\mathrm{dist}(r, \Lambda_{\tau,\sigma})}, \tag{1}$$

  where $\Lambda_{\tau,\sigma} = \{v \in \Lambda : \tau(v) \neq \sigma(v)\}$. (This differs slightly from the notion of strong spatial mixing we defined previously, but only by constant factor losses in C_{SSM}.)

- **Total Influence Decay:** There exist constants $0 < \delta < 1$ and $C_{\mathsf{TID}} > 0$ such that for every vertex $r \in V$, every boundary condition $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, and every radius $R > 0$, we have that

$$\sum_{\substack{v \in V \setminus \Lambda \\ \mathrm{dist}(r,v) = R}} \left\| \mu_v^{\tau, r \leftarrow \mathfrak{b}} - \mu_v^{\tau, r \leftarrow \mathfrak{c}} \right\|_{\mathsf{TV}} \leq C_{\mathsf{TID}} \cdot (1-\delta)^R. \tag{2}$$

In the setting of trees, both of these properties are directly implied by contraction of the tree recursion w.r.t. some potential function, even if there are q-spins with $q > 2$. This contraction was established for colorings in [Che+23].

Our main goal is to translate strong spatial mixing and total influence decay *on trees* into influence bounds for graphs of sufficiently large girth. The vague intuition here is that because the graph has large girth, the ball around every vertex is a tree, for which we will try to apply spatial mixing and total influence decay for trees. It is not at all clear that this should be possible, especially since we are only assuming the girth is some (possibly large) constant, which notably does not scale as the size of the graph increases. We will use a disagreement percolation argument. Throughout, we write $B(r, R) \stackrel{\mathsf{def}}{=} \{v : \mathrm{dist}(r, v) < R\}$ for the (open) radius-R ball around r; similarly, we write $S(r, R) \stackrel{\mathsf{def}}{=} \{v : \mathrm{dist}(r, v) = R\}$ for the associated sphere.

2.1 A First Attempt: One-Shot Coupling the Boundary of a Ball

Let us make a first attempt at bounding the influence of a single vertex in a graph of large girth.

Lemma 2.3. *Suppose Strong Spatial Mixing (see Eq. (1)) holds for the uniform distribution over proper q-colorings of any* tree *of maximum degree Δ. Let $G = (V, E)$ be a graph of maximum degree Δ and girth $g > 0$. Then for every vertex $r \in V$, $\mathfrak{b}, \mathfrak{c} \in [q]$, every radius $R < g/2$, and every boundary condition $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, we have*

$$\left\| \mu_{S(r,R) \setminus \Lambda}^{\tau, r \leftarrow \mathfrak{b}} - \mu_{S(r,R) \setminus \Lambda}^{\tau, r \leftarrow \mathfrak{c}} \right\|_{\mathsf{TV}} \lesssim C_{\mathsf{SSM}} (1-\delta)^R.$$

Note that the threshold $g/2$ is chosen precisely so that the induced subgraph on the ball $B(r, g/2)$ is a tree. A proof of Lemma 2.3 is provided in Appendix A. The trick is to apply Bayes' Theorem to convert the left-hand side into the influence of $S(r, R) \setminus \Lambda$ onto r. Since this influence is defined by pinning all vertices in $S(r, R) \setminus \Lambda$, and since the induced subgraph on $B(r, R)$ is a tree, we can directly apply strong spatial mixing on trees.

In light of Lemma 2.3, a natural attempt at coupling $\mu^{\tau, r \leftarrow \mathfrak{b}}, \mu^{\tau, r \leftarrow \mathfrak{c}}$ is to first sample $\sigma, \sigma' : S(r, R) \setminus \Lambda \to [q]$ from an optimal coupling of the conditional marginals $\mu_{S(r,R) \setminus \Lambda}^{\tau, r \leftarrow \mathfrak{b}}, \mu_{S(r,R) \setminus \Lambda}^{\tau, r \leftarrow \mathfrak{c}}$, and then optimally couple $\mu^{\tau,\sigma}$ and $\mu^{\tau,\sigma'}$.

- **Good Case:** If $\sigma = \sigma'$, then all disagreements will be confined to $B(r, R)$. Indeed, $\mu_{V \setminus B(r,R)}^{\tau,\sigma} = \mu_{V \setminus B(r,R)}^{\tau,\sigma'}$, and so we can couple all vertices outside $B(r, R)$ perfectly. This would yield a Hamming distance of at most $\Delta^R \leq \Delta^{g/2} \leq O(1)$. By Lemma 2.3, this happens with good (but constant) probability, since g is large (but constant).

- **Bad Case:** Otherwise $\sigma \neq \sigma'$. We could try to apply Triangle Inequality to Wasserstein distance, flipping colors from σ to σ' one vertex at a time, and then recursively coupling the new instances. But Lemma 2.3 unfortunately gives us no control on the number of disagreeing vertices in $S(r, R) \setminus \Lambda$; there could be as many as Δ^R, in which case the number of disagreements could propagate uncontrollably. This happens with small (but constant) probability, so we're doomed in this case.

2.2 A More Refined Step-by-Step Coupling

Lemma 2.4. *Suppose Strong Spatial Mixing (see Eq. (1)) and Total Influence Decay (see Eq. (2)) both hold for the uniform distribution over proper q-colorings of any* tree *of maximum degree Δ. Let $G = (V, E)$ be a graph of maximum degree Δ and girth $g > 0$. Then for every vertex $r \in V$, $\mathfrak{b}, \mathfrak{c} \in [q]$, every radius $R < g/2$, and every boundary condition $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, we have the influence decay bound*

$$\sum_{v \in S(r,R) \setminus \Lambda} \left\| \mu_v^{\tau, r \leftarrow \mathfrak{b}} - \mu_v^{\tau, r \leftarrow \mathfrak{c}} \right\|_{\mathsf{TV}} \lesssim C_{\mathsf{TID}}(1-\delta)^R + C_{\mathsf{SSM}}(1-\delta)^{g/2}\Delta^R. \tag{3}$$

The first term $C_{\mathsf{TID}}(1-\delta)^R$ is the bound we would get if G was literally a tree, or if we were allowed to truncate to $B(r, R)$, since this is just total influence decay for trees. We can implement this truncation to radius-R by pinning vertices in the boundary $S(r, g/2) \setminus \Lambda$. While this is not a true truncation, vertices within radius $R \ll g/2$ intuitively cannot tell the difference. This can be formalized by taking advantage of strong spatial mixing on trees, giving an additional error term of $2C_{\mathsf{SSM}}(1-\delta)^{g/2}\Delta^R$. Note that the Δ^R factor tells us that vertices close enough to the boundary $S(r, g/2)$ definitely do feel the difference between a genuine truncation and one achieved by pinning vertices in $S(r, g/2)$. Therefore, we'll eventually take R to be much smaller than $g/2$. The precise proof of Lemma 2.4 is implemented in Appendix A, using Lemma 2.3 as a building block.

Lemma 2.4 gives us "smoother"/"more fine-grained" control compared to Lemma 2.3. The latter only tells us that no discrepancies appear in $S(r, R)$ with good (but still constant) probability; however, if any discrepancies do appear, we have no control on how many. In contrast, Lemma 2.4 already says that if we pick a uniformly random vertex v and couple the conditional marginals optimally, then we get a discrepancy with probability $\leq \frac{\text{Small Constant}}{|S(r,R)|}$; since this probability is inversely proportional to the size of the boundary, at a high level, one can then imagine iterating this for the remaining vertices in $S(r, R)$, going in a random order. We will use bounds of the form in Eq. (3) to build a coupling certifying $O(1)$-coupling independence. This is formalized as follows.

Theorem 2.5 ([Che+23])**.** *Suppose there exists a radius $0 < R < g/2$ and an $\epsilon \lesssim \frac{1}{R \log \Delta}$ such that the following holds: For every graph $G = (V, E)$ of maximum degree Δ, every vertex $r \in V$, $\mathfrak{b}, \mathfrak{c} \in [q]$, and every boundary condition $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, we have the influence bound*

$$\sum_{v \in S(r,R) \setminus \Lambda} \left\| \mu_v^{\tau, r \leftarrow \mathfrak{b}} - \mu_v^{\tau, r \leftarrow \mathfrak{c}} \right\|_{\mathsf{TV}} \leq \epsilon. \tag{4}$$

Then for every $r \in V$, $\mathfrak{b}, \mathfrak{c} \in [q]$, and $\tau : \Lambda \to [q]$,

$$\mathscr{W}_1 \left(\mu^{\tau, r \leftarrow \mathfrak{b}}, \mu^{\tau, r \leftarrow \mathfrak{c}} \right) \leq O\left(\Delta^R\right),$$

and μ^τ is $O(\Delta^R)$-spectrally independent.

Remark 2. Note that this theorem itself does *not* require a girth lower bound, nor correlation decay. However, establishing Eq. (4) is highly nontrivial. Furthermore, it is important that we assume Eq. (4) holds for all pinnings τ. These can be achieved with through correlation decay and a girth assumption, and then applying Lemma 2.4 with R and g sufficiently large (but constant). Finally, we emphasize again that the results presented in this section are not specific to colorings; they work for essentially any q-spin system.

Proof of Theorem 2.5. Given an arbitrary *instance* of the problem, i.e. a graph $G = (V, E)$ of maximum degree Δ, a vertex $r \in V$, a boundary condition $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V \setminus \{r\}$, and a pair of distinct colors $\mathfrak{b}, \mathfrak{c} \in [q]$, we give a general recipe for coupling $\mu^{\tau, r \leftarrow \mathfrak{b}}, \mu^{\tau, r \leftarrow \mathfrak{c}}$:

- **"Base" Case:** Suppose there is no vertex in $S(r, R) \setminus \Lambda$. Then we arbitrarily couple all unpinned vertices within $B(r, R)$, and perfectly couple all unpinned vertices outside of $B(r, R)$. The latter is possible just by conditional independence.
- **"Recursive" Case:** Select a uniformly random vertex $v \in S(r, R) \setminus \Lambda$. Let $(\mathfrak{b}', \mathfrak{c}')$ be drawn from a TV-optimal coupling between $\mu_v^{\tau, r \leftarrow \mathfrak{b}}, \mu_v^{\tau, r \leftarrow \mathfrak{c}}$, and use a $\mathscr{W}_1$-optimal coupling between $\mu_v^{\tau, r \leftarrow \mathfrak{b}, v \leftarrow \mathfrak{b}'}$ and $\mu_v^{\tau, r \leftarrow \mathfrak{c}, v \leftarrow \mathfrak{c}'}$.

Of course, one cannot algorithmically implement this coupling efficiently, but this is not an issue since again, we're just using this as a proof technique. While we don't have access to the $\mathscr{W}_1$-optimal coupling between $\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}$ in the second step, we can (recursively) produce some coupling of $\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}$ which gives us upper bounds: We can just use the exact procedure we just defined to couple $\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}$ and separately couple $\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}$. Composing these two couplings then gives a coupling between $\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}$, and we can use the Triangle Inequality.

More formally, for a fixed vertex v, let ξ_v denote the TV-optimal coupling between $\mu_v^{\tau,r\leftarrow\mathfrak{b}}, \mu_v^{\tau,r\leftarrow\mathfrak{c}}$ from the second step, and write $(\mathfrak{b}', \mathfrak{c}')$ for the random pair of colors drawn from this coupling. Then

$$\mathscr{W}_1\left(\mu^{\tau,r\leftarrow\mathfrak{b}}, \mu^{\tau,r\leftarrow\mathfrak{c}}\right) \le \mathbb{E}_{v\in S(r,R)\setminus\Lambda}\Bigg[\Pr_{\xi_v}[\mathfrak{b}'=\mathfrak{c}'] \cdot \underbrace{\mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}\right)}_{(A)} + \Pr_{\xi_v}[\mathfrak{b}'\neq\mathfrak{c}'] \cdot \underbrace{\mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}\right)}_{(B)}\Bigg].$$

For (A), we know that $\mathfrak{b}'=\mathfrak{c}'$, and so the new pinnings $v\leftarrow\mathfrak{b}', v\leftarrow\mathfrak{c}'$ agree and can be absorbed into τ, yielding a new boundary condition and decreasing the number of unpinned vertices. This is the good case. For (B), we have by the Triangle Inequality that

$$(B) \le \mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}\right) + \mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}\right) + 1.$$

In the first term, we have removed the discrepancy at r, and so we can absorb the pinning $r\leftarrow\mathfrak{b}$ into τ. In the second term, we have removed the discrepancy at v, and so we can absorb the pinning $v\leftarrow\mathfrak{c}'$ into τ, similar to what we did for (A). All in all, we have the reduced the number of unpinned vertices by 1 for each of the instances appearing in the right-hand side. Finally, using the fact that ξ_v is a TV-optimal coupling, we have $\Pr_{\xi_v}[\mathfrak{b}'\neq\mathfrak{c}'] = \left\|\mu_v^{\tau,r\leftarrow\mathfrak{b}} - \mu_v^{\tau,r\leftarrow\mathfrak{c}}\right\|_{\mathsf{TV}}$. Putting all of this together, we obtain

$$\mathscr{W}_1\left(\mu^{\tau,r\leftarrow\mathfrak{b}}, \mu^{\tau,r\leftarrow\mathfrak{c}}\right) \le \mathbb{E}_{v\in S(r,R)\setminus\Lambda}\Bigg[\max_{\mathfrak{c}'}\mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}, \mu_v^{\tau,r\leftarrow\mathfrak{c},v\leftarrow\mathfrak{c}'}\right) + \left\|\mu_v^{\tau,r\leftarrow\mathfrak{b}} - \mu_v^{\tau,r\leftarrow\mathfrak{c}}\right\|_{\mathsf{TV}} \cdot \left(1 + \max_{\mathfrak{b}',\mathfrak{c}'}\mathscr{W}_1\left(\mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{b}'}, \mu_v^{\tau,r\leftarrow\mathfrak{b},v\leftarrow\mathfrak{c}'}\right)\right)\Bigg]. \tag{5}$$

Again, the first term is good for us. To control the second term, we will use our assumed influence bound Eq. (4) to control the expected number of new discrepancy vertices v appearing in $S(r,R)$ from our coupling.

For two nonnegative integers $k, \ell\in\mathbb{N}$ satisfying $\ell\le k$, define the "worst discrepancy"

$$\mathsf{WD}(k,\ell) \stackrel{\mathsf{def}}{=} \max \mathscr{W}_1\left(\mu^{\tau,r\leftarrow\mathfrak{b}}, \mu^{\tau,r\leftarrow\mathfrak{c}}\right),$$

where the maximum is taken over all instances $(G, r, \tau:\Lambda\to[q], \mathfrak{b}, \mathfrak{c})$ such that the number of unpinned vertices $|V\setminus\Lambda\setminus\{r\}|$ is exactly k, and the number of unpinned vertices in the sphere $|S(r,R)\setminus\Lambda|$ is exactly ℓ. We have that

$$\mathsf{WD}(k,\ell) \le \mathsf{WD}(k-1,\ell-1) + \max_{\ell'\le\Delta^R}\{1+\mathsf{WD}(k,\ell')\}\cdot\mathbb{E}_{v\in S(r,R)\setminus\Lambda}\left[\left\|\mu_v^{\tau,r\leftarrow\mathfrak{b}} - \mu_v^{\tau,r\leftarrow\mathfrak{c}}\right\|_{\mathsf{TV}}\right] \qquad \text{(Eq. (5))}$$

$$\le \mathsf{WD}(k-1,\ell-1) + \frac{\epsilon}{\ell}\cdot\max_{\ell'\le\Delta^R}\{1+\mathsf{WD}(k,\ell')\}, \qquad \text{(Eq. (4))}$$

with the base case $\mathsf{WD}(k,0)=\Delta^R$ for all $k\in\mathbb{N}$. To finish the proof, we can solve this recursion to obtain that

$$\mathsf{WD}(k,\ell) \le (1+2\epsilon H(\ell))\cdot\Delta^R, \tag{6}$$

where $H(\ell)\stackrel{\mathsf{def}}{=}\sum_{j=1}^{\ell}\frac{1}{j}\le 1+\log\ell$ denotes the ℓth harmonic number. The calculations for verifying Eq. (6) are available in [Che+23]. Since the largest ℓ can be is Δ^R and we assumed $\epsilon\lesssim\frac{1}{R\log\Delta}$, we are done. □

References

[CG23] Zongchen Chen and Yuzhou Gu. "Fast Sampling of b-Matchings and b-Edge Covers". In: *arXiv preprint arXiv:2304.14289* (2023) (cit. on p. 3).

[Che+23] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. "Strong spatial mixing for colorings on trees and its algorithmic applications". In: *arXiv preprint arXiv:2304.01954* (2023). To Appear in FOCS 2022 (cit. on pp. 3–6).

[CZ23] Xiaoyu Chen and Xinyuan Zhang. "A Near-Linear Time Sampler for the Ising Model with External Field". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2023, pp. 4478–4503. DOI: 10.1137/1.9781611977554.ch170. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch170 (cit. on p. 3).

[GK12] David Gamarnik and Dmitriy Katz. "Correlation decay and deterministic FPTAS for counting colorings of a graph". In: *Journal of Discrete Algorithms* 12 (2012), pp. 29–47 (cit. on p. 3).

[GKM15] David Gamarnik, Dmitriy Katz, and Sidhant Misra. "Strong Spatial Mixing of List Coloring of Graphs". In: *Random Struct. Algorithms* 46.4 (July 2015), pp. 599–613. ISSN: 1042-9832. DOI: 10.1002/rsa.20518 (cit. on p. 3).

[LY13] Pinyan Lu and Yitong Yin. "Improved FPTAS for Multi-spin Systems". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 639–654 (cit. on p. 3).

[Wei06] Dror Weitz. "Counting Independent Sets Up to the Tree Threshold". In: *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*. Seattle, WA, USA, 2006, pp. 140–149 (cit. on p. 3).

A Unfinished Proofs

Proof of Lemma 2.3. To ease notation, let $\nu = \mu^{\tau}_{S(r,R)\setminus\Lambda}$. We have

$$\begin{aligned}
\left\|\nu^{r\leftarrow\mathfrak{b}} - \nu^{r\leftarrow\mathfrak{c}}\right\|_{\mathsf{TV}} &= \frac{1}{2}\sum_{\sigma}\nu^{r\leftarrow\mathfrak{c}}(\sigma)\cdot\left|\frac{\nu^{r\leftarrow\mathfrak{b}}(\sigma)}{\nu^{r\leftarrow\mathfrak{c}}(\sigma)} - 1\right| \\
&\le \frac{1}{2}\max_{\sigma}\left|\frac{\nu(\sigma \mid r\leftarrow\mathfrak{b})}{\nu(\sigma \mid r\leftarrow\mathfrak{c})} - 1\right| \\
&= \frac{1}{2}\max_{\sigma}\left|\frac{\nu(r\leftarrow\mathfrak{b} \mid \sigma)}{\nu(r\leftarrow\mathfrak{b})}\cdot\frac{\nu(r\leftarrow\mathfrak{c})}{\nu(r\leftarrow\mathfrak{c} \mid \sigma)} - 1\right| && \text{(Bayes' Rule)} \\
&= \frac{1}{2}\max_{\sigma}\frac{\nu(r\leftarrow\mathfrak{b} \mid \sigma)}{\nu(r\leftarrow\mathfrak{b})}\cdot\left|\frac{\nu(r\leftarrow\mathfrak{c})}{\nu(r\leftarrow\mathfrak{c} \mid \sigma)} - \frac{\nu(r\leftarrow\mathfrak{b})}{\nu(r\leftarrow\mathfrak{b} \mid \sigma)}\right| \\
&\lesssim \max_{\sigma}\max_{\mathfrak{c}}\left|\frac{\nu(r\leftarrow\mathfrak{c})}{\nu(r\leftarrow\mathfrak{c} \mid \sigma)} - 1\right| && \text{(Marginal Bounds and Triangle Inequality)} \\
&\lesssim \max_{\sigma,\sigma'}\max_{\mathfrak{c}}\left|\frac{\nu(r\leftarrow\mathfrak{c} \mid \sigma)}{\nu(r\leftarrow\mathfrak{c} \mid \sigma')} - 1\right| && \text{(Triangle Inequality)} \\
&\lesssim C_{\mathsf{SSM}}(1-\delta)^{R}. && \text{(Strong Spatial Mixing Eq. (1))}
\end{aligned}$$

□

Proof of Lemma 2.4. Let ξ be a TV-optimal coupling between $\mu^{\tau,r\leftarrow\mathfrak{b}}_{S(r,g/2)}$ and $\mu^{\tau,r\leftarrow\mathfrak{c}}_{S(r,g/2)}$; note that the total variation distance between these two distributions is small by Lemma 2.3. Write (σ,σ') for a sample from ξ. Then for every $v \in S(r,R)\setminus\Lambda$,

$$\begin{aligned}
\left\|\mu^{\tau,r\leftarrow\mathfrak{b}}_{v} - \mu^{\tau,r\leftarrow\mathfrak{c}}_{v}\right\|_{\mathsf{TV}} &\le \mathbb{E}_{(\sigma,\sigma')\sim\xi}\left[\left\|\mu^{\tau,\sigma,r\leftarrow\mathfrak{b}}_{v} - \mu^{\tau,\sigma',r\leftarrow\mathfrak{c}}_{v}\right\|_{\mathsf{TV}}\right] && \text{(Triangle Inequality)} \\
&\le \Pr_{\xi}[\sigma\neq\sigma'] + \Pr_{\xi}[\sigma=\sigma']\cdot\mathbb{E}_{\sigma}\left[\left\|\mu^{\tau,\sigma,r\leftarrow\mathfrak{b}}_{v} - \mu^{\tau,\sigma,r\leftarrow\mathfrak{c}}_{v}\right\|_{\mathsf{TV}}\right],
\end{aligned}$$

where $\mathbb{E}_\sigma$ is an expectation w.r.t. ξ conditioned on $\sigma = \sigma'$. Since ξ is TV-optimal, the first term is $\left\|\mu^{\tau,r\leftarrow\mathfrak{b}}_{S(r,g/2)} - \mu^{\tau,r\leftarrow\mathfrak{c}}_{S(r,g/2)}\right\|_{\mathsf{TV}}$, which is upper bounded by $\lesssim C_{\mathsf{SSM}}(1-\delta)^{g/2}$ using Lemma 2.3. Summing over all $v \in S(r,R) \setminus \Lambda$ and invoking linearity of expectation yields an upper bound of

$$\begin{aligned}
&C_{\mathsf{SSM}}(1-\delta)^{g/2}\Delta^R + \mathbb{E}_\sigma\left[\sum_{v\in S(r,R)\setminus\Lambda} \left\|\mu_v^{\tau,\sigma,r\leftarrow\mathfrak{b}} - \mu_v^{\tau,\sigma,r\leftarrow\mathfrak{c}}\right\|_{\mathsf{TV}}\right] \\
&\leq C_{\mathsf{SSM}}(1-\delta)^{g/2}\Delta^R + C_{\mathsf{TID}}(1-\delta)^R.
\end{aligned}$$

Here, in the final step, we used the fact that pinning all vertices in $S(r,g/2)$ using both τ and σ yields a tree (with boundary conditions prescribed by τ,σ) of depth $\leq g/2$. Hence, we may apply total influence decay on trees to deduce an upper bound of $C_{\mathsf{TID}}(1-\delta)^R$ for the second term. This concludes the proof. □