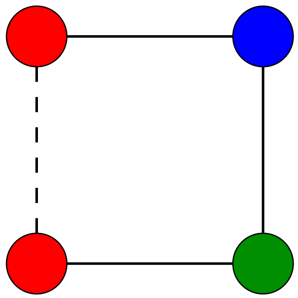
Fast and slow mixing of Markov chains for the ferromagnetic Potts model

Catherine Greenhill

School of Mathematics and Statistics
University of New South Wales

Joint work with Magnus Bordewich (Durham)
and Viresh Patel (Birmingham)

A vertex colouring of a graph $G = (V, E)$ is a map $c : V \to [q]$ such that adjacent vertices must not have the same colour. Here $q \geq 2$ is an integer and $[q] = \{1, 2, \ldots, q\}$ is a set of colours.

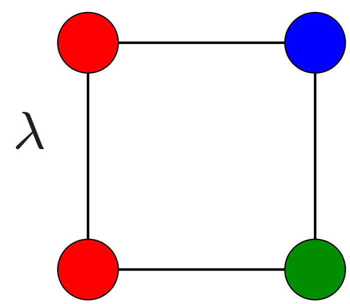


We often wish to sample such a colouring of G uniformly at random.

Instead we can allow all maps $c : V \to [q]$, but encourage adjacent vertices to have distinct colours by giving each colouring σ a weight

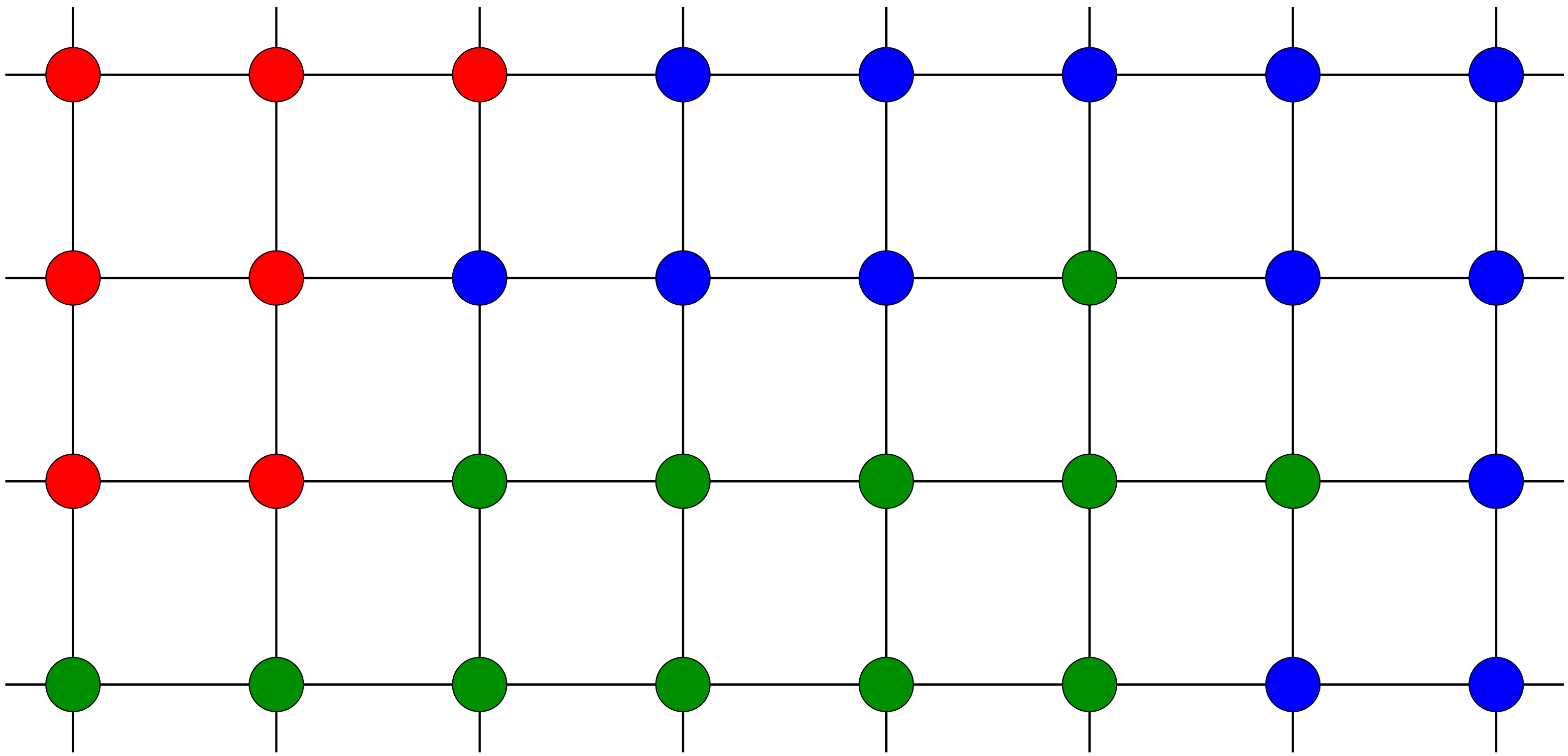
$$w(\sigma) = \lambda^{\#\ \text{mono edges in}\ \sigma},$$

where $\lambda < 1$.



This leads to the antiferromagnetic Potts model. (If $\lambda = 0$ then we recover vertex colourings.)

If instead $\lambda > 1$ then monochromatic edges are encouraged. This leads to the ferromagnetic Potts model, which arose in statistical physics as a model of magnetism.



Let $\Omega = [q]^V$ and fix the "fugacity" $\lambda > 1$.

The Gibbs distribution on Ω is the probability distribution which gives $\sigma \in \Omega$ probability which is proportional to

$$\lambda^{\mu(\sigma)},$$

where $\mu(\sigma)$ is the number of monochromatic edges of G in the colouring σ.

Then σ has probability $\lambda^{\mu(\sigma)}/Z$, where

$$Z = \sum_{\sigma \in \Omega} \lambda^{\mu(\sigma)}$$

is the partition function of the model.

Aim: to sample from Ω according to the Gibbs distribution.

However, this is computationally equivalent to computing the partition function Z exactly.

FACT: Evaluation of Z for a general graph is #P-hard.

This follows from Vertigan & Welsh (1992), since (up to an easy multiplicative constant), Z is an evaluation of the Tutte polynomial $T(G; x, y)$ along the hyperbola $(x-1)(y-1) = q$.

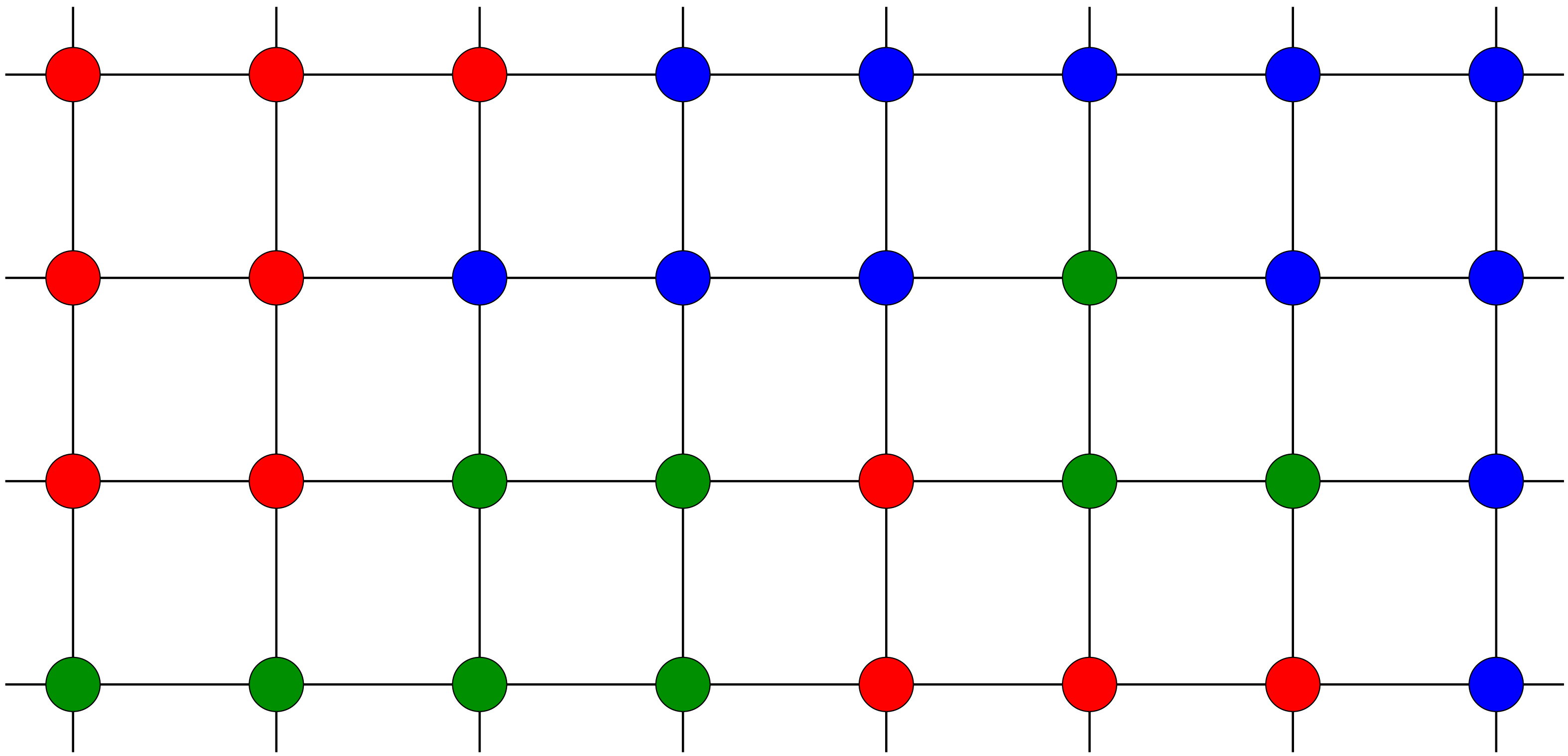
Hence the best we can hope for in polynomial time is approximate sampling. Try a Markov chain: the simplest is called the Glauber dynamics.

From current colouring $\sigma \in \Omega$ do:

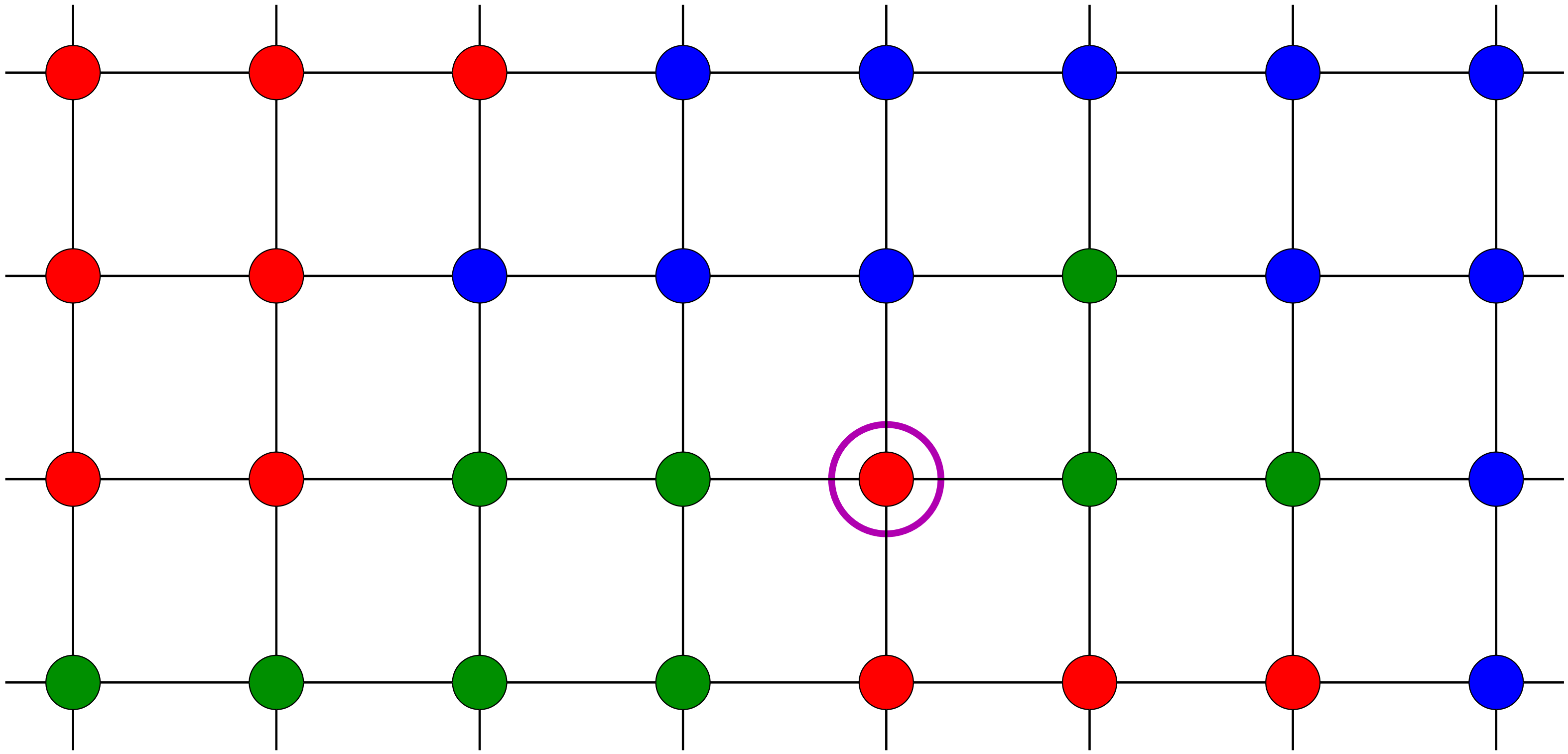
- choose a vertex $v \in V$ uniformly at random,
- choose a colour $c \in [q]$ with probability proportional to

$$\lambda^{\text{number of neighbours of } v \text{ coloured } c},$$

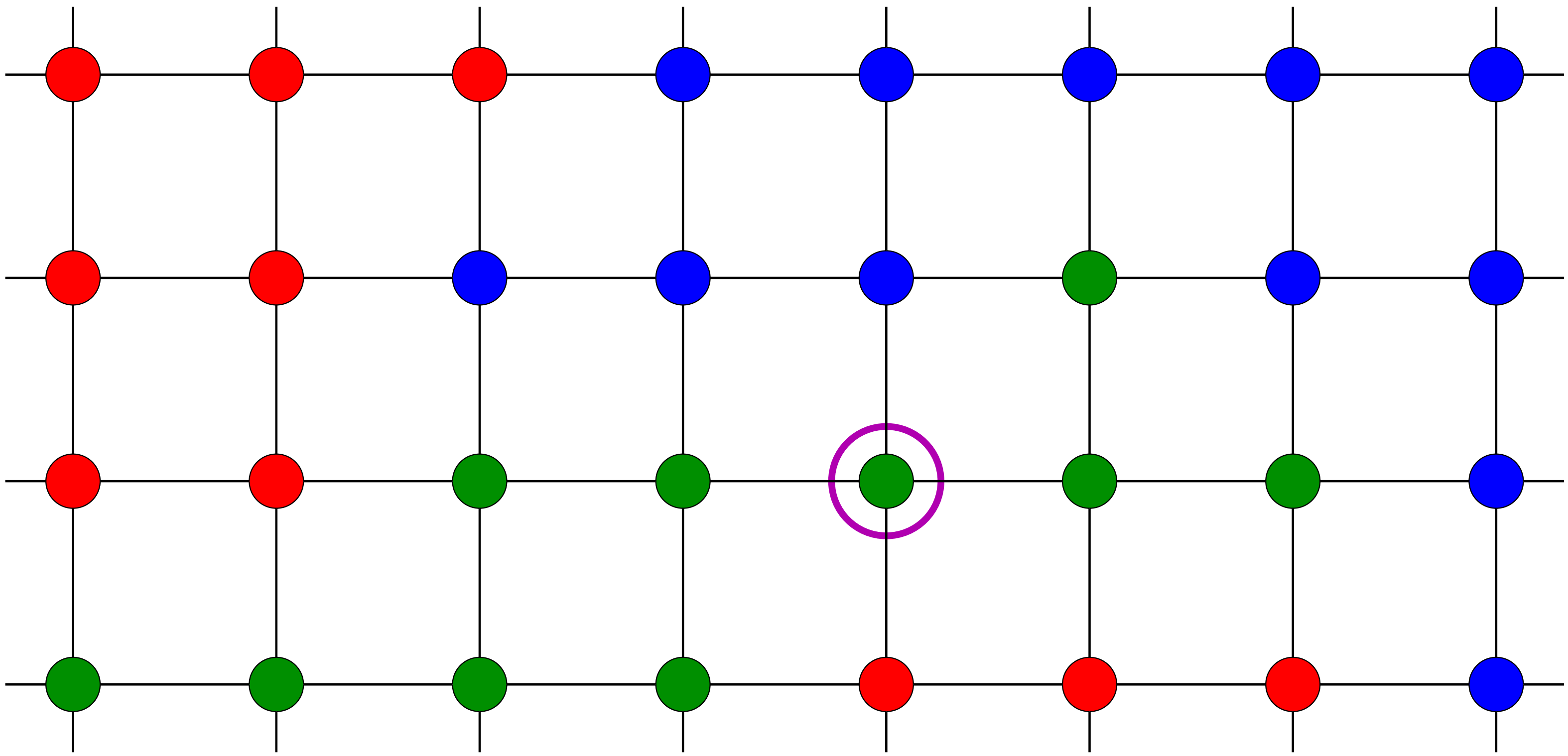
- recolour v with colour c to give the new colouring $\sigma' \in \Omega$.



Choose a vertex v uniformly at random...



Choose a vertex v uniformly at random, and choose a colour $c \in [q]$ with probability proportional to $\lambda^{\text{nr nbs of } v \text{ coloured } c}$.



Choose a vertex v uniformly at random, and choose a colour $c \in [q]$ with probability proportional to $\lambda^{\text{nr nbs of } v \text{ coloured } c}$. Recolour v with colour c.

The stationary distribution of the Glauber dynamics is the Gibbs distribution π. (Some other nice properties guarantee this.)

Start the Glauber dynamics at initial colouring $\sigma_0 \in \Omega$ and run it for t steps, visiting colourings

$$\sigma_0, \sigma_1, \cdots, \sigma_t.$$

The distance from stationarity after t steps can be measured using total variation distance:

$$d_{\mathrm{TV}}(\Pr(\sigma_t = \cdot), \pi) = \frac{1}{2} \sum_{\sigma \in \Omega} |\Pr(\sigma_t = \sigma) - \pi(\sigma)|.$$

How big must t be before this distance is at most ε, for any choice of starting colouring σ_0?

The mixing time of the Glauber dynamics is

$$\tau(\varepsilon) = \max_{\sigma_0 \in \Omega} \min \{T : d_{\mathrm{TV}}(\Pr(\sigma_T = \cdot), \pi) < \varepsilon\}.$$

We consider λ and q as fixed constants.

If $\tau(\varepsilon) \leq \mathrm{poly}(n, \log(\varepsilon^{-1}))$ then we say that the dynamics is rapidly mixing.

If $\tau(1/2e) \geq \exp(\mathrm{poly}(n))$ then we say that the dynamics is torpidly mixing.

Our results:

Theorem 1. Let Δ, $q \geq 2$ be integers and fix $\lambda > 1$ such that

$$q \geq \Delta\lambda^{\Delta} + 1.$$

Then the Glauber dynamics of the q-state Potts model at fugacity λ mixes rapidly for graphs with maximum degree Δ.

Mixing time:

$$\tau(\varepsilon) \leq (\Delta + 1)n \log(n\varepsilon^{-1})$$

(pretty fast).

Proof: Path coupling (Bubley & Dyer, 1997), which builds on Doeblin (1933), Aldous (1983).

(We now write “(q, λ)-Potts” instead of “q-state Potts model at fugacity λ”.)

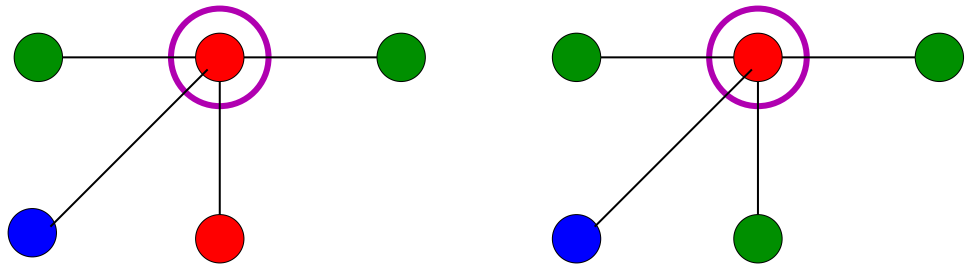
We will define a coupling (X_t, Y_t) for the Glauber dynamics:

- choose a random vertex v;

- X_t and Y_t both recolour v with colour c_X, c_Y respectively, such that c_X and c_Y both have the correct distribution but $\Pr(c_X = c_Y)$ is as large as possible.

Both (X_t) and (Y_t) are faithful copies of the Glauber dynamics.

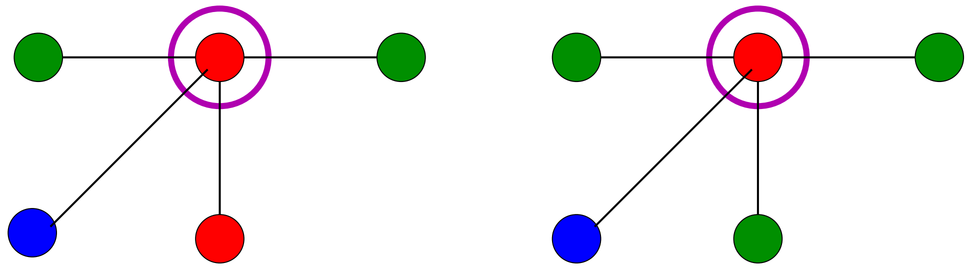
Example: suppose that $\lambda = 2$ and X and Y are as shown:



Then an optimal joint distribution of (c_X, c_Y) is given by solving an assignment problem:

	blue	green	red	
blue				$\frac{1}{4}$
green				$\frac{1}{2}$
red				$\frac{1}{4}$
	$\frac{2}{11}$	$\frac{8}{11}$	$\frac{1}{11}$	

Example: suppose that $\lambda = 2$ and X and Y are as shown:



Then an optimal joint distribution of (c_X, c_Y) is given by solving an assignment problem:

	blue	green	red	
blue	$\frac{2}{11}$	$\frac{3}{44}$	0	$\frac{1}{4}$
green	0	$\frac{1}{2}$	0	$\frac{1}{2}$
red	0	$\frac{7}{44}$	$\frac{1}{11}$	$\frac{1}{4}$
	$\frac{2}{11}$	$\frac{8}{11}$	$\frac{1}{11}$	

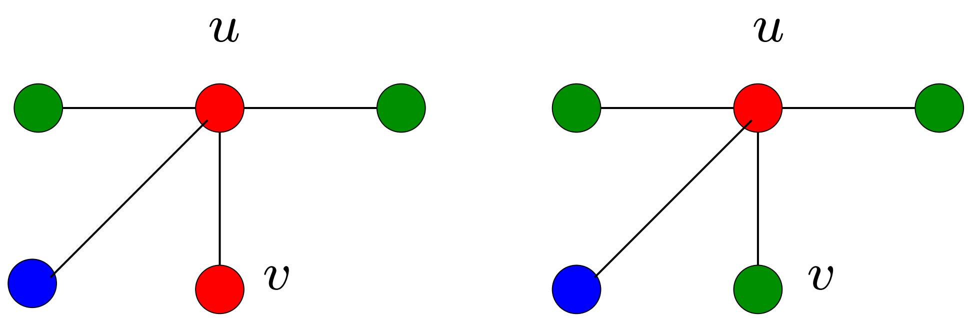
Path coupling allows us to restrict our attention to pairs (X, Y) which differ at just one vertex: that is, $H(X, Y) = 1$ where H denotes the Hamming distance.

If $(X, Y) \mapsto (X', Y')$ under the coupling and

$$\mathbf{E}(H(X', Y') | (X, Y)) \leq \beta$$

for some $\beta < 1$, then (Bubley & Dyer, 1997)

$$\tau(\varepsilon) \leq \frac{\log(n\varepsilon^{-1})}{1 - \beta}.$$



If the disagree vertex v is chosen then $H(X', Y') = 0$. If a neighbour u of v is chosen then

$$\mathbf{E}(H(X', Y')|(X, Y), v) \leq 1 + p$$

where p is the maximum probability that u receives distinct colours in X, Y.

We prove that $p \leq \lambda^{\Delta}/(\lambda^{\Delta} + q - 1)$. Then

$$\mathbf{E}(H(X', Y')|(X, Y)) \leq 1 - \frac{1}{n} + \frac{\Delta p}{n} \leq 1 - \frac{1}{(\Delta + 1)n}$$

using the assumption $q \geq \Delta\lambda^{\Delta} + 1$. $\square$

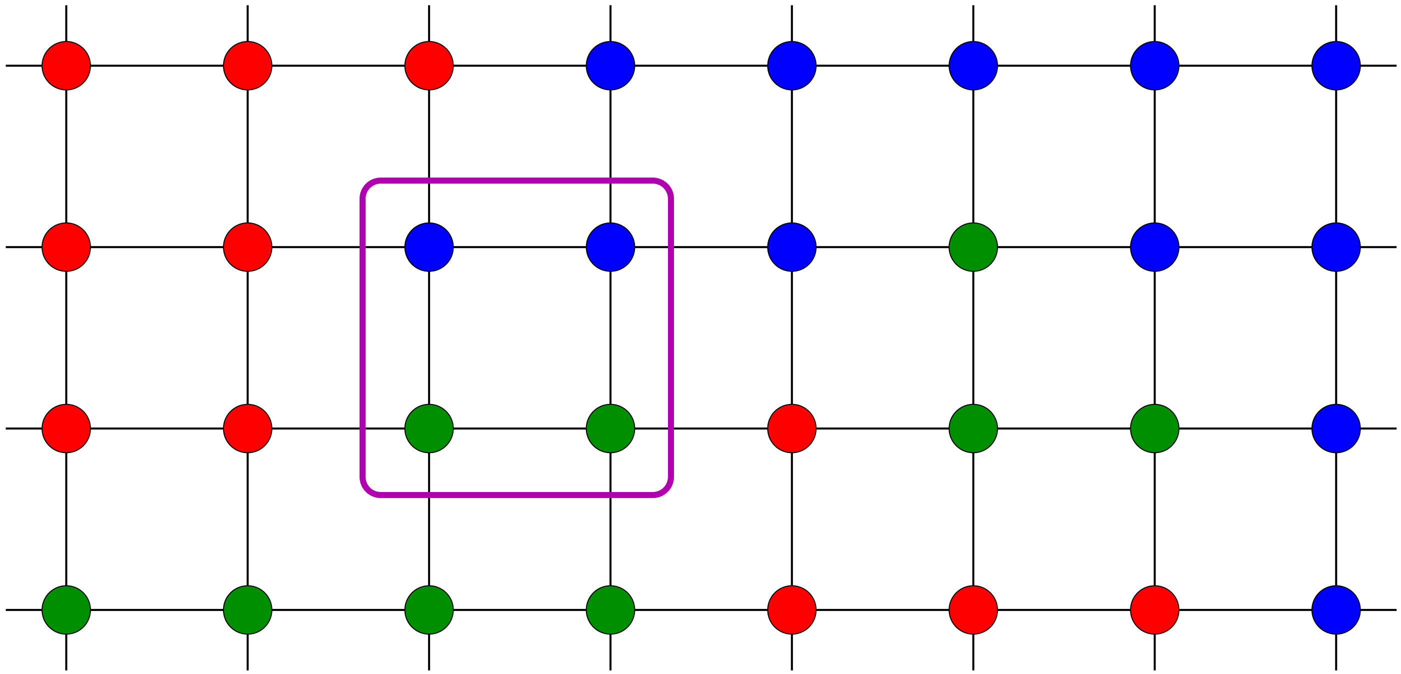
Theorem 2. Let Δ, $q \geq 2$ be integers and fix $\lambda > 1$. For any $\eta > 0$ there is a function $f(\Delta, \eta)$ such that if

$$q > f(\Delta, \eta)\, \lambda^{\Delta - 1 + \eta}$$

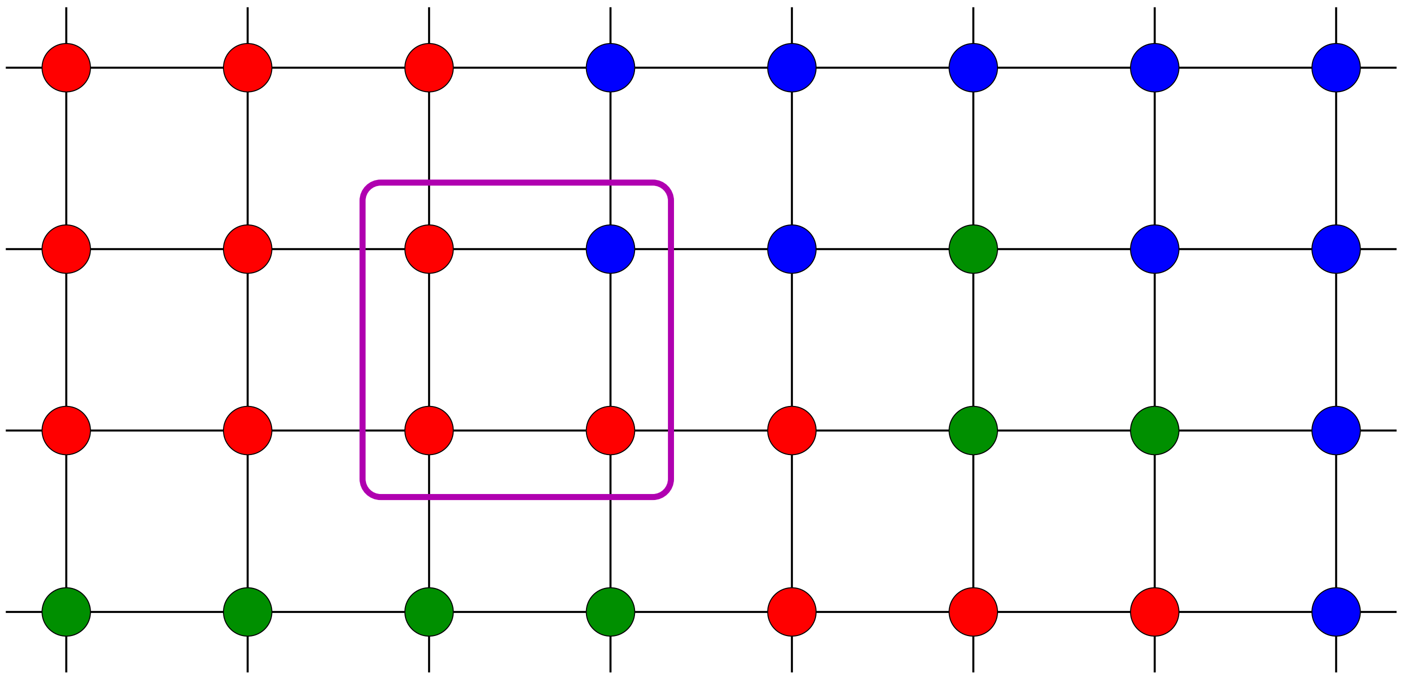
then the Glauber dynamics for (q, λ)-Potts mixes rapidly for graphs with maximum degree Δ.

This is proved by analysing a Markov chain called the block dynamics which updates more than one vertex per step.

For example, consider the set $\mathcal{S}$ of all 2×2 subgrids of the $n \times n$ toroidal grid. Choose a block $S \in \mathcal{S}$ uniformly at random and recolour ALL vertices in S at one step. The distribution on the recolouring is chosen to ensure that the stationary distribution has the Gibbs distribution.



For example, consider the set $\mathcal{S}$ of all 2×2 subgrids of the $n \times n$ toroidal grid. Choose a block $S \in \mathcal{S}$ uniformly at random and recolour ALL vertices in S at one step. The distribution on the recolouring is chosen to ensure that the stationary distribution has the Gibbs distribution.



Let v be a fixed vertex and let ψ_v be the probability that $v \in S$, where S is chosen from $\mathcal{S}$ according to some specified distribution. We prove that when $q \geq b(\mathcal{S})\,\lambda^{d(\mathcal{S})}$ (for some constants $b(\mathcal{S})$, $d(\mathcal{S})$ which we state explicitly), the mixing time of the block dynamics is at most $2\psi^{-1}\log(n\varepsilon^{-1})$, where

$$\psi = \min_{v \in V} \psi_v.$$

Then we apply a comparison theorem of Dyer, Goldberg, Jerrum & Martin (2006) to obtain an upper bound on the mixing time of the Glauber dynamics.

The mixing time we get is horrendous, but it is polynomial.

Comparison via multicommodity flows: for each transition $X \to Y$ of the block dynamics, we define a path

$$\gamma_{XY} : Z_0, Z_1, \ldots, Z_k$$

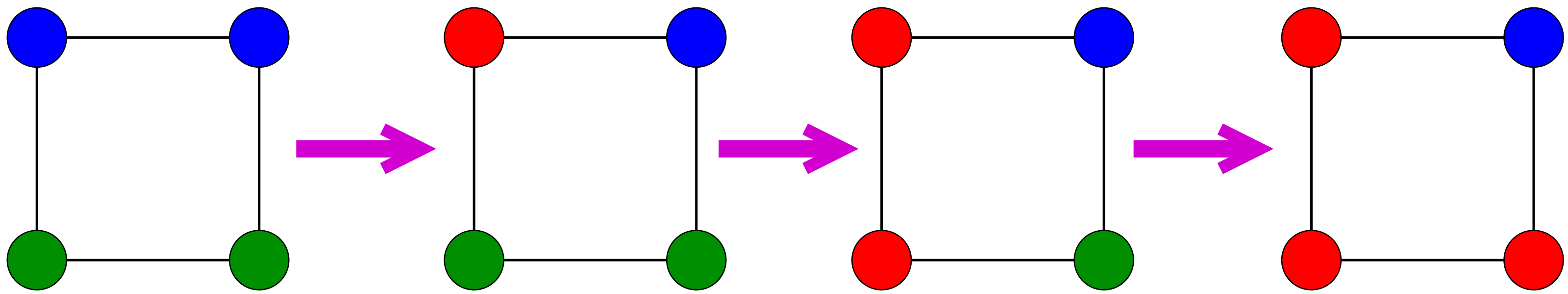
from $X = Z_0$ to $Y = Z_k$, such that $Z_j \to Z_{j+1}$ is a transition of the Glauber dynamics for $j = 0, 1, \ldots, k-1$.

If no transition $Z \to Z'$ of the Glauber dynamics is too overloaded by $\{\gamma_{XY}\}$ then the congestion A of the set of paths is small. The comparison theorem essentially says that

$$\tau_{\mathsf{Glauber}}(\varepsilon) \leq A\, \tau_{\mathsf{block}}(\varepsilon).$$

Our paths are defined by recolouring all vertices recoloured by the block transition $X \to Y$, one at a time in increasing vertex order.

Our paths are defined by recolouring all vertices recoloured by the block transition $X \to Y$, one at a time in increasing vertex order.



It turns out that the congestion A of these paths satisfies

$$A \leq sq^{s+1} \lambda^{\Delta(s+1)}$$

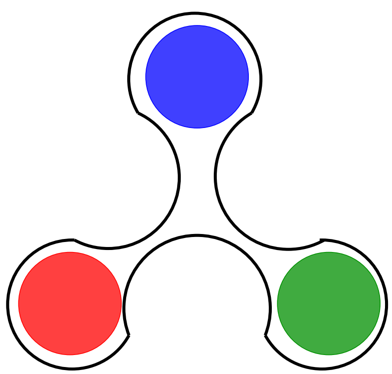
where s is the maximum block size.

Theorem 3. Let Δ, $q \geq 2$ be integers and fix $\lambda > 1$. For any $\eta > 0$ there is a function $g(\Delta, \eta)$ such that if

$$q < g(\Delta, \eta)\, \lambda^{\Delta - 1 - \frac{1}{\Delta - 1} - \eta}$$

then the Glauber dynamics for (q, λ)-Potts mixes torpidly for almost all Δ-regular graphs.

Proof: The proof uses the concept of conductance to show that there are bottlenecks in the state space.



Let σ_0 be the "all red" colouring. Define B_r to be the set of colourings which differ from σ in at most r vertices, and let S_r be those that differ in exactly r vertices (for some convenient r).

We show that for a random Δ-regular graph on n vertices, if

$$q < g(\Delta, \eta)\, \lambda^{\Delta - 1 - \frac{1}{\Delta - 1} - \eta}$$

then

$$\Pr\left(\pi(S_r)/\pi(B_r) \text{ is exponentially small}\right) \to 1 \text{ as } n \to \infty.$$

Hence it takes exponentially many steps for the chain to escape from B_r, for almost all Δ-regular graphs.

Firstly, note that

$$\pi(B_r) \geq \pi(\sigma_0) = \frac{\lambda^m}{Z}.$$

Next we bound $\pi(S_r)$. There are $\binom{n}{r}$ ways to choose the set U of r vertices not coloured red. Then for a fixed U, the contribution to $\pi(S_r)$ is

$$\lambda^{|E(\overline{U})|} Z(G[U], \lambda, q-1).$$

To bound $|E(\overline{U})|$ we perform some calculations in the configuration model, showing that with probability tending to 1 no r-set of vertices induces a subgraph with "too many" edges.

To bound $Z(G[U], \lambda, q-1)$ we proved the following:

Proposition. Let G be a graph with n vertices, m edges and maximum degree Δ. Write $m = a\Delta + b$ where $a = \lfloor m/\Delta \rfloor$ and $0 \le b < \Delta$. For any given $\lambda \ge 1$ we have

$$Z(G, \lambda, q) \le \lambda^b (1 + q^{-1}(\lambda^\Delta - 1))^a q^n.$$

Our proof involved the following probabilistic rearrangement inequality.

Lemma. Let $(X_1, \ldots, X_d)$ be a random, bounded, $\mathbb{N}^d$-valued vector. Suppose that there exists a random variable X such that $X_j \sim X$ for $j = 1, \ldots, d$. Then for all $\lambda \ge 0$ we have

$$\mathbf{E}(\lambda^{X_1 + \cdots + X_d}) \le \mathbf{E}(\lambda^{dX}).$$

The case of $\Delta = 4$ is of particular physical interest:

Proposition Let $q \geq 2$ be an integer and fix $\lambda > 1$. For any $\eta > 0$ there are functions $f(\eta)$ and $g(\eta)$ such that:

(i) if $q > f(\eta)\,\lambda^{3+\eta}$ then the Glauber dynamics for (q, λ)-Potts mixes rapidly for graphs with maximum degree 4,

(ii) if $q > f(\eta)\,\lambda^{2+\eta}$ then the Glauber dynamics for (q, λ)-Potts mixes rapidly for the toroidal grid, and

(iii) if $q < g(\eta)\,\lambda^{\frac{8}{3}-\eta}$ then the Glauber dynamics of (q, λ)-Potts mixes torpidly for almost all 4-regular graphs.

The proof of (ii) uses block dynamics where the blocks are square subgrids of the toroidal grid.

Note: The phase transition for (q, λ)-Potts model on the 2-dimensional grid occurs at $q = (\lambda - 1)^2$, so we expect rapid mixing on the grid for $q > (\lambda - 1)^2$. From (ii) we have $q > f(\eta)\, \lambda^{2+\eta}$, nearly the right power of λ.

Corollary: For sufficiently large λ, there is some number q of colours such that the Glauber dynamics for (q, λ)-Potts mixes rapidly for the toroidal grid, but mixes torpidly for almost all 4-regular graphs.