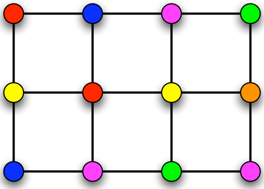
Mixing 101

Mark Jerrum
School of Mathematical Sciences
Queen Mary, University of London

Isaac Newton Institute, Cambridge, 20th March 2008

Example 1: proper colourings (antiferro Potts)

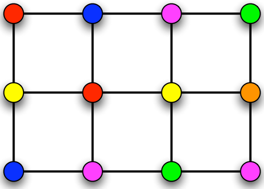
Instance: a graph $G = (V, E)$.



A (vertex) *colouring* of G is a an assignment $\sigma : V \to [q]$ of q "colours" $\{0, \dots, q-1\}$ to the vertices of G; it is *proper* if there are no monochromatic edges. Take "proper" as read.

Example 1: proper colourings (antiferro Potts)

Instance: a graph $G = (V, E)$.

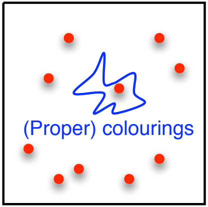


A (vertex) *colouring* of G is a an assignment $\sigma : V \to [q]$ of q "colours" $\{0, \dots, q-1\}$ to the vertices of G; it is *proper* if there are no monochromatic edges. Take "proper" as read.

Problem

Sample a colouring of G uniformly at random (u.a.r.), *efficiently* (and certainly in time polynomial in $n = |V|$).

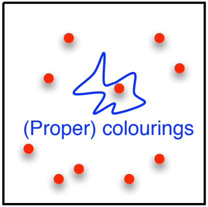
Monte Carlo (Dart throwing)



All colourings (not necessarily proper)

- Until success:
    - Choose $\sigma : V \to [q]$ u.a.r.;
    - If σ is a (proper) colouring, output σ.

Monte Carlo (Dart throwing)



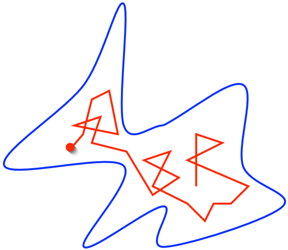
All colourings (not necessarily proper)

- Until success:
  - Choose $\sigma : V \to [q]$ u.a.r.;
  - If σ is a (proper) colouring, output σ.

Observation

Correct distribution, but exponential running time.

Markov chain Monte Carlo



- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Mixing time

The trial just described defines the transition probabilities P of a Markov chain (X_t) on state space

$$\Omega = \{\text{All (proper) } q\text{-colourings of } G\}.$$

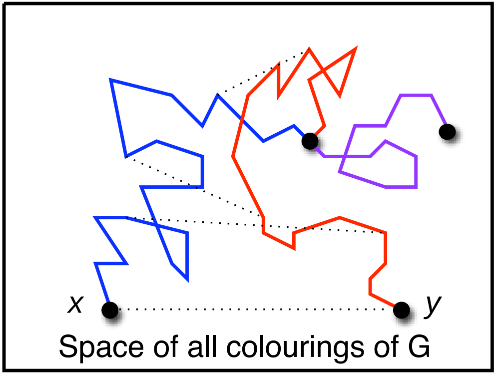
The Markov chain is irreducible and aperiodic, and its stationary distribution π is uniform.

We are interested in the *mixing time* τ of the Markov chain, i.e., the time to convergence to near stationarity:

$$\tau = \max_{x \in \Omega} \min \left\{ t : \| P^t(x, \cdot) - \pi \|_{\mathrm{TV}} \leq e^{-1} \right\},$$

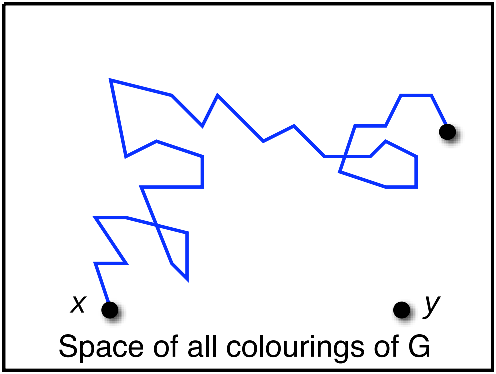
where $\|\varphi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\varphi(x)|$.

Rough guide to coupling



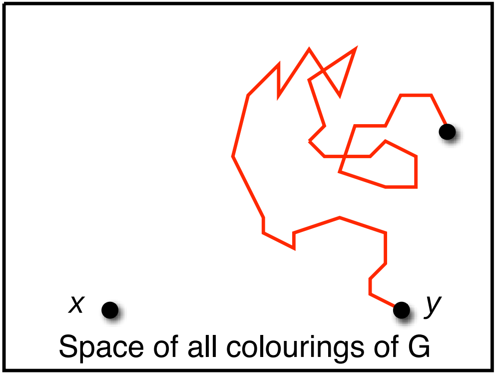
Two "coupled" evolutions of the Markov chain on the same sample space, but with different initial states.

Rough guide to coupling



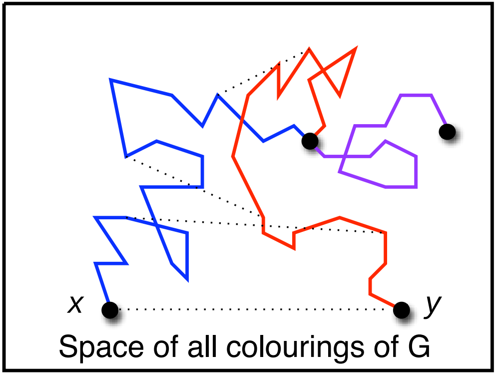
Projecting on the blue component we see a faithful copy. . .

Rough guide to coupling



Ditto projecting on red.

Rough guide to coupling



If the two can be made to coalesce rapidly, then the Markov chain must be rapidly mixing.

Basic coupling lemma

Consider a coupling $\big((X_t, Y_t) \in \Omega^2 : t \in \mathbb{N}\big)$.

Lemma

Suppose

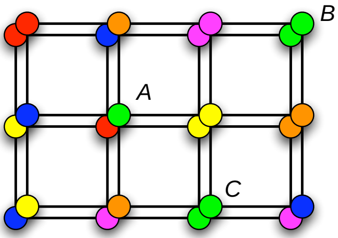
$$\Pr\big(X_t \neq Y_t \mid (X_0, Y_0) = (x_0, y_0)\big) \leq e^{-1}$$

for all choices of starting states (x_0, y_0)*. Then* $\tau \leq t$*.*

[Doeblin 1938, Aldous 1983.]

Coupling: How it might be applied to colourings

Consider a pair of colourings $(X_t, Y_t) \in \Omega^2$.



The coupling:

- Choose the same vertex v in both copies.
- Choose the same colour c in both copies.
- Attempt to recolour vertex v in both X_t and Y_t with colour c; the result is (X_{t+1}, Y_{t+1}).

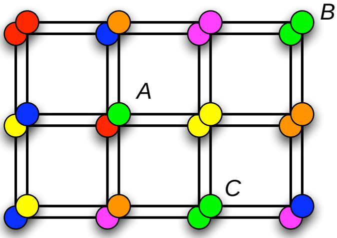
Analysis

Assume $q = 41$ colours and maximum degree 4.

Measure the progress of the coupling in terms of the Hamming distance $H(X_t, Y_t)$.

The are three basic cases, according to which vertex v is selected.

Analysis (continued)



- Type A. A vertex of disagreement. With probability at least $\frac{33}{41}$ the distance decreases by one.
- Type B. A vertex of agreement that is surrounded by other vertices of agreement. No change.
- Type C. A vertex of agreement adjacent to at least one vertex of disagreement. With probability at most $\frac{8}{41}$ the distance increases by one.

Analysis (continued)

Observe:

$$|\text{Type C}| \leq 4 \times |\text{Type A}|.$$

Therefore

$$\begin{aligned} \mathrm{E}\, H(X_{t+1}, Y_{t+1} \mid X_t, Y_t) - H(X_t, Y_t) \\ \leq \left(-\frac{33}{41} + 4 \times \frac{8}{41} \right) \frac{1}{n} H(X_t, Y_t), \end{aligned}$$

or

$$\mathrm{E}\, H(X_{t+1}, Y_{t+1} \mid X_t, Y_t) \leq \left(1 - \frac{1}{41n} \right) H(X_t, Y_t).$$

Analysis (concluded)

$$
\begin{aligned}
\mathrm{E}\, H(X_t, Y_t \mid X_0, Y_0) &\le \left(1 - \frac{1}{41n}\right)^t H(X_0, Y_0) \\
&\le \left(1 - \frac{1}{41n}\right)^t n \\
&\le e^{-1},
\end{aligned}
$$

for $t = \lceil 41n(1 + \ln n) \rceil$.

So, by the Coupling Lemma, the mixing time is $O(n \log n)$.

Poincaré inequality

. . . looks something like this:

$$\int_{\Omega} |\nabla f(x)|^2 \, dx \geq \lambda \int_{\Omega} f(x)^2 \, dx$$

for all sufficiently nice functions f with zero mean on some bounded connected domain Ω.

Poincaré inequality

... looks something like this:

$$\int_\Omega |\nabla f(x)|^2 \, dx \geq \lambda \int_\Omega f(x)^2 \, dx$$

for all sufficiently nice functions f with zero mean on some bounded connected domain Ω.

Theorem

By abstract nonsense, the Poincaré constant λ is strictly positive.

Poincaré inequality (discrete domains)

The discrete analogue for a finite Markov chain with transition probabilities P and and stationary distribution π is

$$\mathcal{E}_P(f,f) \geq \lambda \operatorname{Var}_\pi f$$

where

$$\mathcal{E}_P(f,f) = \sum_{x,y\in\Omega} \pi(x)P(x,y)(f(x)-f(y))^2$$

is the *Dirichlet form* associated with P.

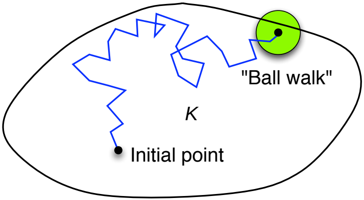
Relevance to mixing

Theorem

$$\tau = O\left(\lambda^{-1} \log \pi_{\min}^{-1}\right).$$

Example 2: sampling from a convex body

[Dyer, Frieze & Kannan, 1991], [Lovász & Simonovits, 1997].

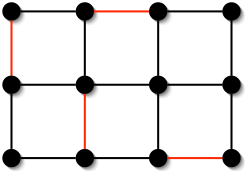


Poincaré constant is significant provided the convex body $K \subseteq \mathbb{R}^n$ is not "long and thin", yielding $\tau = \mathrm{poly}(n, \mathrm{diam}\, K)$.

Technically hard! "Abstract nonsense" is not sufficient. Also the transition from continuous to discrete causes significant headaches.

Example 3: matchings (monomer-dimer)

Instance: a graph $G = (V, E)$.



A *matching* is a collection $M \subseteq E$ of vertex-disjoint edges.

$$\pi(M) = w^{|M|}/Z, \quad \text{where } Z = \sum_M w^{|M|},$$

and $w \in \mathbb{R}^+$ is a weight or "fugacity". Task: Sample from π, efficiently (certainly in time polynomial in $n = |V|$).

Canonical paths/Multi-commodity flow

For every pair of states $x, y \in \Omega$, define a *canonical path* γ_{xy} from x to y using valid transitions of the MC.
"Congestion constant" ϱ:

$$\sum_{\gamma_{xy} \ni (z,z')} \pi(x)\pi(y)\,|\gamma_{xy}| \leq \varrho\,\pi(z)P(z,z'), \quad \forall z, z'.$$

Canonical paths/Multi-commodity flow

For every pair of states $x, y \in \Omega$, define a *canonical path* γ_{xy} from x to y using valid transitions of the MC.
"Congestion constant" ϱ:

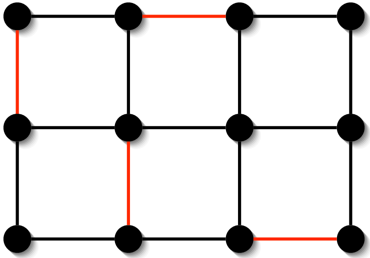
$$\sum_{\gamma_{xy} \ni (z,z')} \pi(x)\pi(y)\,|\gamma_{xy}| \leq \varrho\,\pi(z)P(z,z'), \quad \forall z, z'.$$

Theorem (Diaconis, Stroock; Sinclair)

$$\lambda \geq \varrho^{-1}, \text{ so } \tau = O(\varrho \log \pi_{\min}^{-1}).$$

The allowed transitions

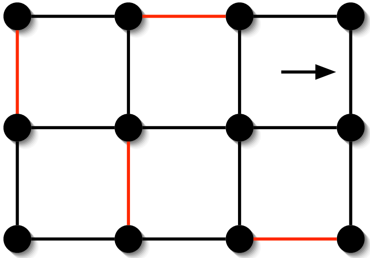
A convenient selection of transitions is: “add”, “delete” and “displace”.



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

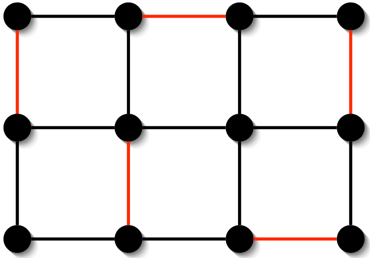
A convenient selection of transitions is: "add", "delete" and "displace".



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

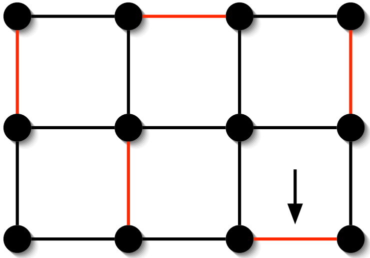
A convenient selection of transitions is: “add”, “delete” and “displace”.



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

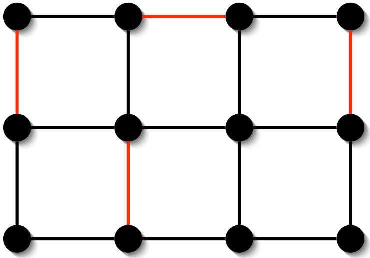
A convenient selection of transitions is: “add”, “delete” and “displace”.



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

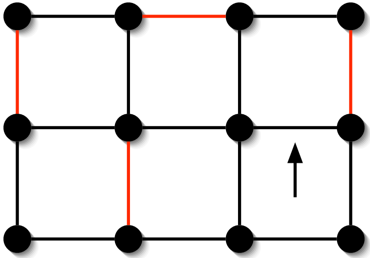
A convenient selection of transitions is: “add”, “delete” and “displace”.



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

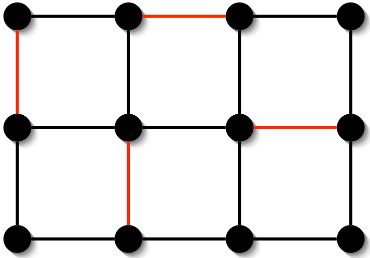
A convenient selection of transitions is: "add", "delete" and "displace".



[Broder, 1986; J. & Sinclair, 1988]

The allowed transitions

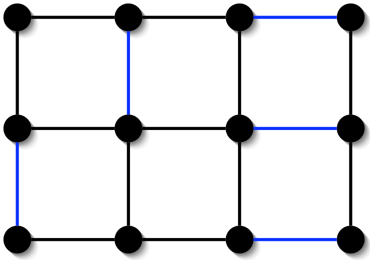
A convenient selection of transitions is: "add", "delete" and "displace".



[Broder, 1986; J. & Sinclair, 1988]

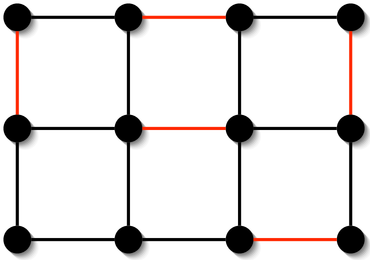
Canonical paths for matchings

To get from the blue matching. . .



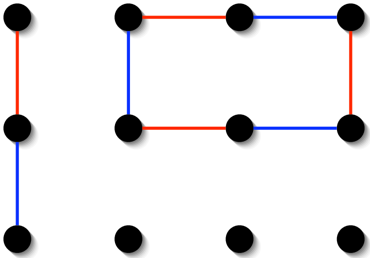
Canonical paths for matchings

. . . to the red matching. . .



Canonical paths for matchings

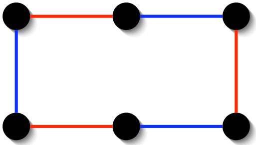
... first superimpose red and blue (symmetric difference)...



and then "unwind" each component (path or cycle).

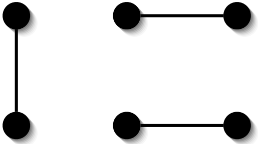
"Unwinding" a cycle

The cycle:



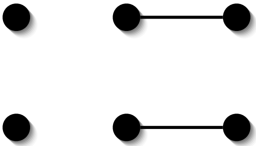
“Unwinding” a cycle

Initial matching:



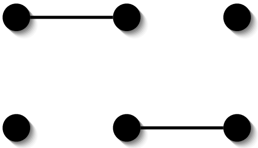
“Unwinding” a cycle

After 1 step:



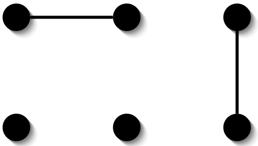
“Unwinding” a cycle

After 2 steps:



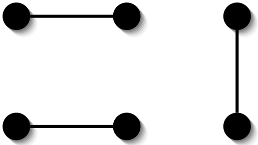
"Unwinding" a cycle

After 3 steps:



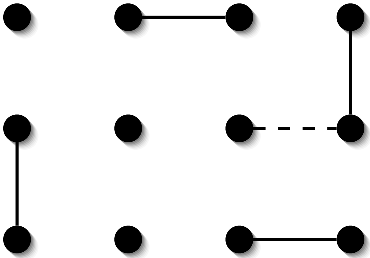
“Unwinding” a cycle

After 4 steps (final matching):



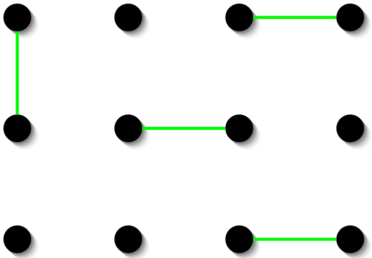
Encoding a canonical path through a transition

A transition:



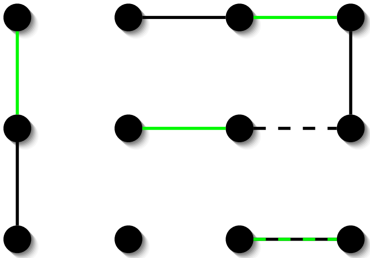
Encoding a canonical path through a transition

An encoding (matching):



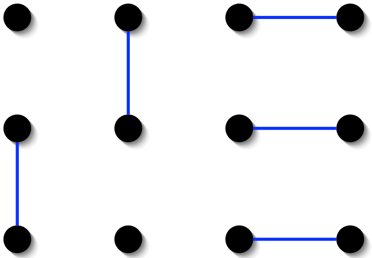
Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



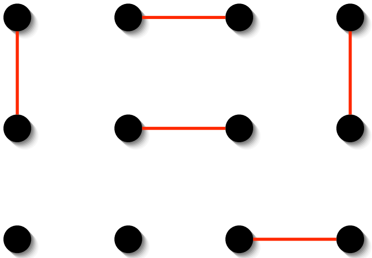
Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



Calculating the congestion

This encoding argument shows that the number of canonical paths passing through a given transition is roughly equal to the size of the state space.

Pursuing the calculation in detail yields:

Theorem (J. & Sinclair)

$\varrho = O(nm\bar{w}^2)$, *where* $n = |V|$, $m = |E|$ *and* $\bar{w} = \max\{w, 1\}$.

Corollary

$\tau = O\big(n^2 m\bar{w}^2(\log n + |\log w|)\big)$.

Using samples to estimate a partition function

Suppose G is a graph and N_k the number of matchings of cardinality k in G. Say we want to estimate $N = \sum_k N_k$, the total number of matchings of all sizes.
Define the partition function

$$Z(\lambda) = Z_G(\lambda) = \sum_k N_k \lambda^k.$$

Then our task is to compute $Z_G(1) = N$. Idea: write

$$Z(1) = Z(0) \times \frac{Z(\lambda_1)}{Z(0)} \times \frac{Z(\lambda_2)}{Z(\lambda_1)} \times \cdots \times \frac{Z(\lambda_{t-1})}{Z(\lambda_{t-2})} \times \frac{Z(1)}{Z(\lambda_{t-1})}$$

where $O < \lambda_1 < \lambda_2 < \cdots < \lambda_{t-1} < 1$, and estimate the individual ratios.

Estimating a partition function (continued)

Let $Y(M) = (\lambda_{i+1}/\lambda_i)^{|M|}$, where M is a matching chosen at random from the distribution

$$\pi(M) = \lambda_i^{|M|}/Z(\lambda_i).$$

Then

$$\begin{aligned}
\mathbb{E}_{\lambda_i}[Y] &= \frac{1}{Z(\lambda_i)} \sum_k N_k \lambda_i^k \left(\frac{\lambda_{i+1}}{\lambda_i}\right)^k \\
&= \frac{1}{Z(\lambda_i)} \sum_k N_k \lambda_{i+1}^k \\
&= \frac{Z(\lambda_{i+1})}{Z(\lambda_i)}.
\end{aligned}$$

Estimating a partition function (concluded)

Set $\lambda_{i+1} = (1 + n^{-1})\lambda_i$, where $2n$ is the number of vertices in G.
Then $1 \leq Y(M) \leq e$, and

$$\frac{\operatorname{Var} Y}{(\mathbb{E}\, Y)^2} \leq \frac{(e-1)^2}{4e}.$$

So not many samples are needed to estimate each ratio, in fact $O(n^2)$ in total.

Logarithmic Sobolev inequality

. . . looks something like this:

$$\int_\Omega |\nabla f(x)|^2 \, dx \geq \alpha \int_\Omega f(x)^2 \ln(f(x)^2) \, dx$$

where $f \geq 0$ is a sufficiently nice function satisfying $\int_\Omega f(x)^2 \, dx = 1$.

Log-Sobolev inequality (discrete domains)

The discrete analogue for a finite Markov chain with transition probabilities P and and stationary distribution π is

$$\mathcal{E}_P(f,f) \geq \alpha\,\mathcal{L}_\pi(f),$$

where

$$\mathcal{L}_\pi(f) = \mathrm{E}_\pi\left[f^2(\ln f^2 - \ln(\mathrm{E}_\pi f^2))\right]$$

and $\mathcal{E}_P(f,f)$ is the Dirichlet form associated with P, as before.

Log-Sobolev inequality (discrete domains)

The discrete analogue for a finite Markov chain with transition probabilities P and and stationary distribution π is

$$\mathcal{E}_P(f,f) \geq \alpha\,\mathcal{L}_\pi(f),$$

where

$$\mathcal{L}_\pi(f) = \mathrm{E}_\pi\left[f^2(\ln f^2 - \ln(\mathrm{E}_\pi f^2))\right]$$

and $\mathcal{E}_P(f,f)$ is the Dirichlet form associated with P, as before.

Theorem

$$\tau = O\left(\alpha^{-1}\log\log\pi_{\min}^{-1}\right).$$

Example 4: bases of a balanced matroid

... of which spanning trees in a graph is a special case.

The "bases-exchange" walk is a natural Markov chain on the bases of a matroid.

Using the inductive structure of bases with respect to the operations of deletion and contraction, one can compute a lower bound on the log-Sobolev constant in the case of *balanced* matroids.

Example 4: bases of a balanced matroid

... of which spanning trees in a graph is a special case.

The "bases-exchange" walk is a natural Markov chain on the bases of a matroid.

Using the inductive structure of bases with respect to the operations of deletion and contraction, one can compute a lower bound on the log-Sobolev constant in the case of *balanced* matroids.

Theorem (J. and Son, building on Feder and Mihail.)

The log-Sobolev constant of the bases-exchange walk for balanced matroids is $\alpha = \Omega(1/nm)$, *so the mixing time is* $\tau = O(nm \log m)$, *where* m *is the size of the ground set and* n *is the rank.*

Some other successes

- Proper colourings of a bounded degree graph, a.k.a. antiferromagnetic Potts model [Various].
- Partition function of the Ferromagnetic Ising model [J. and Sinclair].
- Linear extensions of a partial order. [Khachiyan and Karzanov], [Bubley and Dyer].
- Feasible solutions to an instance of the knapsack problem [Morris and Sinclair].
- Perfect matchings in a bipartite graph [J., Sinclair and Vigoda].

A selection of open problems

- Is there a polynomial-time algorithm for sampling perfect matchings in a *general* graph?
- Is there an algorithm for sampling perfect matchings in a bipartite graph that is efficient in practice?
- Is there a polynomial-time algorithm for sampling contingency tables?
- Can one sample proper colourings efficiently when $q \geq (1+\varepsilon)\Delta$?
- Is the bases-exchange walk rapidly mixing for all matroids?
- Develop combinatorial techniques for bounding the log-Sobolev constant.