Modern mixing, spatial and temporal

Mark Jerrum

Queen Mary, University of London

Day on Probabilities in Theoretical Computer Science
Institut Henri Poincaré, 16th September 2024

Sampling, perfect and approximate

- The computational complexity of sampling structures (and the related problem of approximately counting them) is an active area of study.
- There are two main threads to this activity: perfect sampling and approximate sampling.
- The recent major advances have been in approximate sampling based on Markov chain simulation.
- As I have not been involved in any of these, I am well placed to give an impressionistic overview.
- I'll finish off with a modest application of my own.

A running example: the hard-core gas

- Let G be a finite undirected graph with n vertices and maximum degree at most Δ.
- Denote by Ω the set of all independent sets of G.
- Let $\lambda > 0$ be a real parameter, and let the weight of independent set $I \in \Omega$ be $\mathtt{wt}(I) = \lambda^{|I|}$.
- Define the partition function by $Z = Z(G, \lambda) = \sum_{I \in \Omega} \mathtt{wt}(I)$.
- The 'hard-core' distribution assigns probability $\mathtt{wt}(I)/Z$ to independent set I.

Glauber dynamics

Consider the following dynamics (Markov chain) on independent sets of a graph G:

- Let the current state (independent set) be I.
- Select a vertex $v \in V(G)$ uniformly at random.
- If $\Gamma_G(v) \cap I = \emptyset$ then:
  - with probability $\lambda/(1+\lambda)$, set $I' \leftarrow I \cup \{v\}$;
  - with the remaining probability set $I' \leftarrow I \setminus \{v\}$;
- else: set $I' \leftarrow I$.
- The new state is I'.

The limiting distribution is the hard-core distribution.

Mixing time

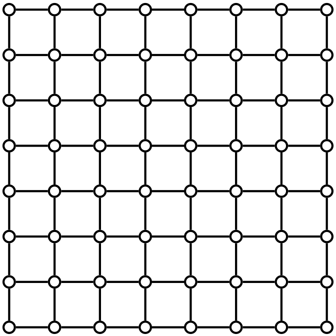
The *mixing time* is the number of steps to required to come close to the stationary distribution (in total variation distance).
The question is: how does the mixing time scale as a function of the number n of vertices in G?

- Optimal ($O(n \log n)$)?
- Polynomial ($O(n^c)$ for some c)?
- Exponential ($\Omega(\exp(cn))$ for some $c > 0$)?

As λ increases, the typical independent sets tend to increase in density, and we expect the mixing time to increase. We are interested in what happens as $\lambda \to \infty$.

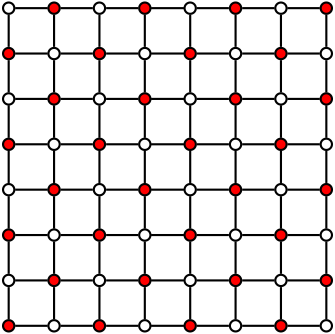
Phase transitions and mixing time

Typical behaviour is exhibited by the square lattice.



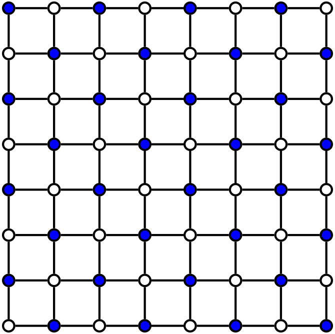
Phase transitions and mixing time

Typical behaviour is exhibited by the square lattice. For large enough λ, the system settles for long periods in states that are perturbations of this red independent set:



Phase transitions and mixing time

Typical behaviour is exhibited by the square lattice. For large enough λ, the system settles for long periods in states that are perturbations . . . or of this blue independent set:



Prehistory

Fix Δ. Let G be a graph of maximum degree Δ on n vertices.

- If λ is sufficiently small as a function of Δ, then the mixing time of Glauber dynamics is $O(n \log n)$. For example, if $\Delta = 4$ then $\lambda < 1$ suffices by a simple coupling argument [Luby and Vigoda 1999].

Prehistory

Fix Δ. Let G be a graph of maximum degree Δ on n vertices.

- If λ is sufficiently small as a function of Δ, then the mixing time of Glauber dynamics is $O(n \log n)$. For example, if $\Delta = 4$ then $\lambda < 1$ suffices by what seems now to be a simple coupling argument [Luby and Vigoda 1999].

Prehistory

Fix Δ. Let G be a graph of maximum degree Δ on n vertices.

- If λ is sufficiently small as a function of Δ, then the mixing time of Glauber dynamics is $O(n \log n)$. For example, if $\Delta = 4$ then $\lambda < 1$ suffices by a simple coupling argument [Luby and Vigoda 1999].
- If λ is sufficiently large then the mixing time of Glauber dynamics is exponential. Not only that, the problem of sampling independent sets becomes NP-hard. When $\Delta = 25$, it suffices to take $\lambda = 1$ [Dyer, Frieze and Jerrum, 2002].

History

As before, let G be a graph of maximum degree Δ on n vertices. Let $\lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$.

- If $\lambda < \lambda_\Delta$, then there is a polynomial-time algorithm for sampling from the hard-core distribution. Also there is a *deterministic* algorithm for approximating the partition function [Weitz, 2002].
- The problem of sampling independent sets is NP-hard when $\lambda > \lambda_\Delta$ [Galanis, Štefankovič and Vigoda, 2011; Sly and Sun 2012].

What is the relevance of λ_Δ? It marks the onset of a phase transition in a random regular bipartite graph of degree Δ. Alternatively, it marks the boundary of correlation decay on an infinite regular tree of degree Δ.

Back to the future: Glauber dynamics

- Although Weitz's algorithm runs in time polynomial in n, the degree of the polynomial is large, and becomes larger as λ approaches λ_Δ.
- Can we achieve the same end more simply and more efficiently by simulating Glauber dynamics?
- We hinted on the previous slide that bipartite expanders are the worst case.
- This suggests that analysing the hard core model on trees of bounded degree will be an important component in the analysis.

Arriving at the present day: the influence matrix

Let $V = \{1, 2, \ldots, n\}$, and μ be a distribution on $\{0, 1\}^V$. Think of an independent set as a function $\sigma : V \to \{0, 1\}$, and μ as being the hard-core distribution on $G = (V, E)$.

The $n \times n$ *influence matrix* is defined as

$$\Psi(i, j) = \mu[\sigma(j) = 1 \mid \sigma(i) = 1] - \mu[\sigma(j) = 1 \mid \sigma(i) = 0].$$

Let $S \subset V$ and $\tau : S \to \{0, 1\}$ be a partial configuration on S. Also let μ^τ be the conditional distribution on $\{0, 1\}^{V \setminus S}$ obtained by 'pinning' to τ on S. Then

$$\Psi^\tau(i, j) = \mu^\tau[\sigma(j) = 1 \mid \sigma(i) = 1] - \mu^\tau[\sigma(j) = 1 \mid \sigma(i) = 0].$$

Spectral independence

Definition

If, for all pinnings τ, the maximum eigenvalue of Ψ^τ is less than $1 + \eta$, we say that μ is *η-spectrally independent*.

Our hope is that the matrix entries $\Psi^\tau(i, j)$ tend rapidly to 0 as the graph distance between i and j increases. Then we should have η-spectral independence for some constant η.

Spectral independence implies optimal mixing

Theorem

Suppose that G is a graph of maximum degree Δ and that the distribution μ on $\{0,1\}^V$ is η-spectrally independent for some constant η. Subject to a local side condition (satisfied by the hard-core distribution) the mixing time of Glauber dynamics satisfies $T_{\mathrm{mix}} = O(n \log n)$, where the constant depends on η, Δ and a parameter arising from the side condition.

See Chen, Liu and Vigoda (2021); building on Anari, Liu, Oveis Gharan and Vinzant (2024), Cryan, Guo and Mousa (2021) and others.

Can show that the hard-core distribution is η-spectrally indendent when $\lambda < \lambda_\Delta$, Technology comes from Weitz (2002).

Alternative view: couplings with small Wasserstein distance

Another useful criterion for optimal mixing is the existence of a (probabilistic) coupling of two random configurations with different boundary conditions.

In the context of the hard-core distribution: Given a graph G and a vertex $v \in V(G)$ we need to show that we can couple a random red independent set R that includes v with a random blue independent set B that excludes v in such a way that R and B are close. Specifically, in expectation, R and B should disagree on only a constant number of vertices.

The existence of such a coupling readily implies spectral independence.

Example: k-matchings

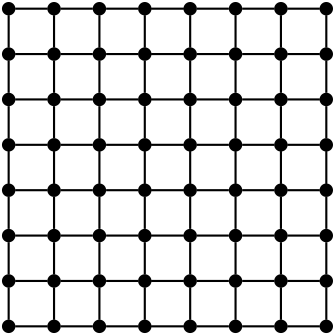
Van den Berg and Brouwer (2000) showed how to do this for matchings in graphs, which easily translates to independent sets in claw-free graphs.

Chen and Gu used this criterion very effectively in analysing Glauber dynamics for more general edge-based models. They showed, for example, that Glauber dynamics for k-matchings has optimal $O(n \log n)$ mixing. (A k-*matching* in a graph G is a subset $A \subset E(G)$ of the edges of G such that the subgraph $(V(G), A)$ has maximum degree k.)

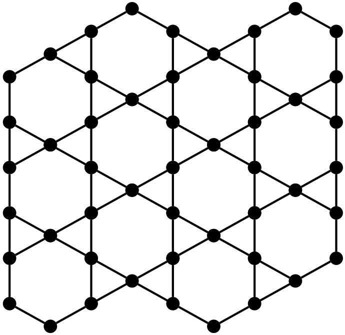
Given an edge e which is to be included in a red k-matching and excluded from a blue k-matching, we need to couple red and blue k-matchings iteratively, growing out from e. We will see the process in action in the context of the hard-core distribution.

Back to hard-core: phase transitions and mixing

Typical behaviour is exhibited by the square lattice. For large enough λ, the system settles for long periods into one of two 'phases'.



Exceptional behaviour is exhibited by the kagome lattice. There is no phase transition and optimal mixing survives at arbitrarily large λ.



Some graph theory: claw-free graphs

The *claw* is the complete bipartite graph $K_{1,3}$. A *claw-free* graph is one that does not contain the claw as an induced subgraph.

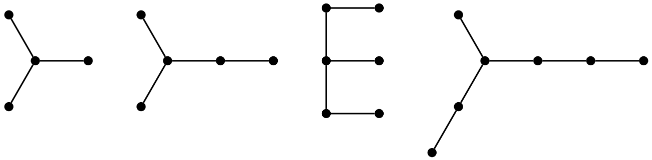
Matthews (PhD Thesis, 2008) showed that the mixing time of Glauber dynamics is polynomial (in n and λ) on claw-free graphs.

Under the additional assumption that G has maximum degree bounded by Δ, Chen and Gu (2024) showed optimal mixing $O(n \log n)$. (The implied constant depends on Δ and λ.)

The kagome lattice is claw-free.

H-free graphs of bounded degree

In the study of independent sets on hereditary graph classes, subdivided claws play a crucial role.



Subdivided claws: the claw itself, the fork, the E and the skew star.

For example, Abrishami, Chudnovsky, Dibek and Rzążewski (2022) exhibited a polynomial-time algorithm for the maximum cardinality independent set problem on bounded-degree graphs that exclude a subdivided claw (any subdivided claw).

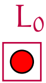
Main result

We consider classes of graphs that are H-free and have maximum degree bounded by Δ.

- If H is not a path or subdivided claw, then there is an infinite sequence of graphs on which, for sufficiently large λ, the mixing time of Glauber dynamics is exponential in n.
- If H is a path, there are only a finite number of problem instances, so this case is uninteresting.
- If H is a subdivided claw, the mixing time is $O(n \log n)$. The implicit constant is dependent on λ, Δ and H.

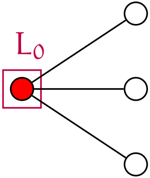
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.



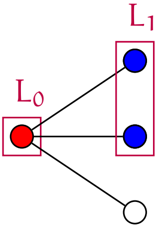
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0;



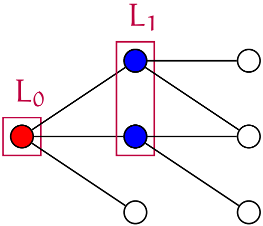
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.



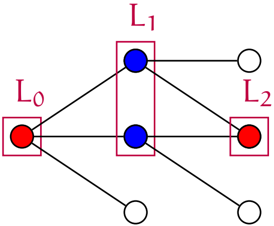
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.
- Find neighbours of L_1;



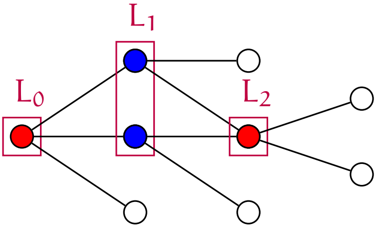
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.
- Find neighbours of L_1; reveal which are red. This is the set L_2.



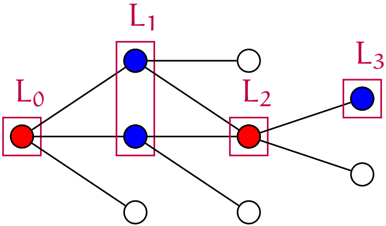
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.
- Find neighbours of L_1; reveal which are red. This is the set L_2.
- . . . and so on . . .



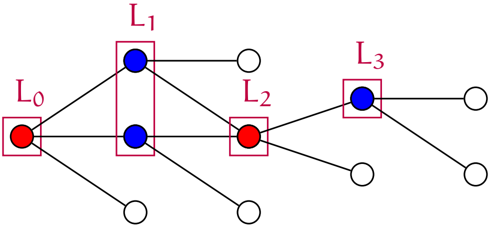
Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.
- Find neighbours of L_1; reveal which are red. This is the set L_2.
- ... and so on ...



Coupling red and blue independent sets iteratively

- Colour v red. $L_0 = \{v\}$.
- Find neighbours of L_0; reveal which are blue. This is the set L_1.
- Find neighbours of L_1; reveal which are red. This is the set L_2.
- ... and so on ...
- Terminate when some L_i is empty.



- The remainder of the red and blue independent sets may be chosen equal!

Key fact

Suppose we apply this breadth-first search procedure just described to a bounded degree graph that excludes a subdivided claw.

We can show that the 'front' has bounded size.

So on each iteration the probability of termination is bounded away from 0!

Main sources

- Daniel Štefankovič and Eric Vigoda, Lecture Notes on Spectral Independence and Bases of a Matroid: Local-to-Global and Trickle-Down from a Markov Chain Perspective. arXiv:2307.13826.
- Zongchen Chen and Yuzhou Gu, Fast Sampling of $\mathfrak{b}$-Matchings and $\mathfrak{b}$-Edge Covers. arXiv:2304.14289.
- Mark Jerrum, Glauber dynamics for the hard-core model on bounded-degree H-free graphs. arXiv:2404.07615.