A Note on the Glauber Dynamics for Sampling Independent Sets

Eric Vigoda
Division of Informatics
King's Buildings
University of Edinburgh
Edinburgh EH9 3JZ
vigoda@dcs.ed.ac.uk

Submitted: September 25, 2000; Accepted: January 17, 2001.
MR Subject Classifications: 68W20, 60J10

Abstract

This note considers the problem of sampling from the set of weighted independent sets of a graph with maximum degree Δ. For a positive fugacity λ, the weight of an independent set σ is $\lambda^{|\sigma|}$. Luby and Vigoda proved that the Glauber dynamics, which only changes the configuration at a randomly chosen vertex in each step, has mixing time $O(n \log n)$ when $\lambda < \frac{2}{\Delta-2}$ for triangle-free graphs. We extend their approach to general graphs.

1 Introduction

Given a graph $G = (V, E)$ with maximum degree Δ, we are interested in sampling from the set Ω of independent sets weighted by a positive fugacity λ. The weight of an independent set σ is $w(\sigma) = \lambda^{|\sigma|}$. Our focus is the associated probability measure

$$\mu(\sigma) = \frac{w(\sigma)}{\sum_{\sigma' \in \Omega} w(\sigma')}.$$

This corresponds to the hard-core lattice gas model from statistical physics (e.g., see [4]).

We study the *Glauber dynamics*, a very simple Markov Chain whose stationary distribution is the desired distribution. The transitions of the chain consist of selecting a vertex uniformly at random and modifying the configuration only at the chosen vertex. The goal is to analyze the time required for the dynamics to get close to its stationary distribution, known as its *mixing time* (formally defined in Section 2).

Luby and Vigoda [5] proved the Glauber dynamics has mixing time $O(n \log n)$ when $\lambda < \frac{2}{\Delta-2}$ on any triangle-free graph. This result was extended to general graphs in a preliminary version of this note [7] (the proof presented here is simpler than the original version). An alternative approach was pursued by Dyer and Greenhill who analyzed a slightly different, though also quite simple, chain (roughly speaking, they consider the Glauber dynamics with an extra slide-type move). They prove $O(n \log n)$ mixing time of their dynamics for the same range of λ for all graphs. Dyer and Greenhill [3] and Randall and Tetali [6] show that bounds on the mixing time of this modified dynamics imply bounds on the mixing time of the original Glauber dynamics. (See the remark at the end of Section 3 for a discussion regarding the case $\lambda = \frac{2}{\Delta-2}$.)

Before stating our main theorem we formally define the Markov chain for the Glauber dynamics. For the purposes of the proof, the chain is defined on the set of all subsets of V, not just independent sets. It will soon be clear that the results hold for the identical chain only defined on the set of independent sets. Given a configuration $\sigma_t \subset V$, the transitions $\sigma_t \to \sigma_{t+1}$ of the dynamics are as follows:

- Choose a vertex v uniformly at random from V.
- With probability $\frac{\lambda}{1+\lambda}$, attempt to add v into the configuration. Specifically, let $\sigma' = \sigma \cup \{v\}$. If no neighbors of v are in the configuration σ', then set $\sigma_{t+1} = \sigma'$, otherwise set $\sigma_{t+1} = \sigma_t$.
- With probability $\frac{1}{1+\lambda}$, remove v from the configuration, i.e., set $\sigma_{t+1} = \sigma_t \setminus \{v\}$.

Observe that if σ_t is an independent set, then σ_{t+1} is also an independent set. Therefore if our initial configuration is an independent set then this chain behaves as if the state space is solely the set of independent sets. Moreover, starting from an arbitrary set we eventually reach an independent set. Thus the mixing time of this chain bounds the mixing time of the chain with identical transition rules and whose state space is the set of independent sets. Redefining the state space to the set of all subsets is a crucial element in extending the original proof to graphs with triangles. The technique, first used by Bubley and Dyer [1], is often used in analyzing Markov chains for sampling proper k-colorings.

Theorem 1 *For an arbitrary graph G with maximum degree Δ, the Glauber dynamics has mixing time $O(n \log n)$ when $\lambda < \frac{2}{\Delta-2}$.*

In contrast, for $\lambda = 1$ and all $\Delta \geq 6$, Dyer, Frieze and Jerrum [2] proved there exists a bipartite graph for which the mixing time of the Glauber dynamics is exponential in the size of this graph.

2 Background

Consider a Markov chain (P, Ω, π) where P denotes the transition matrix, Ω denotes the state space, and π denotes the unique stationary distribution. Recall that the variation distance between two distributions μ, ν on Ω is defined as

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

We are interested in the mixing time T_{mix} defined as the time to get close to the stationary distribution,

$$T_{mix} = \max_{x \in \Omega} \min\{t : d_{TV}(P^t(x,\cdot), \pi(\cdot)) \leq 1/e\}.$$

We define a coupling in order to bound the mixing time. A coupling is a stochastic process (σ_t, η_t) on $\Omega \times \Omega$ such that each of the processes individually is a faithful copy of the original Markov chain, and if $\sigma_t = \eta_t$ then $\sigma_{t+1} = \eta_{t+1}$. The goal is to define a coupling so that they quickly coalesce regardless of their initial configuration. The time at which they first coalesce bounds the mixing time.

Path coupling is an important tool in designing and analyzing a useful coupling. Using path coupling it is sufficient to consider the coupling for only certain pairs of configurations. Let $S \subset \Omega \times \Omega$ where $\sigma \sim \tau$ denotes $(\sigma, \tau) \in S$. For $\sigma, \eta \in \Omega$, let $\rho(\sigma, \eta)$ denote the set of simple paths $\tau_0 = \sigma \sim \tau_1 \sim \ldots \sim \tau_k = \eta$.

Theorem 2 (Bubley and Dyer [1]) *Let Φ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$ such that, for all $\sigma, \eta \in \Omega$, there exists a path $\tau \in \rho(\sigma, \eta)$ with*

$$\Phi(\sigma, \eta) = \sum_i \Phi(\tau_i, \tau_{i+1}).$$

Suppose there exists a constant $\beta < 1$ and a coupling (σ_t, η_t) of the Markov chain such that, for all $\sigma_t \sim \eta_t$,

$$E[\Phi(\sigma_{t+1}, \eta_{t+1})] \leq \beta \Phi(\sigma_t, \eta_t).$$

Then the mixing time is bounded by

$$T_{mix} \leq \frac{\log(eD)}{1 - \beta}.$$

3 Analysis

We begin with a few definitions. Let Ω' denote the state space of the chain we analyze, i.e., Ω' is the set of all subsets of vertex set V. For a configuration $\sigma \subset \Omega'$ and vertex $v \in V$ where $v \notin \sigma$, let $\sigma_v = \sigma \cup \{v\}$. Let $\Gamma(x)$ denote the set of neighbors of vertex x, and $\Delta_x = |\Gamma(x)|$. Furthermore, let $T(v, w)$ denote the set of triangles containing both v and w, i.e., $T(v, w) = \Gamma(v) \cap \Gamma(w)$.

We say a neighbor w of v is *blocked* if $w \notin \sigma$ and $\Gamma(w) \cap \sigma \neq \emptyset$. Let $B_\sigma(v)$ denote the set of blocked neighbors of v with respect to the pair of configurations σ and σ_v. Conversely, $\overline{B}_\sigma(v) = \Gamma(v) \setminus B_\sigma(v)$. The distance for neighboring configurations is defined as

$$\Phi(\sigma, \sigma_v) = \Delta_v - c|B_\sigma(v)|, \quad \text{where } c = \frac{\lambda \Delta}{\lambda \Delta + 2}.$$

For an arbitrary pair of configurations σ, η, the distance is defined as

$$\Phi(\sigma, \eta) = \min_{\tau \in \rho(\sigma,\eta)} \sum_i \Phi(\tau_i, \tau_{i+1}).$$

It is clear that the distance function is symmetric, non-negative, zero only when the configurations are identical (since $c < 1$), and satisfies the triangle inequality. Thus Φ is a metric as is required by the path coupling theorem.

Before proving the main theorem, we identify how the re-definition of the state space to all subsets of vertex set V is used in the proof. Consider a graph which is simply a triangle on vertices v, w, x. For the pair of configurations $\sigma = \{v\}, \sigma' = \{w\}$, we need to capture the notion that x is blocked with respect to σ, σ'. More precisely, we have

$$\Phi(\sigma, \sigma') \leq \Phi(\{v\}, \{v, w\}) + \Phi(\{w\}, \{v, w\}) = 2(2 - c).$$

In contrast, suppose the state space was was simply the set of independent sets. Since $\{v, w\}$ is not an independent set, we would instead have $\Phi(\sigma, \sigma') = 4$. A similar calculation is used in equation (3) of the proof.

Proof of Theorem 1:

Throughout the proof consider a pair of neighboring configurations σ, σ_v (where $v \notin \sigma$). For $\sigma_t = \sigma, \eta_t = \sigma_v$, let $E[\Delta\Phi] = E[\Phi(\sigma_{t+1}, \eta_{t+1})|\sigma_t, \eta_t] - \Phi(\sigma_t, \eta_t)$. Let w denote a neighbor of v and x a vertex distance two from v, i.e., $x \in \Gamma(\Gamma(v)) \setminus \Gamma(v)$.

The coupling is simply the identity, i.e., each configuration attempts the same move at every step. It will be useful to consider the effect of individual moves. Towards this end, we introduce the following definition,

$$E[\Delta^{+z}\Phi] = E[\Delta\Phi| \text{ both Markov chains attempt to add } z \text{ at time } t].$$

Similarly define $E[\Delta^{-z}\Phi]$ for the move which attempts to remove z from both sets, and denote the net effect of all moves on z as follows,

$$(1 + \lambda)E[\Delta^z\Phi] = \lambda E[\Delta^{+z}\Phi] + E[\Delta^{-z}\Phi].$$

Since the distance can only change if a move occurs on v, some w, or some x, we then have that

$$n(1 + \lambda)E[\Delta\Phi] = (1 + \lambda)E[\Delta^v\Phi] + (1 + \lambda)\sum_w E[\Delta^w\Phi] + (1 + \lambda)\sum_x E[\Delta^x\Phi]. \qquad (1)$$

We can now analyze the effect of individual moves. Consider a move which:

Transitions on v:

Attempting to remove v always works in both sets, however it can only be added if $\Gamma(v) \cap \sigma = \emptyset$. If such moves are successful then both configurations are identical, thus the distance decreases by $\Phi_t = -\Delta_v + c|B_\sigma(v)|$. More formally we have the following,

$$(1 + \lambda)E[\Delta^v\Phi] = \begin{cases} (1 + \lambda)(-\Delta_v + c|B_\sigma(v)|) & \text{if } \Gamma(v) \cap \sigma = \emptyset \\ -\Delta_v + c|B_\sigma(v)| & \text{otherwise} \end{cases} \qquad (2)$$

Transitions on w, a neighbor of v:

Consider the move that attempts to add w into both configurations. Clearly this does not effect things when $w \in \sigma$. Similarly when w is blocked, the move does not work in either chain. In the remaining case of $w \notin \sigma, w \in \overline{B}_\sigma(v)$, the move succeeds in exactly one of the chains resulting in σ moving to $\sigma \cup \{w\}$ while σ_v remains unchanged. To calculate the effect of this move we begin with the following observations,

$$\begin{aligned}\Phi(\sigma \cup \{w\} \cup \{v\}, \sigma \cup \{w\}) &\leq \Phi(\sigma, \sigma \cup \{v\}) = \Phi_t \\ \Phi(\sigma \cup \{w\} \cup \{v\}, \sigma \cup \{v\}) &= \Delta_w - c|B_\sigma(w) \cup T(v,w)|. \end{aligned} \tag{3}$$

The term $T(v,w)$ is included in the last equality since every $w' \in T(v,w)$ is blocked by vertex v. We can now compute the effect of attempting to add w,

$$\begin{aligned} & E[\Delta^{+w}\Phi | w \in \overline{B}_\sigma(v), w \notin \sigma] \\ &= \Phi(\sigma \cup \{w\}, \sigma \cup \{v\}) - \Phi_t \\ &\leq \Phi(\sigma \cup \{w\}, \sigma \cup \{w\} \cup \{v\}) + \Phi(\sigma \cup \{w\} \cup \{v\}, \sigma \cup \{v\}) - \Phi_t \\ &\leq \Delta_w - c|B_\sigma(w) \cup T(v,w)|. \end{aligned} \tag{4}$$

It remains to consider the effect of removing w. This move might unblock some w' which are adjacent to both v and w,

$$E[\Delta^{-w}\Phi | w \in \sigma] = c|\{w' \in T(v,w) : \Gamma(w') \cap \sigma = \{w\}\}|. \tag{5}$$

Transitions on x, where x is distance two from v:

Since x is not a neighbor of v, the move either works in both or neither configurations. Removing x might unblock a set of neighbors of v,

$$E[\Delta^{-x}\Phi | x \in \sigma] \leq c|\{w \in \Gamma(v) : \Gamma(w) \cap \sigma = \{x\}\}|. \tag{6}$$

Meanwhile adding x will block those $w \in \Gamma(x)$ which are currently unblocked (i.e., $w \in \overline{B}_\sigma(v) \cap \Gamma(x)$). Observe that x can only be added if $\Gamma(x) \cap \sigma = \emptyset$ which implies $x \in \overline{B}_\sigma(w), w \notin \sigma$ for all $w \in \overline{B}_\sigma(v) \cap \Gamma(x)$. Thus,

$$E[\Delta^{+x}\Phi | x \notin \sigma] \leq -c|\{w \in \overline{B}_\sigma(v) \setminus \sigma : w \in \Gamma(x), x \in \overline{B}_\sigma(w)\}|. \tag{7}$$

Let $\delta()$ denote the Kronecker delta function which takes value one if the argument is true and zero otherwise. Putting inequalities (2)-(7) into equation (1) yields the following,

$$\begin{aligned} & n(1+\lambda)E[\Delta\Phi] \\ &\leq (1 + \lambda\delta(\Gamma(v) \cap \sigma = \emptyset))(-\Delta_v + c|B_\sigma(v)|) \\ &\quad + \sum_{w \in \overline{B}_\sigma(v) \setminus \sigma} \lambda(\Delta_w - c|B_\sigma(w) \cup T(v,w)|) && (8) \\ &\quad + \sum_{w \in \sigma} c|\{w' \in T(v,w) : \Gamma(w') \cap \sigma = \{w\}| && (9) \\ &\quad + \sum_{x \in \sigma} c|\{w \in \Gamma(v) : \Gamma(w) \cap \sigma = \{x\}| && (10) \\ &\quad + \sum_{x} -c\lambda|\{w \in \overline{B}_\sigma(v) \setminus \sigma : w \in \Gamma(x), x \in \overline{B}_\sigma(w)\}| && (11) \end{aligned}$$

We can consolidate (8) and (11) in the following manner,

$$\begin{aligned}
&\sum_{w\in\overline{B}_\sigma(v)\setminus\sigma} \lambda(\Delta_w - c|B_\sigma(w)\cup T(v,w)|)\\
&+\sum_x -c\lambda|\{w\in\overline{B}_\sigma(v)\setminus\sigma : w\in\Gamma(x), x\in\overline{B}_\sigma(w)\}|\\
&= \sum_{w\in\overline{B}_\sigma(v)\setminus\sigma} \lambda(\Delta_w - c|B_\sigma(w)\cup T(v,w)| - c|\{x\in\overline{B}_\sigma(w)\}|)\\
&= \sum_{w\in\overline{B}_\sigma(v)\setminus\sigma} \lambda(\Delta_w - c|B_\sigma(w)\cup T(v,w)| - c|\overline{B}_\sigma(w)\setminus\{v\}\setminus T(v,w)|)\\
&= \sum_{w\in\overline{B}_\sigma(v)\setminus\sigma} \lambda(\Delta_w(1-c) + c\delta(v\in\overline{B}_\sigma(w)))\\
&\leq |\overline{B}_\sigma(v)\setminus\sigma|\lambda(\Delta(1-c)+c\delta(\Gamma(v)\cap\sigma=\emptyset))\\
&\leq (|\overline{B}_\sigma(v)| - \delta(\Gamma(v)\cap\sigma\neq\emptyset))\lambda(\Delta(1-c)+c\delta(\Gamma(v)\cap\sigma=\emptyset)) \qquad (12)
\end{aligned}$$

We can also combine (9) and (10),

$$\begin{aligned}
&\sum_{w\in\sigma} c|\{w'\in T(v,w):\Gamma(w')\cap\sigma=\{w\}\}|\\
&\quad+\sum_{x\in\sigma} c|\{w:\Gamma(w)\cap\sigma=\{x\}\}| = c|\{w\in B_\sigma(v): |\Gamma(w)\cap\sigma|=1\}|\\
&\qquad\leq c|B_\sigma(v)|. \qquad (13)
\end{aligned}$$

Putting inequalities (12) and (13) back into (8) - (11), using the identity $c(2+\lambda) = \lambda(\Delta(1-c)+c)$, along with algebraic manipulations imply the following chain of inequalities for the case $\Gamma(v)\cap\sigma=\emptyset$,

$$\begin{aligned}
&n(1+\lambda)E[\Delta\Phi]\\
&\leq (1+\lambda)(-\Delta_v + c|B_\sigma(v)|) + |\overline{B}_\sigma(v)|\lambda(\Delta(1-c)+c) + c|B_\sigma(v)|\\
&= -\Delta_v(1+\lambda) + c(2+\lambda)|B_\sigma(v)| + |\overline{B}_\sigma(v)|\lambda(\Delta(1-c)+c)\\
&= -\Delta_v(1+\lambda) + c(2+\lambda)\Delta_v\\
&= \frac{\Delta_v}{\Delta\lambda+2}[\lambda(\Delta-2)-2].
\end{aligned}$$

For the case $\Gamma(v)\cap\sigma\neq\emptyset$ we instead use the identity $2c=\lambda\Delta(1-c)$, obtaining the following chain of inequalities,

$$\begin{aligned}
n(1+\lambda)E[\Delta\Phi] &\leq -\Delta_v + c|B_\sigma(v)| + (|\overline{B}_\sigma(v)|-1)\lambda\Delta(1-c) + c|B_\sigma(v)|\\
&= -\Delta_v + 2|B_\sigma(v)|c + (|\overline{B}_\sigma(v)|-1)\lambda\Delta(1-c)\\
&= -\Delta_v + (\Delta_v-1)\lambda\Delta(1-c)\\
&\leq \frac{\Delta_v}{\Delta\lambda+2}[\lambda(\Delta-2)-2]. \qquad (14)
\end{aligned}$$

Clearly $E[\Delta\Phi] < 0$ when $\lambda < \frac{2}{\Delta-2}$. We now want to use this bound on $E[\Delta\Phi]$ with the path coupling theorem to get a bound on the mixing time. Recall that the path coupling theorem uses a bound on $\beta = \max_{\sigma_t,\eta_t} \beta_{\sigma_t,\eta_t}$ where $\sigma_t \sim \eta_t$ and

$$E[\Phi(\sigma_{t+1}, \eta_{t+1})] = \beta_{\sigma_t,\eta_t}\Phi(\sigma_t, \eta_t).$$

We determine β_{σ_t,η_t} in terms of $E[\Delta\Phi(\sigma_t, \eta_t)] = E[\Delta\Phi]$ as follows:

$$\begin{aligned}
\beta_{\sigma_t,\eta_t}\Phi(\sigma_t, \eta_t) &= E[\Phi(\sigma_{t+1}, \eta_{t+1})] \\
(\beta_{\sigma_t,\eta_t} - 1)\Phi(\sigma_t, \eta_t) &= E[\Phi(\sigma_{t+1}, \eta_{t+1})] - \Phi(\sigma_t, \eta_t) = E[\Delta\Phi] \\
\beta_{\sigma_t,\eta_t} &= 1 + \frac{E[\Delta\Phi]}{\Phi(\sigma_t, \eta_t)}.
\end{aligned}$$

Using $\Phi(\sigma, \sigma_v) \leq \Delta_v$ and our bound on $E[\Delta\Phi]$ from inequality (14), we get a bound on β:

$$\beta \leq 1 + \frac{1}{n(1+\lambda)}\frac{\lambda(\Delta-2)-2}{(\Delta\lambda+2)}.$$

Using $\Phi' = \frac{\Phi}{c}$ which is integer-valued and plugging the bound on β into the path coupling theorem, for $\lambda < \frac{2}{\Delta-2}$ we obtain:

$$T_{mix} \leq \frac{(1+\lambda)(\Delta\lambda+2)}{2-\lambda(\Delta-2)} n \log(\frac{ne\Delta}{c}).$$

■

Remark:

The proof technique used in this paper does not immediately enable us to conclude that the Glauber dynamics has polynomial mixing time at the point $\lambda = \frac{2}{\Delta-2}$. We have $E[\Delta\Phi] \leq 0$ when $\lambda = \frac{2}{\Delta-2}$, thus it suffices to prove that for an arbitrary pair of configurations σ_t, η_t, $Pr[\Phi(\sigma_{t+1}, \eta_{t+1}) \neq \Phi(\sigma_t, \eta_t)] > 1/poly(n)$. However, this is non-trivial because of the way in which the distance function was defined for non-neighboring states. The approach taken in the original version of this note [7] did handle this case. In that version, path coupling was not used, instead a distance function was defined and analyzed directly for an arbitrary pair of states. The approach of Dyer and Greenhill, which uses path coupling with a simpler metric, does easily handle this case $\lambda = \frac{2}{\Delta-2}$.

References

[1] R. Bubley and M. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. In *38th Annual Symposium on Foundations of Computer Science*, pages 223–231, Miami Beach, FL, October 1997. IEEE.

[2] M. E. Dyer, A. M. Frieze, and M. R. Jerrum. On counting independent sets in sparse graphs. In *40th Annual Symposium on Foundations of Computer Science*, pages 210–217, New York City, NY, 1999. IEEE.

[3] M. E. Dyer and C. Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.

[4] H. O. Georgii, O. Häggström, and C. Maes. The random geometry of equilibrium phases. In C. Domb and J. Lebowitz, editors, *Phase Transitions and Critical Phenomena*. Academic Press, to appear.

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

[6] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.*, 41(3):1598–1615, 2000.

[7] E. Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets: Part II. Technical Report TR-99-003, International Computer Science Institute, January 1999.