INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

Fast Convergence of the Glauber Dynamics for Sampling Independent Sets: Part II

Eric Vigoda *

TR-99-003

January 28, 1999

Abstract

This work is a continuation of [4]. The focus is on the problem of sampling independent sets of a graph with maximum degree δ. The weight of each independent set is expressed in terms of a fixed positive parameter $\lambda \leq \frac{2}{\delta-2}$, where the weight of an indepednent set σ is $\lambda^{|\sigma|}$. The Glauber dynamics is a simple Markov chain Monte Carlo method for sampling from this distribution.

In [4], we showed fast convergence of this dynamics for triangle-free graphs. This paper proves fast convergence for arbitrary graphs.

*Computer Science Division, University of California at Berkeley, and International Computer Science Institute.
Supported in part by National Science Foundation Fellowship.

1 Introduction

For a more general introduction and a discussion of related work we refer the reader to the companion work [4]. The aim of this work is given a graph $G = (V, E)$ to efficiently sample from the probability measure μ_G defined on the set of indepedent sets $\Omega = \Omega_G$ of G weighted by a positive parameter λ. Specifically, for all independent sets $\sigma \in \Omega$,

$$\mu_G(\sigma) = \frac{\lambda^{|\sigma|}}{Z_G}$$

where $Z_G = \sum_{\sigma \in \Omega_G} \lambda^{|\sigma|}$ is the normalizing factor.

The Glauber dynamics [2] from statistical physics is a very simple Markov chain MC defined on Ω. The transition probability matrix P of MC is defined as follows.

From an independent set σ:

- Choose a vertex v of G uniformly at random.
- Let
$$\sigma' = \begin{cases} \sigma \cup \{v\} & \text{with probability } \frac{\lambda}{1+\lambda} \\ \sigma \setminus \{v\} & \text{with probability } \frac{1}{1+\lambda} \end{cases}$$
- If σ' is a valid independent set, move to state σ' otherwise remain at state σ.

The stationary distribution of MC is the probability measure μ_G. To sample from this distribution, we simply start at an arbitrary independent set and follow the random walk defined by P until we reach stationarity. We show that MC quickly converges to its stationary distribution in a manner we formalize later. In particular, for graphs with maximum degree δ, the number of steps of this random walk that are required to achieve stationarity, known as the mixing time, is $O(n \log n\delta)$ for $\lambda < \frac{2}{\delta-2}$ and $O(n^3\delta^2)$ for $\lambda = \frac{2}{\delta-2}$.

The simpler proof for the case of triangle-free graphs was given in [4].

2 Background

2.1 Markov chain fundamentals

Before getting into the proof, we need to review some background material and machinery. Consider a discrete-time Markov chain (X_t) with transition probability matrix P defined on a finite state space Ω. We let P_i denote the chain starting in state i and $P_i^t(j)$ is the probability this chain is in state j after t steps. A classical theorem of stochastic processes states that if P has the following properties:

- *aperiodicity*: $\gcd\{t : P_i^t(i) > 0\} = 1$ for all $i \in \Omega$
- *irreducibility*: the graph with P as its adjacency matrix is connected

then the chain (X_t) has a unique limiting distribution, referred to as the stationary distribution π, i.e.

$$\lim_{t\to\infty} P_i^t(j) = \pi_j \text{ for all } i, j \in \Omega$$

In fact, π can easily be determined if the Markov chain is time-reversible, i.e. satisfies all local-balance equations. Specifically, a distribution π is the stationary distribution if it satisfies the following:

$$\pi(i)P_i(j) = \pi(j)P_j(i) \text{ for all } i, j \in \Omega$$

Our goal is to bound the time until the chain is sufficiently close to the stationary distribution. The traditional bound on the distance from stationarity is the variation distance,

$$d_{TV}(P_i^t, \pi) = \frac{1}{2} \sum_{j \in \Omega} |P_i^t(j) - \pi(j)|$$

We are interested in the mixing time, τ:

$$\tau(\epsilon) = \max_i \min\{t : d_{TV}(P_i^{t'}, \pi) \leq \epsilon \text{ for all } t' \geq t\}$$

2.2 Coupling

The analysis relies on coupling to bound the mixing time. Coupling contructs a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that separately X_t, Y_t are copies of the original Markov chain and if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

The following standard result ties together coupling and the mixing time. In the theorem, σ' (similarly τ') refers to the state of the chain in state σ after one step of the Markov chain.

Theorem 1 *Let Φ be an integer-valued function defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$. Suppose there exists a $\beta \leq 1$ and a coupling of the Markov chain MC such that for all $\sigma, \tau \in \Omega$:*

$$E[\Phi(\sigma', \tau')] \leq \beta \Phi(\sigma, \tau)$$

If $\beta < 1$ then the mixing time is

$$\tau(\epsilon) \leq \frac{\log(D\epsilon^{-1})}{1 - \beta}$$

If $\beta = 1$ and there exists a positive α such that for all t,

$$Pr[\Phi(\sigma', \tau') \neq \Phi(\sigma, \tau)] \geq \alpha$$

then the mixing time is

$$\tau(\epsilon) \leq \frac{2D^2}{\alpha} \log(\epsilon^{-1})$$

Proof: The proof for the case when $\beta < 1$ is quite simple. Since Φ is non-negative and integer-valued, we have that for all X_0, Y_0

$$Pr[X_t \neq Y_t] \leq E[\Phi(X_t, Y_t)] \leq \beta^t \Phi(X_0, Y_0) \leq \beta^t D.$$

The Coupling Lemma of [1, lemma 3.6] says that the variation distance is bounded above by the probability the chains have not coupled. Using the lemma, taking logarithms, and rearranging terms gives the desired result.

We refer the reader to [3] for the proof of the case $\beta = 1$. ■

3 Arbitrary Graphs

The following section considers a pair of independent sets σ, η and will omit obvious references to them as parameters to functions.

We first need a bit of notation. Let D denote the set of disagree vertices, i.e. D is the symmetric difference between σ and η. The set of agree vertices is $A = V \setminus D$. We use $D_v = \Gamma(v) \cap D$ to denote the disagree neighbors of a vertex v and d_v is the cardinality of D_v. Similarly, $A_v = \Gamma(v) \cap A$.

Let $c = \frac{\delta\lambda}{\delta\lambda+2}$. We now define a potential function between σ and η:

$$\alpha_v = \left\{ \begin{array}{ll} \delta_v & \text{if } v \in D \\ 0 & \text{otherwise} \end{array} \right\}$$

$$\beta_v = \left\{ \begin{array}{ll} -cd_v & \text{if there exists a neighbor } w \text{ of } v \text{ such that } w \in \sigma, w \in \eta \\ -c(d_v - 1) & \text{if there is no such } w \text{ and } d_v > 1 \\ 0 & \text{otherwise} \end{array} \right\}$$

$$\Phi = \sum_v [\alpha_v + \beta_v]$$

4 Analysis

We now analyze the expected change Φ after the next transition of the Markov chain. Our coupling is simply the identity, i.e. each chain attempts the same move at every step. For simplicity, we rescale everything by a factor $n(1+\lambda)$.

$$n(1+\lambda)E[\Delta\Phi] = n(1+\lambda)\sum_v E[\Delta\alpha_v] + E[\Delta\beta_v]$$

We first try to manipulate the terms in the expected change in Φ to ease the analysis.

Observe that for an agree vertex v if all of its' neighbors agree then α_v is still 0 after the next move. For an agree vertex w, we can then amortize the expected change in α_w over its disagree neighbors as follows.

$$\sum_v E[\Delta\alpha_v] = \sum_{v\in D} E[\Delta\alpha_v] + \sum_{w\in A_v} \frac{1}{d_w} E[\Delta\alpha_w] \quad (1)$$

To simplify our accounting, we divide $E[\Delta\beta_v]$ as follows.

Let,

$$E[\Delta^w \beta_v] = E[\Delta\beta_v | \text{Markov chain transitions on } w]$$

We then have the following from the definition of β_v.

$$\begin{aligned} \textstyle\sum_v E[\Delta\beta_v] = \quad & = \textstyle\sum_{v,w} E[\Delta^w \beta_v] \\ & = \textstyle\sum_v \sum_{w\in\Gamma(v)} E[\Delta^w \beta_v] \end{aligned}$$

For an agree vertex w, we can now try to amortize the expected change in β_w over its disagree neighbors.

$$\sum_v E[\Delta\beta_v] = \sum_{v\in D}\left[\sum_{w\in\Gamma(v)} E[\Delta^w\beta_v] + \sum_{w\in A_v} E[\Delta^v\beta_w]\right] + \sum_{w\in A}\sum_{x\in A_w} E[\Delta^w\beta_x]$$

Observe that the following is also true.

$$\sum_v E[\Delta\beta_v] = \sum_{v\in D}\left[\sum_{w\in\Gamma(v)} E[\Delta^w\beta_v] + \sum_{w\in A_v}\left(E[\Delta^v\beta_w] + \sum_{x\in A_w}\frac{1}{d_w}E[\Delta^w\beta_x]\right)\right] + \sum_{w\in A}\sum_{x\in A_w, d_x=0} E[\Delta^x\beta_w]$$

Observe that if a vertex v agrees, all neighbors w of v agree, and all neighbors of w agree, then we are guaranteed that β_v is 0 and after the next move will still be 0.

$$\sum_v E[\Delta\beta_v] = \sum_{v\in D}\left[\sum_{w\in\Gamma(v)} E[\Delta^w\beta_v] + \sum_{w\in A_v}\left(E[\Delta^v\beta_w] + \sum_{x\in A_w}\frac{1}{d_w}E[\Delta^w\beta_x] + \sum_{x\in A_w, d_x=0}\frac{1}{d_w}E[\Delta^x\beta_w]\right)\right] \quad (2)$$

Using (1) and (2), we can divide the expected change in Φ over the disagree vertices as follows.

$$E[\Delta\Phi] = \sum_{v\in D}\left[E[\Delta\alpha_v] + \sum_{w\in\Gamma(v)}\phi_v(w)\right]$$

where $\phi_v(w)$ is the following:

$$\phi_v(w) = \begin{cases} E[\Delta^w\beta_v] & \text{if } w\in D \\ \frac{1}{d_w}E[\Delta\alpha_w] + E[\Delta^w\beta_v] + E[\Delta^v\beta_w] + \frac{1}{d_w}\left(\sum_{x\in A_w}E[\Delta^w\beta_x] + \sum_{x\in A_w, d_x=0}E[\Delta^x\beta_w]\right) & \text{if } w\in A \end{cases}$$

The analysis will show that for each disagree vertex v, $E[\Delta\alpha_v] + \sum_{w\in\Gamma(v)}\phi_v(w) \le 0$.

Consider a disagree vertex v, without loss of generality we assume $v\in\sigma, v\notin\eta$. Recall, $\alpha_v = \delta_v$. The move that attempts to remove v from both independent sets definitely causes v to agree and has weight 1. We also know that the move which attempts to add v to both sets works in both sets if $d_v = 0$. This move occurs with weight λ. Thus,

$$E[\Delta\alpha_v] = \begin{cases} -(1+\lambda)\delta_v & \text{if } d_v = 0 \\ -\delta_v & \text{otherwise} \end{cases} \quad (3)$$

For a disagree vertex v and a neighbor w, we analyze $\phi_v(w)$ based on the following cases:

- $w\in D$:

  We only need to consider $E[\Delta^w\beta_v]$. The only move that changes the configuration at w is attempting to remove w from both independent sets. Since neighbors of v are either disagree vertices or out of both independent sets, removing w only changes β_v if $d_v > 1$. In this case, we then have that

  $$\phi_v(w) = \begin{cases} c & \text{if } d_v > 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

- $w\in A$ and w has a neighbor z such that $z\in\eta$:

  In this scenario there are no moves that change the configuration at w. Thus for all neighbors x of w,

  $$E[\Delta\alpha_w] = E[\Delta^w\beta_v] = E[\Delta^w\beta_x] = 0$$

  Now let us consider $E[\Delta^v\beta_w]$. Notice that if $d_v = 0$ then v agrees after the moves which attempt to add or remove v from both independent sets. Whereas if $d_v > 0$ then v only agrees after the move which attempts to remove it from both sets. In the worst case, these moves cause β_w to increase by c.

  $$E[\Delta^v\beta_w] \le \begin{cases} c(1+\lambda) & \text{if } d_v = 0 \\ c & \text{otherwise} \end{cases}$$

  We still need to consider $E[\Delta^x\beta_w]$ where x is an agree neighbor of w and all the neighbors of x agree. Attempting to add x to both independent sets can only decrease β_w. Also, attempting to remove x from both sets can only have an effect if x is already in both sets. If there is one such x in both sets, then removing it from both sets may increase β_w by c:

$$\sum_{x \in A_v, d_x = 0} E[\Delta^x \beta_w] \leq c$$

Since the worst case is when $d_w = 1$, we have that

$$\phi_v(w) \leq \begin{cases} c(2+\lambda) & \text{if } d_v = 0 \\ 2c & \text{otherwise} \end{cases}$$

- $w \in A$, and no neighbors of x are in η:

  We know that $\alpha_w = 0$. There is one move that will cause w to disagree. Specifically, attempting to add w to both independent sets will work in exactly one of the sets. This increases α_w by δ_w. Similarly, if $d_v > 0$ this move will decrease β_v by c.

$$E[\Delta \alpha_w] = \lambda \delta_w$$

$$E[\Delta^w \beta_v] = \begin{cases} -c\lambda & \text{if } d_v > 0 \\ 0 & \text{otherwise} \end{cases}$$

  Now look at $E[\Delta^v \beta_w]$. If $d_w > 1$ then the move which attempts to remove v from both independent sets will increase β_w by c. Whereas attempting to add v to both sets will not effect β_w.

$$E[\Delta^v \beta_w] = \begin{cases} c & \text{if } d_w > 0 \\ 0 & \text{otherwise} \end{cases}$$

  Consider an agree neighbor x of w. Suppose x has some disagree neighbors or a neighbor which is in both independent sets. Then the move which attempts to add w to both independent sets and causes w to disagree will decrease β_x by c. In the other case for x that all of its neighbors are out of both independent sets then the move that adds x to both sets decreases β_w by c and occurs with weight λ. Thus, for all agree neighbors x of w, either: (i) $E[\Delta^w_x] = -c\lambda$ or (ii) $E[\Delta^x \beta_w] = -c\lambda$ and $d_x = 0$.

$$\sum_{x \in A_v} E[\Delta^w \beta_x] + \sum_{x \in A_v, d_x = 0} E[\Delta^x \beta_w] = -c\lambda(\delta_w - d_w)$$

  For this case we have that

$$\phi_v(w) = \begin{cases} \lambda\delta_w - c\lambda(\delta_w - 1) & \text{if } d_w = 1, d_v = 0 \\ \lambda\delta_w - c\lambda(\delta_w - 1) - c\lambda & \text{if } d_w = 1, d_v > 0 \\ \frac{1}{d_w}(\lambda\delta_w - c\lambda(\delta_w - d_w)) + c & \text{if } d_w > 1, d_v = 0 \\ \frac{1}{d_w}(\lambda\delta_w - c\lambda(\delta_w - d_w)) + c - c\lambda & \text{if } d_w > 1, d_v > 0 \end{cases}$$

  Recall our setting of $c = \frac{\delta\lambda}{\delta\lambda+2}$.

  We leave it to the reader to verify that once again the worst case is when $d_w = 1$.

$$\phi_v(w) \leq \begin{cases} \lambda\delta_w - c\lambda(\delta_w - 1) & \text{if } d_v = 0 \\ \lambda\delta_w - c\lambda(\delta_w - 1) - c\lambda & \text{otherwise} \end{cases}$$

Notice that for our setting of c, the following are true:

$$\begin{aligned} (2+\lambda)c &\geq \lambda\delta_w - c\lambda(\delta_w - 1) \\ 2c &\geq \lambda\delta_w - c\lambda(\delta_w - 1) - c\lambda \end{aligned}$$

We then have that for $w \in A$,

$$\phi_v(w) \leq \begin{cases} (2+\lambda)c & \text{if } d_v = 0 \\ 2c & \text{otherwise} \end{cases} \tag{5}$$

Using (3), (4), and (5), we have the following:

$$E[\Delta\alpha_v] + \sum_{w \in \Gamma(v)} \phi_v(w) \leq \begin{cases} -(1+\lambda)\delta_v + (2+\lambda)c\delta_v & \text{if } d_v = 0 \\ -\delta_v + 2c(\delta_v - 1) & \text{if } d_v = 1 \\ -\delta_v + 2c(\delta_v - d_v) + d_v c & \text{otherwise} \end{cases}$$

Using the facts that for $d_v > 1$,

$$-(1+\lambda)\delta_v + (2+\lambda)c\delta_v = -\delta_v + 2c(\delta_v - 1) \geq -\delta_v + 2c(\delta_v - d_v) + d_v c$$

$$\frac{\delta_v}{\delta\lambda + 2}[\lambda(\delta - 2) - 2] \geq -(1+\lambda)\delta_v + (2+\lambda)c\delta_v$$

Therefore,

$$E[\Delta\Phi] \leq \frac{1}{n(1+\lambda)} \frac{\sum_{v \in D} \delta_v}{\delta\lambda + 2}[\lambda(\delta - 2) - 2]$$

Notice that $E[\Delta\Phi] < 0$ when $\lambda < \frac{2}{\delta - 2}$.
We now want to rephrase this bound on $E[\Delta\Phi]$ to get a bound on $\beta = \max_{\sigma,\tau}$ where

$$E[\phi(\sigma', \tau')] = \beta_{\sigma,\tau}\Phi(\sigma, \tau)$$

We bound β as follows:

$$\begin{aligned} \beta_{\sigma,\tau}\Phi(\sigma,\tau) &= E[\Phi(\sigma', \tau')] \\ (\beta_{\sigma,\tau} - 1)\Phi(\sigma,\tau) &= E[\Phi(\sigma', \tau')] - \Phi(\sigma,\tau) = E[\Delta\Phi] \\ \beta_{\sigma,\tau} &= 1 + \frac{E[\Delta\Phi]}{\Phi} \end{aligned}$$

Since $\Phi \leq \sum_{v \in D} \delta_v$, we get a bound on β of:

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

We rescale Φ by $\Phi' = \frac{\Phi}{c}$ to make it integer-valued. Thus, $\Phi' \leq \frac{n\delta}{c}$. Plugging these bounds into theorem 1 we get that when $\lambda < \frac{2}{\delta - 2}$,

$$\tau(\epsilon) \leq \frac{n(1+\lambda)(\delta\lambda + 2)}{2 - \lambda(\delta - 2)} \log(\frac{n\delta}{c\epsilon})$$

Using the fact that $\delta \geq 3, \lambda \leq \frac{2}{\delta - 2}$, we get $\lambda \leq 2, \delta\lambda \leq 6, c \geq \frac{1}{3}$.
We can now simplify the bound on the mixing time. For $\lambda = (1 - \alpha)\frac{2}{\delta - 2}$, where α is positive,

$$\tau(\epsilon) \leq \frac{48n}{\alpha} \log(3n\delta/\epsilon)$$

Theorem 2 *For graphs of maximum degree δ, MC mixes in time $O(\frac{n}{\alpha}\log(n\delta/\epsilon))$ when $\lambda = (1 - \alpha)\frac{2}{\delta - 2}$ for positive $\alpha < 1$.*

Corollary 3 *For lattices of degree δ, the limiting Gibbs measure is unique when $\lambda < (1-\alpha)\frac{2}{\delta - 2}$ for fixed positive $\alpha < 1$.*

When $\lambda = \frac{2}{\delta-2}$ we have that $\beta = 1$. We simply need to bound $\alpha = Pr[\Delta\Phi \neq 0]$.

For any disagree vertex v, consider the move that attempts to remove v from both sets. This move reduces α_v by δ_v. Also, this move may increase β_w for w which are neighbors of v. Since $c < 1$, we are guaranteed this move changes Φ by at least $\delta_v(1-c) > 0$. Therefore, $\alpha \geq \frac{1}{n(1+\lambda)}$.

When $\lambda = \frac{2}{\delta-2}$,

$$\tau(\epsilon) \leq \frac{2n^3\lambda^2(1+\lambda)}{c^2}\log(\epsilon^{-1}) \leq 54n^3\delta^2\log(\epsilon^{-1})$$

Theorem 4 *For graphs of maximum degree δ, MC mixes in time $O(n^3\delta^2\log(\epsilon^{-1}))$ when $\lambda = \frac{2}{\delta-2}$.*

References

[1] D.J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilites XVII*, pages 243–297. Springer-Verlag, 1983. Lecture Notes in Math. 986.

[2] Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter and Co., Berlin-New York, 1988.

[3] Michael Luby, Dana Randall, and Alistair Sinclair. Markov chain algorithms for planar lattice structures (extended abstract). In *36th Annual Symposium on Foundations of Computer Science*, pages 150–159, Milwaukee, Wisconsin, 23–25 October 1995. IEEE.

[4] Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets: Part I. *Random Structures and Algorithms*, 1999. To appear in special issue based on DIMACS workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science.