

On Markov chains for independent sets*

Martin Dyer and Catherine Greenhill
School of Computer Studies
University of Leeds
Leeds LS2 9JT
United Kingdom

Submitted: 8 December 1997. Revised: 16 February 1999
Second revision: 8 November 1999

Abstract

Random independent sets in graphs arise, for example, in statistical physics, in the hard-core model of a gas. In 1997, Luby and Vigoda described a rapidly mixing Markov chain for independent sets, which we refer to as the Luby–Vigoda chain. A new rapidly mixing Markov chain for independent sets is defined in this paper. Using path coupling, we obtain a polynomial upper bound for the mixing time of the new chain for a certain range of values of the parameter λ . This range is wider than the range for which the mixing time of the Luby–Vigoda chain is known to be polynomially bounded. Moreover, the upper bound on the mixing time of the new chain is always smaller than the best known upper bound on the mixing time of the Luby–Vigoda chain for larger values of λ (unless the maximum degree of the graph is 4). An extension of the chain to independent sets in hypergraphs is described. This chain gives an efficient method for approximately counting the number of independent sets of hypergraphs with maximum degree two, or with maximum degree three and maximum edge size three. Finally, we describe a method which allows one, under certain circumstances, to deduce the rapid mixing of one Markov chain from the rapid mixing of another, with the same state space and stationary distribution. This method is applied to two Markov chains for independent sets, a simple insert/delete chain and the new chain, to show that the insert/delete chain is rapidly mixing for a wider range of values of λ than was previously known.

1 Introduction

An independent set in a graph G is a subset of vertices, no two of which are adjacent. In statistical physics, independent sets arise in the hard-core model of a gas, where a vertex is a possible site for a particle and adjacent sites cannot be simultaneously occupied. The probability $\pi(X)$ that a particular configuration X is observed is proportional to $\lambda^{|X|}$ for some positive parameter λ . The normalising factor, $Z(\lambda)$, is called the *partition function* of the system. The two main tasks are evaluating the partition function and sampling according to the probability distribution π . Both of these problems can be

*Supported by ESPRIT Working Group RAND2.

solved in polynomial time using the Markov Chain Monte Carlo method wherever a rapidly mixing Markov chain for independent sets is available (see, for example [18]).

The simplest Markov chain for independent sets is the so called insert/delete chain. This chain was only known to be rapidly mixing for small values of λ . A slightly more complicated chain was proposed by Luby and Vigoda [21]. We refer to this chain as the Luby–Vigoda chain. A rapidly mixing Markov chain for independent sets of fixed size was introduced by Bubley and Dyer [3]; in this paper we concentrate on results relating to the set $\mathcal{I}(G)$ of *all* independent sets in the given graph.

In [21], Luby and Vigoda used the coupling method to establish the rapid mixing of their chain. Here we state tighter bounds on the mixing time of the Luby–Vigoda chain which can be established using the *path coupling* method [3]. Then a new Markov chain for independent sets is defined. This chain was independently discovered by Luby, Mitzenmacher and Vigoda, as mentioned in [20].

The new chain is an improvement on the Luby–Vigoda chain in two areas. The upper bound on the mixing time of the new chain is polynomial for a wider range of values of λ than the best known upper bound on the mixing time of the Luby–Vigoda chain. Moreover, the new chain has better time bounds than the Luby–Vigoda chain for “large” values of λ (unless the maximum degree Δ of the graph is 4). We do not know whether this apparent improvement is an artefact of the analysis, or whether the new chain really is more rapidly mixing (or rapidly mixing for a wider class of graphs) than the Luby–Vigoda chain.

To compare the time bounds of the new chain and the Luby–Vigoda chain, we calculate the ratio of the best known upper bounds of the respective mixing times. In fact we show that this ratio tends to infinity as λ increases. The Luby–Vigoda chain has better time bounds on Δ -regular graphs for small values of λ , but never by more than a factor of two. While comparing best known upper bounds on the mixing times is not entirely satisfactory (we cannot claim, for example, that one chain mixes more rapidly than the other), in the context of Markov chain Monte Carlo algorithms it is the upper bound which determines the number of steps of the chain to be simulated per sample.

Like the Luby–Vigoda chain, the new chain can be used to approximately count independent sets in graphs of degree at most four, and we give better time bounds for this than those given in [21]. Specifically, the time bound given in [21] for this task is $O(n^3 \log(n))$, while our chain has time bound $O(n \log(n))$ if the graph has maximum degree three, and $O(n^2 \log(n))$ if the graph has maximum degree four, where n is the number of vertices of the graph.

The new chain is easily extended to act on independent sets of hypergraphs. The upper bound on the mixing time increases with the size of the largest edge in the hypergraph. We show that the new chain can be used to approximately count (in polynomial time) independent sets in hypergraphs where either each vertex is contained in at most two edges, or each vertex is contained in at most three edges and each edge contains at most three vertices. The former problem is related to the problem of approximately counting edge covers of a graph. The bound on the mixing time of the new chain is $\Omega^*(n^2)$ smaller than the bound on the mixing time of the only previously available Markov chain for edge covers in graphs (where the $\Omega^*(\cdot)$ notation hides factors of $\log(n)$). As far as we are aware, the Markov chain presented here is the first which can be used to approximately count the number of independent sets in hypergraphs

of maximum degree three, maximum edge size three. We show that the problem of counting independent sets in graphs with maximum degree 3 is $\#P$ -complete, and that it is $\#P$ -complete to count independent sets in hypergraphs where every edge has at most three vertices and each vertex is contained in at most three edges.

To conclude, we show how the rapid mixing of one Markov chain can, under certain conditions, be used to deduce the rapid mixing of another Markov chain with the same state space and stationary distribution. This method is useful in the following situation. Suppose that a known Markov chain for a given state space is rapidly mixing. If a simpler Markov chain exists with the same state space and stationary distribution, one might prefer to work with this simple chain when performing simulations (such as in a Markov chain Monte Carlo algorithm). If the specified conditions are met, we can deduce that the simple chain is also rapidly mixing.

The method involves relating the spectral gap of the two chains, using a linear relationship between the entries of the transition matrices of the chains. It gives a simple alternative to the approaches described in [7, 8, 9, 25]. We illustrate our method using the insert/delete chain and the new chain. This allows us to demonstrate that the simple insert/delete chain is rapidly mixing for a much wider range of values of λ than was previously known.

The plan of the paper is as follows. In the remainder of the introduction we review the path coupling method. In Section 2 we introduce some notation and describe the insert/delete chain and the Luby–Vigoda chain. The new chain is described in Section 3 and a proof by path coupling is given to show that it is rapidly mixing for a wider range of values of λ than the Luby–Vigoda chain is known to be. An extension of the new chain to independent sets of hypergraphs is described in Section 4. Classes of hypergraphs for which this chain mixes rapidly when $\lambda = 1$ are discussed. In Section 5 we develop and apply the new method referred to above which allows us, under certain conditions, to deduce the rapid mixing of one Markov chain from the rapid mixing of another with the same state space and stationary distribution.

Note added: Since submitting this paper, we have learnt that Luby and Vigoda [20] have analysed the insert/delete chain, using path coupling with respect to a specially defined metric. They have shown that the insert/delete chain is rapidly mixing for the same range of values as the new chain described in this paper. (The proof in [20] only considers triangle-free graphs, and the proof for arbitrary graphs uses coupling.) The upper bound on the mixing time that they obtain (in the triangle-free case) is a constant times larger than the upper bound on the mixing time that we prove for the new chain. For example, suppose that the maximum degree of the graph is 4. As λ tends to 1, the upper bound on the mixing time of the Markov chain given in [20] is almost three times bigger than the upper bound on the mixing time of the chain described in this paper.

1.1 A review of path coupling

Let Ω be a finite set and let \mathcal{M} be a Markov chain with state space Ω , transition matrix P and unique stationary distribution π . In order for a Markov chain to be useful for almost uniform sampling or approximate counting, it must converge quickly towards its stationary distribution π . We make this notion more precise below. If the initial state of the Markov chain is x then the distribution of the chain at time t is given by

$P^t_x(y) = P^t(x, y)$. The *total variation distance* of the Markov chain from π at time t , with initial state x , is defined by

$$d_{\text{TV}}(P^t_x, \pi) = \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|.$$

Following Aldous [1], let $\tau_x(\varepsilon)$ denote the least value T such that $d_{\text{TV}}(P^t_x, \pi) \leq \varepsilon$ for all $t \geq T$. The *mixing time* of \mathcal{M} , denoted by $\tau(\varepsilon)$, is defined by $\tau(\varepsilon) = \max \{\tau_x(\varepsilon) : x \in \Omega\}$. A Markov chain will be said to be *rapidly mixing* if the mixing time is bounded above by some polynomial in n and $\log(\varepsilon^{-1})$, where n is a measure of the size of the elements of Ω . (Throughout this paper all logarithms are to base e .) Sometimes it is convenient to specify an *initial distribution* μ for the Markov chain, rather than an initial state. Then the total variation distance at time t is given by

$$d_{\text{TV}}(\mu P^t, \pi) = \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{z \in \Omega} \mu(z) P^t(z, y) - \pi(y) \right|.$$

One can easily prove that

$$d_{\text{TV}}(\mu P^t, \pi) \leq \max \{d_{\text{TV}}(P^t_x, \pi) : x \in \Omega\} \quad (1)$$

for all initial distributions μ .

There are relatively few methods available to prove that a Markov chain is rapidly mixing. One such method is *coupling*. A *coupling* for \mathcal{M} is a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that each of (X_t) , (Y_t) , considered marginally, is a faithful copy of \mathcal{M} . The *Coupling Lemma* (see for example, Aldous [1]) states that the total variation distance of \mathcal{M} at time t is bounded above by $\text{Prob}[X_t \neq Y_t]$, the probability that the process has not *coupled*. The difficulty in applying the Coupling Lemma lies in obtaining an upper bound for this probability. In the *path coupling* method, introduced by Bubley and Dyer [3], one need only define and analyse a coupling on a subset S of $\Omega \times \Omega$. Choosing the set S carefully can considerably simplify the arguments involved in proving rapid mixing by coupling. The path coupling method is described in the next theorem, taken from [12]. We use the term *path* to describe a sequence of elements of the state space, which *need not* necessarily be a sequence of possible transitions of the Markov chain.

Theorem 1.1 *Let δ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \dots, D\}$. Let S be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path*

$$X_t = Z_0, Z_1, \dots, Z_r = Y_t$$

between X_t and Y_t such that $(Z_l, Z_{l+1}) \in S$ for $0 \leq l < r$ and

$$\sum_{l=0}^{r-1} \delta(Z_l, Z_{l+1}) = \delta(X_t, Y_t).$$

Define a coupling $(X, Y) \mapsto (X', Y')$ of the Markov chain \mathcal{M} on all pairs $(X, Y) \in S$. Suppose that there exists $\beta \leq 1$ such that $\mathbf{E}[\delta(X', Y')] \leq \beta \delta(X, Y)$ for all $(X, Y) \in S$. If $\beta < 1$ then the mixing time $\tau(\varepsilon)$ of \mathcal{M} satisfies

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

If $\beta = 1$ and there exists $\alpha > 0$ such that $\text{Prob}[\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)] \geq \alpha$ for all t , then

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\alpha} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Remark 1.1 In many applications the set S is defined by

$$S = \{(X, Y) \in \Omega \times \Omega : \delta(X, Y) = 1\}.$$

Here one need only define and analyse a coupling on pairs at distance 1 apart.

2 Known Markov chains for independent sets in graphs

Let $G = (V, E)$ be a graph. A subset X of V is an *independent set* if $\{v, w\} \notin E$ for all $v, w \in X$. Let $\mathcal{I}(G)$ be the set of all independent sets in a given graph G and let λ be a positive number. The partition function $Z = Z(\lambda)$ is defined by

$$Z = Z(\lambda) = \sum_{X \in \mathcal{I}(G)} \lambda^{|X|}.$$

The function π defined by

$$\pi(X) = \frac{\lambda^{|X|}}{Z}$$

is a probability measure on $\mathcal{I}(G)$. The two main tasks are to approximately evaluate the partition function $Z(\lambda)$ and to approximately sample from $\mathcal{I}(G)$ according to the distribution π . When $\lambda = 1$ the task of approximately evaluating $Z(\lambda)$ is equivalent to approximately counting $\mathcal{I}(G)$. Note that it is NP-hard to compute $Z(\lambda)$ to within any polynomial factor whenever $\lambda > c/\Delta$ for some constant $c > 0$ (unless $\text{NP} = \text{RP}$). For a proof of this result see [21, Theorem 4].

For the remainder of the paper we will assume that the maximum degree Δ of the graph G is at least 3. This assumption is justified by the following theorem, the proof of which is easy and omitted.

Theorem 2.1 *Let G be a graph with maximum degree Δ and let $\mathcal{I}(G)$ be the set of all independent sets in G . Suppose that $0 \leq \Delta \leq 2$. Then we can evaluate $Z(\lambda)$ exactly (in polynomial time) for all $\lambda > 0$. Moreover there exists a polynomial-time procedure for sampling from $\mathcal{I}(G)$ according to the distribution π .*

For $\Delta \geq 3$ the tasks of approximately evaluating $Z(\lambda)$ and approximately sampling from $\mathcal{I}(G)$ according to π can be performed using a rapidly mixing Markov chain with state space $\mathcal{I}(G)$ and stationary distribution π . For a description of how this can be achieved, see for example [18]. The simplest Markov chain on $\mathcal{I}(G)$ which converges to the stationary distribution π is the so-called *insert/delete* chain. If X_t is the state at time t then the state at time $t + 1$ is determined by the following procedure:

- (i) choose a vertex v uniformly at random from V ,
- (ii) (Delete) if $v \in X_t$ then let $X_{t+1} = X_t \setminus \{v\}$ with probability $1/(1 + \lambda)$,
 (Insert) if $v \notin X_t$ and v has no neighbours in X_t then let $X_{t+1} = X_t \cup \{v\}$
 with probability $\lambda/(1 + \lambda)$,
 otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic with stationary distribution equal to π . Given $X, Y \in \mathcal{I}(G)$, let $H(X, Y)$ denote the *Hamming distance* between X and Y which equals $|X \setminus Y| + |Y \setminus X|$. A bound on the mixing time of the insert/delete chain is stated below. This bound is obtained using the path coupling method on states at distance 1 apart. The details are omitted.

Theorem 2.2 *Let G be a graph with maximum degree Δ . The insert/delete Markov chain is rapidly mixing for $\lambda \leq 1/(\Delta - 1)$. When $\lambda < 1/(\Delta - 1)$ the mixing time $\tau(\varepsilon)$ of the insert/delete chain satisfies*

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

When $\lambda = 1/(\Delta - 1)$ the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil 2n^2 e(1 + \lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Remark 2.1 Given the above bound on the mixing time of the insert/delete chain, we can only guarantee rapid mixing at $\lambda = 1$ when the input graph has maximum degree 2. Since $|\mathcal{I}(G)|$ can be calculated exactly for these graphs (see Theorem 2.1), this would suggest that the insert/delete chain is not useful in terms of approximate counting. However, the results of Section 5 below show that the insert/delete chain is rapidly mixing for a wider range of values of λ than stated above. In particular, we show that the insert/delete chain can be used to approximately count independent sets in graphs with maximum degree at most four. *Note added:* Since submission of this paper, Luby and Vigoda [20] have proved rapid mixing of the insert/delete chain for the same range of values of λ as we consider in Section 5, but with a better bound on the mixing time.

In [21], Luby and Vigoda proposed the following Markov chain on state space $\mathcal{I}(G)$, which we shall denote by $\mathcal{LV}(\mathcal{I}(G))$. If the Markov chain is at state X_t at time t , the state at time $t + 1$ is determined by the following procedure:

- (i) choose an edge $e = \{v, w\}$ from E uniformly at random,
- (ii)
 - with probability $\lambda/(1 + 2\lambda)$ let $X' = (X_t \cup \{v\}) \setminus \{w\}$,
 - with probability $\lambda/(1 + 2\lambda)$ let $X' = (X_t \cup \{w\}) \setminus \{v\}$,
 - with probability $1/(1 + 2\lambda)$ let $X' = X_t \setminus \{v, w\}$,
- (iii) if $X' \in \mathcal{I}(G)$ then let $X_{t+1} = X'$ else let $X_{t+1} = X_t$.

The Luby–Vigoda chain is ergodic with stationary distribution π . In [21], the chain is shown to be rapidly mixing for $\Delta \geq 4$, $\lambda \leq 1/(\Delta - 3)$, using the coupling method. A bound for the mixing time of the Luby–Vigoda chain is given below which is an improvement on that stated in [21]. The Luby–Vigoda chain is also rapidly mixing when $\Delta = 3$, $\lambda \leq 1$, as stated in [21]. A bound for the mixing time of the chain in this case is stated below. Both bounds were obtained using path coupling on pairs at distance 1 apart. The details are omitted, but note that the bounds improve those given in [21] by a factor of $\Omega(n)$.

Theorem 2.3 *Let G be a graph with maximum degree Δ and minimum degree δ . The Markov chain $\mathcal{LV}(\mathcal{I}(G))$ is rapidly mixing for $\Delta = 3$, $\lambda \leq 1$ and for $\Delta \geq 4$, $\lambda \leq 1/(\Delta - 3)$. When $\Delta = 3$ and $\lambda < 1$ the mixing time satisfies*

$$\tau_{LV}(\varepsilon) \leq \frac{1 + 2\lambda}{\delta(1 - \lambda)} |E| \log(n\varepsilon^{-1}).$$

When $\Delta \geq 4$ and $\lambda < 1/(\Delta - 3)$ the mixing time $\tau_{LV}(\varepsilon)$ satisfies

$$\tau_{LV}(\varepsilon) \leq \frac{1 + 2\lambda}{\delta(1 - (\Delta - 3)\lambda)} |E| \log(n\varepsilon^{-1}).$$

When (a) $\Delta = 3$ and $\lambda = 1$ or (b) $\Delta \geq 4$ and $\lambda = 1/(\Delta - 3)$ the mixing time satisfies

$$\tau_{LV}(\varepsilon) \leq \lceil 2ne |E|(1 + 2\lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Remark 2.2 The Luby–Vigoda chain gives an efficient algorithm for approximately counting of independent sets in graphs with maximum degree at most four.

Remark 2.3 In all but the boundary cases (i.e. when $\Delta > 3$ and $\lambda = 1/(\Delta - 3)$, or $\Delta = 3$ and $\lambda = 1$), the bound on the mixing time of the Markov chain $\mathcal{LV}(\mathcal{I}(G))$ is proportional to $|E|/\delta$, where δ is the minimum degree of the graph G . Therefore the chain may be Δ times more rapidly mixing for a Δ -regular graph than for another graph with maximum degree Δ . It is possible to improve this situation by embedding G into a Δ -regular multigraph $G' = (V, E')$, adding self-loop edges at vertices with degree less than Δ . A Markov chain on $\mathcal{I}(G')$ can be defined which is rapidly mixing for $\lambda \leq 1/(\Delta - 3)$ when $\Delta > 3$, and for $\lambda \leq 1$ if $\Delta = 3$. Moreover, the mixing time of this chain can be bounded above by the upper bound established in Theorem 2.3 for the Luby–Vigoda chain on regular graphs, in the following situations: when $\Delta \geq 4$ and $1/(2\Delta - 5) \leq \lambda \leq 1/(\Delta - 3)$, or when $\Delta = 3$ and $1/3 \leq \lambda \leq 1$. The details are omitted.

Remark 2.4 Recently, Dyer, Frieze and Jerrum [13] have given a moderate value of Δ , above which approximate counting is impossible unless $\text{RP}=\text{NP}$. Moreover, they show that, for $\Delta \geq 6$, no random walk can converge in polynomial time if it adds to, or deletes from, the independent set only a “small” number of vertices. Thus, for random walks of the type described here, the issue of rapid mixing is only open for $\Delta = 5$.

3 A new Markov chain for independent sets in graphs

In this section we define a new chain $\mathcal{M}(\mathcal{I}(G))$ with state space $\mathcal{I}(G)$, the set of all independent sets in the graph G . The new Markov chain can perform the moves of the insert/delete chain, and can also perform a new kind of move, called a *drag* move. If X_t is the state at time t then the state at time $t + 1$ is determined by the following procedure:

- (i) choose a vertex v uniformly at random from V ,
- (ii) (Delete) if $v \in X_t$ then let $X_{t+1} = X_t \setminus \{v\}$ with probability $1/(1 + \lambda)$,
 (Insert) if $v \notin X_t$ and v has no neighbours in X_t then
 let $X_{t+1} = X_t \cup \{v\}$ with probability $\lambda/(1 + \lambda)$,
 (Drag) if $v \notin X_t$ and v has a unique neighbour u in X_t then
 let $X_{t+1} = (X_t \cup \{v\}) \setminus \{u\}$ with probability $\lambda/(4(1 + \lambda))$,
 otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic with stationary distribution π . Recall the definition of the *line graph* $L(\Gamma)$ of a graph Γ , with a vertex for every edge in Γ and edges in $L(\Gamma)$ between adjacent edges in Γ . An independent set in $L(\Gamma)$ corresponds to a *matching* in Γ : that is, a set of edges no two of which are adjacent. The Markov chain $\mathcal{M}(\mathcal{I}(G))$ defined above corresponds almost exactly to the Markov chain on matchings described in [18, p. 495] (in the matchings chain, a drag move is performed with probability 1).

The set of transitions which this Markov chain can perform is identical to the set of transitions performed by the Luby–Vigoda chain. However, the probabilities with which these transitions are performed can be very different. For example, the Luby–Vigoda chain will insert the vertex v with probability which depends on the degree of v . The probability of insertion in the new chain is independent of the degree of the vertex.

We prove below that $\mathcal{M}(\mathcal{I}(G))$ is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$ using the path coupling method on pairs at Hamming distance 1 apart. This result improves upon that of Luby–Vigoda when $\Delta = 3$ or $\Delta \geq 5$. A comparison of the best known upper bounds for the mixing times of the two chains is given at the end of this section.

Theorem 3.1 *Let G be a graph with maximum degree Δ . The Markov chain $\mathcal{M}(\mathcal{I}(G))$ is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$. When $\lambda < 2/(\Delta - 2)$ the mixing time $\tau(\varepsilon)$ satisfies*

$$\tau(\varepsilon) \leq \frac{2(1 + \lambda)}{2 - (\Delta - 2)\lambda} n \log(n\varepsilon^{-1}).$$

When $\lambda = 2/(\Delta - 2)$ the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil 2n^2 e(1 + \lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Proof. Let $(X_t, Y_t) \in \mathcal{I}(G) \times \mathcal{I}(G)$ be given. We can certainly construct a path

$$X_t = Z_0, \dots, Z_r = Y_t$$

between X_t and Y_t such that $H(Z_i, Z_{i+1}) = 1$ for $0 \leq i < r$ and $Z_i \in \mathcal{I}(G)$ for $0 \leq i \leq r$, where $r = H(X_t, Y_t)$. The path may be defined by removing each element of $X_t \setminus Y_t$

one by one until the independent set $X_t \cap Y_t$ is obtained, then by adding each element of $Y_t \setminus X_t$ in turn. Therefore it suffices to define a coupling on elements (X, Y) such that $H(X, Y) = 1$. Let X and Y be independent sets which differ just at a vertex v with degree d . Without loss of generality, assume that $v \in X \setminus Y$. We now define the coupling on (X, Y) . Choose a vertex $w \in V$ uniformly at random. Suppose first that w does not satisfy any of the following conditions:

- (i) $w = v$,
- (ii) $\{w, v\} \in E$ and w has no neighbour in Y ,
- (iii) $\{w, v\} \in E$ and w has a unique neighbour in Y .

Then perform the following procedure:

- if $w \in X$ then let $(X', Y') = (X \setminus \{w\}, Y \setminus \{w\})$ with probability $1/(\lambda + 1)$,
- if $w \notin X$ and w has no neighbour in X then let $(X', Y') = (X \cup \{w\}, Y \cup \{w\})$ with probability $\lambda/(1 + \lambda)$,
- if $w \notin X$ and w has a unique neighbour u in X then let

$$(X', Y') = ((X \cup \{w\}) \setminus \{u\}, (Y \cup \{w\}) \setminus \{u\})$$

with probability $\lambda/(4(1 + \lambda))$,

- otherwise let $(X', Y') = (X, Y)$.

It follows from the fact that w does not satisfy Conditions (i)–(iii) that procedure defines a coupling of $\mathcal{M}(\mathcal{I}(G))$ and that $H(X', Y') = 1$ with probability 1. The coupling in the special cases will now be described and the contribution made to $\mathbf{E}[H(X', Y') - 1]$ will be calculated.

If $w = v$ then let $(X', Y') = (X, X)$ with probability $\lambda/(1 + \lambda)$, otherwise let $(X', Y') = (Y, Y)$. Here $H(X', Y') = 0$ with probability 1.

Suppose next that w satisfies Condition (ii). Define the coupling here as follows:

- with probability $\lambda/(4(1 + \lambda))$ let $(X', Y') = (Y \cup \{w\}, Y \cup \{w\})$,
- with probability $3\lambda/(4(1 + \lambda))$ let $(X', Y') = (X, Y \cup \{w\})$,
- otherwise let $(X', Y') = (X, Y)$.

Here $H(X', Y') = 0$ with probability $\lambda/(4(1 + \lambda))$ and $H(X', Y') = 2$ with probability $3\lambda/(4(1 + \lambda))$, otherwise $H(X', Y') = 1$.

Finally suppose that w satisfies Condition (iii). If u is the unique neighbour of w which is an element of Y then let $(X', Y') = (X, (Y \cup \{w\}) \setminus \{u\})$ with probability $\lambda/(4(1 + \lambda))$, otherwise let $(X', Y') = (X, Y)$. Here $H(X', Y') = 3$ with probability $\lambda/(4(1 + \lambda))$ and $H(X', Y') = 1$ otherwise.

Let d' be the number of neighbours w of v which have neighbours in the independent set Y . Then $d - d'$ elements of V satisfy Condition (ii) and at most d' elements of V satisfy Condition (iii). Combining these calculations, we obtain

$$\begin{aligned} \mathbf{E} [H(X', Y') - 1] &\leq \frac{1}{n} \left(\frac{(d - d')\lambda}{2(1 + \lambda)} + \frac{d'\lambda}{2(1 + \lambda)} - 1 \right) \\ &= \frac{1}{n} \left(\frac{d\lambda}{2(1 + \lambda)} - 1 \right) \\ &\leq \frac{1}{n} \left(\frac{\Delta\lambda}{2(1 + \lambda)} - 1 \right). \end{aligned} \tag{2}$$

Let β be defined by

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

Then (2) states that $\mathbf{E} [H(X', Y')] \leq \beta$. Now $\beta \leq 1$ whenever $\lambda \leq 2/(\Delta - 2)$. When $\lambda < 2/(\Delta - 2)$ the chain $\mathcal{M}(\mathcal{I}(G))$ is rapidly mixing with mixing time $\tau(\varepsilon)$ given by

$$\tau(\varepsilon) \leq \frac{2(1 + \lambda)}{2 - (\Delta - 2)\lambda} n \log(n\varepsilon^{-1}),$$

by Theorem 1.1. When $\lambda = 2/(\Delta - 2)$ we must estimate the probability that the Hamming distance changes in any given step. Rather than apply the second part of Theorem 1.1 directly, we will derive an improved bound. Suppose that $H(X, Y) = i$. Then there exist i vertices v such that $v \in X \setminus Y$ or $v \in Y \setminus X$. With probability $1/(1 + \lambda)$ we will delete v from one independent set, decreasing the Hamming distance between X and Y . Therefore the probability that the Hamming distance changes is at least $i/(n(1 + \lambda))$. We estimate the expected time that it takes a coupling started at (X, Y) to decrease the Hamming distance to $i - 1$. By [19, Lemma 4], the expected time to reach Hamming distance $i - 1$ is at most

$$\frac{n(1 + \lambda)}{i} (2(n - i + 1) - 1) < 2n^2(1 + \lambda)/i.$$

Therefore the expected time to couple is bounded above by

$$2n^2(1 + \lambda) \sum_{i=1}^n \frac{1}{i}$$

which is less than $2n^2(1 + \lambda)(\log(n) + 1)$. It follows using Markov's inequality that the probability that we have not coupled after $T = \lceil 2n^2 e(1 + \lambda)(\log(n) + 1) \rceil$ steps is at most e^{-1} . If we perform s independent coupling trials of T steps then the probability that we have not coupled at the end of these sT steps is at most e^{-s} . It follows that the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil 2n^2 e(1 + \lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil,$$

as stated. □

Remark 3.1 The probability that $\mathcal{M}(\mathcal{I}(G))$ performs a given drag move is $\lambda/(4(1+\lambda))$. We could instead have defined a family of Markov chains by letting this drag probability be $p = p(\lambda)$. However, it is easy to show that our chosen value of p minimises the upper bound on the mixing time, for all values of λ .

Remark 3.2 Like the Luby–Vigoda chain, the new chain gives an efficient algorithm for approximately counting independent sets in graphs with maximum degree at most four. The upper bound for the mixing time for this task is $\Omega(n)$ times smaller than that given in [21] if $\Delta = 4$, and $\Omega(n^2)$ times smaller if $\Delta = 3$.

3.1 Comparing upper bounds on mixing times

We conclude this section by comparing our improved upper bounds for the mixing time of the Luby–Vigoda chain $\mathcal{LV}(\mathcal{I}(G))$ and the upper bounds for the mixing time of the new chain $\mathcal{M}(\mathcal{I}(G))$, as given in Theorems 2.3 and 3.1. In both cases, the upper bound given is the best known upper bound for the mixing time of the chain. First let us make a general definition.

Definition 3.1 Let \mathcal{M}_i be a Markov chain on a state space Ω with mixing time $\tau_i(\varepsilon)$, and suppose that $B_i(\varepsilon)$ is the best known upper bound on $\tau_i(\varepsilon)$ for $i = 1, 2$. If $B_1(\varepsilon) < B_2(\varepsilon)$ then \mathcal{M}_1 is said to be UB-superior to \mathcal{M}_2 (here “UB” stands for “upper bound”).

Suppose that \mathcal{M}_1 is UB-superior to \mathcal{M}_2 . Of course, it does not necessarily follow that $\tau_1(\mathcal{M}_1) < \tau_2(\mathcal{M}_2)$, since the upper bound $B_2(\varepsilon)$ could be far too high. However, in a Markov chain Monte Carlo setting, the Markov chains \mathcal{M}_i are used to perform approximate sampling from the underlying state space Ω . In order to ensure that the final state is drawn from a distribution which is within ε of stationary in total variation distance, the Markov chain \mathcal{M}_i must be simulated for $B_i(\varepsilon)$ steps, as this is the best known upper bound on the mixing time. Thus the approximate sampling algorithm based on \mathcal{M}_1 is faster than the approximate sampling algorithm based on \mathcal{M}_2 (and similarly for approximate counting).

In some other contexts, the concept of UB-superiority may be less useful. For example, suppose that the Markov chain is being used in a perfect sampling algorithm such as Fill’s algorithm [14] or coupling from the past [24]. Here, the running time of the perfect sampling algorithm depends on the mixing time of the Markov chain, rather than on any upper bound for the mixing time.

Now let us return to the comparison of the upper bounds of the new Markov chain and the Luby–Vigoda chain for independent sets. Let $r(\lambda)$ be the ratio of the upper bounds, with the bound for $\tau_{LV}(\varepsilon)$ as the numerator and the bound for $\tau(\varepsilon)$ as the denominator. If $r(\lambda) > 1$ then the new chain is UB-superior, while if $r(\lambda) < 1$ then the Luby–Vigoda chain is UB-superior. For convenience we restrict ourselves to the case that G is Δ -regular. Note however that, by Remark 2.3, the comparison is also valid for nonregular graphs if $\Delta \geq 4$, $1/(2\Delta - 5) \leq \lambda \leq 1/(\Delta - 3)$ or $\Delta = 3$, $1/3 \leq \lambda \leq 1$. A summary of the results for $\lambda < 1$, $\Delta = 3$, $\lambda \leq 1$, $\Delta = 4$ and $\lambda < 1/(\Delta - 3)$, $\Delta \geq 5$ is given in Figure 3.1.

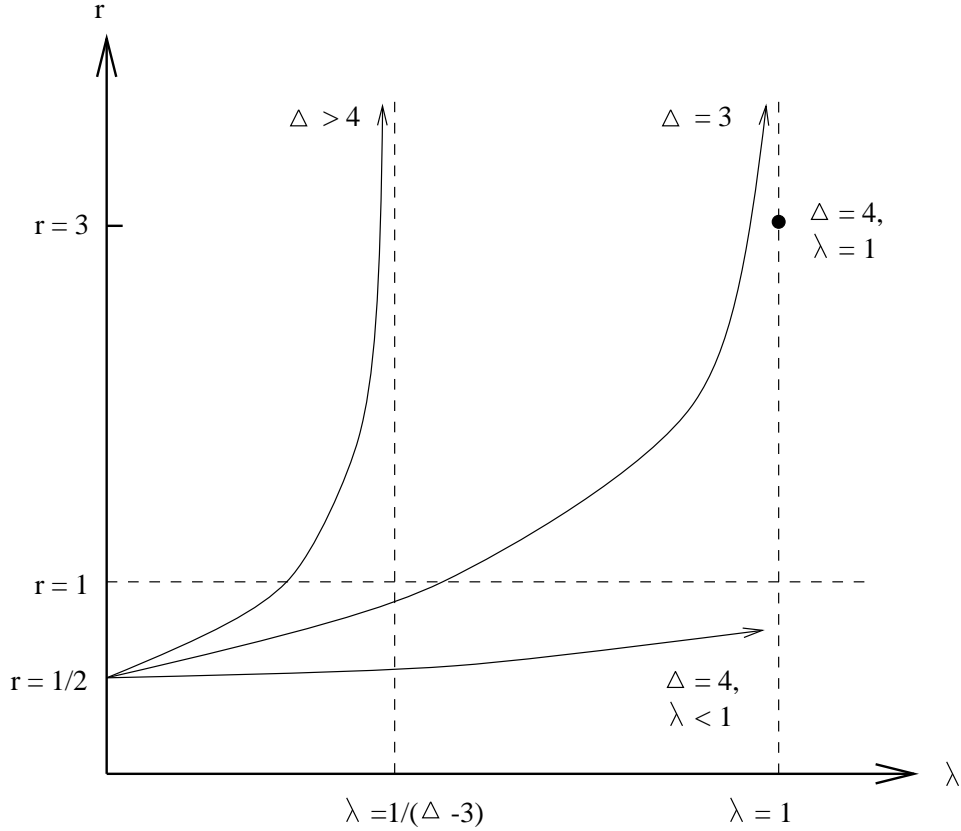


Figure 1: The ratio r of the upper bounds for the mixing times, $\mathcal{LV} : \mathcal{M}$, for various values of λ

In all cases, the ratio $r(\lambda)$ is a monotonic increasing function of λ . Hence, as λ increases, we may move from values where the Luby–Vigoda chain is UB-superior to values where the new chain is UB-superior. Moreover, $r(\lambda) \rightarrow 1/2$ as $\lambda \rightarrow 0$. This shows that the upper bound on the mixing time of the new chain is at most twice as large as the upper bound on the mixing time of the Luby–Vigoda chain. In contrast, if $\Delta \neq 4$ then $r(\lambda)$ tends to infinity as λ grows. This corresponds to situations where the upper bound on the mixing time of the Luby–Vigoda chain is arbitrarily larger than the upper bound on the mixing time of the new chain. Note that, in applications, it is generally the larger values of λ which are of more interest.

First suppose that $\Delta = 3$ and $\lambda < 1$. Then the ratio $r(\lambda)$ is given by

$$r(\lambda) = \frac{(1 + 2\lambda)(2 - \lambda)}{4(1 - \lambda)(1 + \lambda)} = \frac{2 + 3\lambda - 2\lambda^2}{4 - 4\lambda^2}.$$

The Luby–Vigoda chain is UB-superior if $0 < \lambda < 1/2$, and the new chain is UB-superior if $1/2 < \lambda < 1$. As λ tends to 1 the ratio tends to infinity. For example, when $\lambda = 1/4$ the ratio is $7/10$, when $\lambda = 3/4$ the ratio is $25/14$ and when $\lambda = 99/100$ the ratio is

15049/398. When $\lambda = 1$ the upper bound of the Luby–Vigoda chain is $\Omega^*(n)$ times bigger than the upper bound for the new chain. When $1 < \lambda \leq 2$ the new chain is rapidly mixing but it is not known whether the Luby–Vigoda chain is rapidly mixing.

Next consider the case $\Delta = 4$. If $\lambda < 1$ then the the ratio $r(\lambda)$ is given by

$$r(\lambda) = \frac{1 + 2\lambda}{2 + 2\lambda},$$

which is less than 1 for all values of $\lambda < 1$. Therefore the Luby–Vigoda chain is UB-superior when $\Delta = 4$ for all $\lambda < 1$. The ratio tends to $3/4$ as λ tends to 1. When $\lambda = 1$ the ratio $r(\lambda)$ is given by

$$r(\lambda) = \frac{\lceil 12n^2 e (\log(n) + 1) \rceil}{\lceil 4n^2 e (\log(n) + 1) \rceil},$$

which is approximately equal to three for large values of n . This suggests that the upper bound on the mixing time of the Luby–Vigoda chain is roughly three times as large as the upper bound on the mixing time of the new chain in the boundary case $\lambda = 1$.

Finally suppose that $\Delta \geq 5$. If $\lambda < 1/(\Delta - 3)$ then the ratio $r(\lambda)$ of the upper bounds satisfies

$$r(\lambda) = \frac{(1 + 2\lambda)(2 - (\Delta - 2)\lambda)}{4(1 + \lambda)(1 - (\Delta - 3)\lambda)}.$$

It is not difficult to check that $r(\lambda) < 1$ when $\lambda = 2/(3\Delta - 8)$, and $r(\lambda) > 1$ when $\lambda = 2/(3\Delta - 10)$. Therefore there exists λ_0 such that $2/(3\Delta - 8) < \lambda_0 < 2/(3\Delta - 10)$, the Luby–Vigoda chain is UB-superior for $0 < \lambda < \lambda_0$ and the new chain is UB-superior for $\lambda_0 < \lambda < 1/(\Delta - 3)$. Note that the latter interval is nonempty, as $\Delta \geq 5$. As λ tends to $1/(\Delta - 3)$ the ratio tends to infinity. For example, suppose that $\Delta = 8$. When $\lambda = 1/9$ the ratio r equals $33/40$, when $\lambda = 1/7$ the ratio is $9/8$ and when $\lambda = 24/125$ the ratio is $9169/1490$. If $\lambda = 1/(\Delta - 3)$ then the upper bound on the mixing time of the Luby–Vigoda chain is $\Omega^*(n)$ times larger than the upper bound on the mixing time of the new chain. When $1/(\Delta - 3) < \lambda \leq 2/(\Delta - 2)$ the chain $\mathcal{M}(\mathcal{I}(G))$ is known to be rapidly mixing but the Luby–Vigoda chain has no direct proof of rapid mixing.

4 A Markov chain for independent sets in hypergraphs

In this section it is shown that the Markov chain $\mathcal{M}(\mathcal{I}(G))$ can be easily extended to operate on the set of all independent sets of a hypergraph. As we shall see, this gives an efficient method for approximately counting the number of independent sets of two classes of hypergraphs: those with maximum degree two and those with maximum degree three, maximum edge size three.

A *hypergraph* $G = (V, E)$ consists of a vertex set V and a set E of *edges*, where each edge is a subset of V . We will insist that every edge contains at least 2 elements. Let

$$m = \max \{|e| : e \in E\},$$

the size of the largest edge. Then the hypergraph G is a graph if and only if $m = 2$. Say that w is a *neighbour* of v if there exists an edge which contains both v and w . We use the *degree* function d defined by

$$d(v) = |\{e \in E : v \in e\}|$$

for all $v \in V$. Let the maximum degree Δ of G be defined by

$$\Delta = \max \{d(v) : v \in V\}.$$

An *independent set* in the hypergraph G is a subset of the vertex set, no subset of which is an edge in E . Consider a map $X : V \rightarrow \{0, 1\}$. An edge e is said to be a *flaw* in X if and only if $X(v) = 1$ for all $v \in e$. Let $f(X)$ be the number of flaws in the map X . An independent set corresponds to a map $X : V \rightarrow \{0, 1\}$ with no flaws, where the vertex v belongs to the independent set if and only if $X(v) = 1$. Let $\mathcal{I}(G)$ be the set of all independent sets in G , defined by

$$\mathcal{I}(G) = \{X : V \rightarrow \{0, 1\} : f(X) = 0\}.$$

If $v \in e$ and $X(v) = 0$ but $X(u) = 1$ for all $u \in e \setminus \{v\}$, then v is said to be *critical* for the edge e in X . In this situation e is said to be a *critical edge*.

Let λ be a positive parameter. Just as in the graph case, we use the distribution π on $\mathcal{I}(G)$ where $\pi(X)$ is proportional to $\lambda^{|X|}$. We now give a definition of the Markov chain $\mathcal{M}(\mathcal{I}(G))$ which agrees with the chain described in Section 3 whenever G is a graph. For ease of notation, let $p = (m - 1)\lambda / (2m(\lambda + 1))$. If the chain is in state X_t at time t , the next state X_{t+1} is determined according to the following procedure:

- (i) choose a vertex $v \in V$ uniformly at random,
- (ii)
 - if $v \in X_t$ then let $X_{t+1} = X_t \setminus \{v\}$ with probability $1/(\lambda + 1)$,
 - if $v \notin X_t$ and v is not critical in X_t for any edge then let $X_{t+1} = X_t \cup \{v\}$ with probability $\lambda/(1 + \lambda)$,
 - if $v \notin X_t$ and v is critical in X_t for a *unique* edge e then with probability p choose $w \in e \setminus \{v\}$ uniformly at random and let $X_{t+1} = (X_t \cup \{v\}) \setminus \{w\}$,
 - otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic. Let P be the transition matrix of $\mathcal{M}(\mathcal{I}(G))$. If $Y = X \cup \{v\}$ then $\lambda P(X, Y) = P(Y, X)$ and hence $\pi(X)P(X, Y) = \pi(Y)P(Y, X)$. Suppose that $v \notin X$ and v is critical in X for a unique edge e . Let $w \in e \setminus \{v\}$ and let $Y = (X \cup \{v\}) \setminus \{w\}$. Then $P(X, Y) = p/(|e| - 1)$. Now w is critical for e in Y . The fact that X has no flaws implies that w is not critical for any other edge in Y . Therefore $P(Y, X) = P(X, Y)$ and so $\pi(X)P(X, Y) = \pi(Y)P(Y, X)$. These observations show that $\mathcal{M}(\mathcal{I}(G))$ is reversible with respect to π , so π is the stationary distribution of $\mathcal{M}(\mathcal{I}(G))$.

The following theorem proves that the Markov chain $\mathcal{M}(\mathcal{I}(G))$ is rapidly mixing for

$$\lambda \leq m / ((m - 1)\Delta - m),$$

using the path coupling method. When $m = 2$ both the chain, and the result, are identical to those of Theorem 3.1.

Theorem 4.1 *Let G be a hypergraph with maximum degree Δ and maximum edge size m . The Markov chain \mathcal{M} is rapidly mixing for $\lambda \leq m/((m-1)\Delta - m)$. When $\lambda < m/((m-1)\Delta - m)$ the mixing time $\tau(\varepsilon)$ satisfies*

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

When $\lambda = m/((m-1)\Delta - m)$ the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil 2n^2 e(1+\lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

Proof. As in Theorem 3.1, we can couple the pair (X_t, Y_t) of independent sets along a path of length $H(X_t, Y_t)$, where consecutive elements of the path are at Hamming distance 1 apart. Let X and Y be independent sets which differ just at a vertex v with degree d . Without loss of generality, assume that $v \in X \setminus Y$. We now define the coupling on (X, Y) . Choose a vertex $w \in V$ uniformly at random. Suppose first that w *does not satisfy* any of the following conditions:

- (i) $w = v$,
- (ii) w is a neighbour of v and there exists an edge e such that $\{v, w\} \subseteq e$ and e is the *only* edge for which w is critical in X ,
- (iii) w is a neighbour of v which is critical in X for more than one edge but is critical in Y for a *unique* edge e .

Then perform the following procedure:

- if $w \in X$ then let $(X', Y') = (X \setminus \{w\}, Y \setminus \{w\})$ with probability $1/(\lambda + 1)$,
- if $w \notin X$ and w is not critical in X for any edge then let

$$(X', Y') = (X \cup \{w\}, Y \cup \{w\})$$

with probability $\lambda/(1 + \lambda)$,

- if $w \notin X$ and w is critical in X for a *unique* edge e then with probability p choose $u \in e \setminus \{w\}$ uniformly at random and let

$$(X', Y') = ((X \cup \{w\}) \setminus \{u\}, (Y \cup \{w\}) \setminus \{u\}),$$

- otherwise let $(X', Y') = (X, Y)$.

It follows from the fact that w does not satisfy any of Conditions (i)–(iii) that this procedure is indeed a coupling for $\mathcal{M}(\mathcal{I}(G))$ and that $H(X', Y') = 1$ with probability 1. The coupling in the special cases will now be described and the contribution made to $\mathbf{E}[H(X', Y') - 1]$ will be calculated.

If $w = v$ then let $(X', Y') = (X, X)$ with probability $\lambda/(1 + \lambda)$, otherwise let $(X', Y') = (Y, Y)$. Here $H(X', Y') = 0$ with probability 1.

Suppose next that w satisfies Condition (ii). Define the coupling here as follows:

- with probability p choose $u \in e \setminus \{w\}$ uniformly at random and let

$$(X', Y') = ((X \cup \{w\}) \setminus \{u\}, Y \cup \{w\}),$$

- with probability $\lambda/(1 + \lambda) - p$ let $(X', Y') = (X, Y \cup \{w\})$,
- otherwise let $(X', Y') = (X, Y)$.

Now $H(X', Y') = 0$ with probability $p/(|e| - 1)$, as this occurs if and only if v is the element of $e \setminus \{w\}$ chosen uniformly at random. For all other elements of $e \setminus \{w\}$ we obtain $H(X', Y') = 2$. With probability $\lambda/(1 + \lambda) - p$ we let $(X', Y') = (X, Y \cup \{w\})$ giving $H(X', Y') = 2$. With probability $1/(1 + \lambda)$ there is no change in the Hamming distance. Therefore, in this case,

$$\begin{aligned} \mathbf{E} [H(X', Y') - 1] &= -\frac{p}{(|e| - 1)} + \frac{p(|e| - 2)}{(|e| - 1)} + \left(\frac{\lambda}{\lambda + 1} - p \right) \\ &= \frac{\lambda}{1 + \lambda} - \frac{2p}{(|e| - 1)} \\ &\leq \frac{\lambda}{1 + \lambda} - \frac{2p}{(m - 1)} \\ &= 2p, \end{aligned}$$

using the definitions of m, p .

Finally suppose that w satisfies Condition (iii). Then w is critical in X for at least one edge containing v and for *exactly* one edge not containing v . With probability p choose $u \in e \setminus \{w\}$ uniformly at random and let

$$(X', Y') = (X, (Y \cup \{w\}) \setminus \{u\}),$$

otherwise let $(X', Y') = (X, Y)$. Since $v \notin e$ it follows that $H(X, (Y \cup \{w\}) \setminus \{u\}) = 3$ for all $u \in e \setminus \{w\}$. Hence $H(X', Y') = 3$ with probability p and $H(X', Y') = 1$ otherwise.

Let d' be the number of vertices $w \in V$ which satisfy Condition (ii) and let d'' be the number of vertices $w \in V$ which satisfy Condition (iii). Now any vertex w which satisfies Condition (ii) or (iii) is critical in X for some edge e which contains v . Moreover, each such edge can contribute at most one such vertex w . It follows that $d' + d'' \leq d$. Combining these calculations we obtain

$$\begin{aligned} \mathbf{E} [H(X', Y') - 1] &\leq \frac{1}{n} (2pd' + 2pd'' - 1) \\ &\leq \frac{1}{n} (2pd - 1) \\ &\leq \frac{1}{n} (2p\Delta - 1) \\ &= \frac{1}{n} \left(\frac{\Delta(m - 1)\lambda}{m(\lambda + 1)} - 1 \right). \end{aligned} \tag{3}$$

Let β be defined by

$$\beta = 1 - \frac{m - ((m - 1)\Delta - m)\lambda}{nm(1 + \lambda)}.$$

Then (3) can be rearranged to give $\mathbf{E}[H(X', Y')] \leq \beta$. Now $\beta \leq 1$ whenever $\lambda \leq m / ((m-1)\Delta - m)$. When $\lambda < m / ((m-1)\Delta - m)$ the chain \mathcal{M} is rapidly mixing with mixing time $\tau(\varepsilon)$ given by

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

by the first part of Theorem 1.1. When $\lambda = m / ((m-1)\Delta - m)$ we proceed as in the proof of Theorem 3.1. Suppose that $H(X, Y) = i$ for two independent sets X, Y . Then for i choices of vertices we can ensure that $H(X', Y') = i - 1$ with probability at least $1/(1+\lambda)$. This implies that the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil 2n^2 e(1+\lambda)(\log(n) + 1) \rceil \lceil \log(\varepsilon^{-1}) \rceil,$$

as stated. □

We have just shown that the Markov chain $\mathcal{M}(\mathcal{I}(G))$, when extended to act on hypergraphs, is rapidly mixing for $\lambda \leq m / ((m-1)\Delta - m)$. Hence this chain gives an efficient algorithm for approximately counting independent sets whenever $(m-1)\Delta \leq 2m$; that is, for

- (i) $\Delta \in \{3, 4\}$, $m = 2$,
- (ii) $\Delta = 2$, $m \geq 2$,
- (iii) $\Delta = 3$, $m = 3$.

In (i) we find graphs with maximum degree 4, as mentioned in Remark 3.2. We discuss the remaining cases below.

First consider a hypergraph $H = (V, E)$ with maximum degree two and maximum edge size m . Let $n = |V|$ and $q = |E|$. Then $q \leq n$. Now H is the dual hypergraph of a graph $G = (V_G, E_G)$ with maximum degree m , where $|V_G| = q$ and $|E_G| = n$. (The *dual* hypergraph of a hypergraph is obtained by transposing the incidence matrix, i.e., the (0,1)-matrix with rows corresponding to vertices, columns corresponding to edges and a 1 entry in the (v, e) position if and only if $v \in e$.) An independent set in H corresponds to a subset $X \subseteq E_G$ which satisfies the following condition: for every vertex $v \in V_G$ there exists an edge $e \in E_G$ such that $v \in e$ and $e \notin X$. The complement $E_G \setminus X$ of X is an *edge cover* of the graph G . Counting edge coverings in graphs is a $\#P$ -complete problem (for a description of the decision version of the problem see [15] and for a sketch of the proof of $\#P$ -completeness see [2]). In [2] a rapidly mixing Markov chain for a generalisation of this problem is described. The mixing time of this chain is $O(n(n^2 + q^3) \log(\varepsilon^{-1}))$, while the mixing time of $\mathcal{M}(\mathcal{I}(G))$ is bounded above by $mn \log(n\varepsilon^{-1})$. Therefore approximately counting edge coverings in a graph can be done $\Omega^*(n^2)$ times faster using a Markov chain Monte Carlo algorithm based on the Markov chain $\mathcal{M}(\mathcal{I}(G))$ (see Section 3.1).

Now consider the problem of counting independent sets in hypergraphs with maximum degree 3 and maximum edge size 3. We now sketch a proof that counting independent sets in graphs of maximum degree 3 is $\#P$ -complete (for full details see [16, 11]).

Theorem 4.2 Let $\#\text{INDEPENDENT-SETS}(3)$ denote the following counting problem. An instance is a graph with maximum degree three and the problem is to calculate the number of independent sets in that graph. Then $\#\text{INDEPENDENT-SETS}(3)$ is $\#P$ -complete.

Sketch of proof. The problem of counting the number of independent sets in graphs with maximum degree Δ is $\#P$ -complete for $\Delta \geq 4$, as follows from [29] or [26, Theorem 3.1]. We give a polynomial-time reduction from this problem to $\#\text{INDEPENDENT-SETS}(3)$. Let $G = (V_G, E_G)$ be a graph with maximum degree Δ where $\Delta \geq 3$. We construct a graph $H = (V, E)$ with maximum degree three and edge bipartition $E = E' \cup E''$ by replacing every vertex $v \in V_G$ of degree greater than three by a chain of new vertices v_1, \dots, v_{d-2} , each of degree three. Each neighbour of v in G is joined to exactly one v_j by an edge in E' , while the edge $\{v_j, v_{j+1}\}$ belongs to E'' for $1 \leq j < d - 2$. All other edges in G are placed in E' . Clearly H can be formed from G in polynomial time. For $0 \leq t \leq m$, $0 \leq i \leq t$ let $n_{t,i}$ denote the number of subsets X of V such that

- (i) if $e \in E'$ then $e \not\subseteq X$,
- (ii) $|\{e \in E'' : e \cap X = \emptyset\}| = i$,
- (iii) $|\{e \in E'' : e \subseteq X\}| = t - i$.

It is not difficult to see that $|\mathcal{I}(G)| = \sum_{i=0}^m n_{m,i}$.

Let $m = |E''|$. Denote by C_r denote the chain graph with r edges, where $r \geq 1$. Let f_r denote the r th Fibonacci number, defined inductively by $f_1 = 1$, $f_2 = 1$ and $f_{r+2} = f_r + f_{r+1}$ for $r \geq 1$. Then $|\mathcal{I}(C_r)| = f_{r+3}$. For $1 \leq r \leq m + 1$ for the graph H_r , which is the r -stretch of H with respect to E'' . That is, form H_r from H by replacing every edge in E'' by a copy of C_r . The family of graphs $\{H_r : 1 \leq r \leq m + 1\}$ can be formed from H in polynomial time. By counting the number of independent sets in C_r which contain both, exactly one, or neither of the endpoints, we obtain an equation of the form

$$|I(H_r)| = \sum_{t=0}^m \sum_{i=0}^t n_{t,i} f_{r+1}^i f_{r-1}^{t-i} f_r^{m-t}.$$

This can be written as a polynomial in x_r , where $x_r = f_r/f_{r+1}$. Since these quotients are distinct (see for example [30, Lemma A.1]), standard interpolation techniques ensure that the coefficients of this polynomial can be found in polynomial time. Therefore we can evaluate the polynomial at $x_r = 0$, giving the value

$$\sum_{i=0}^m n_{m,i} = |\mathcal{I}(G)|.$$

This completes the polynomial-time reduction from $\#\text{INDEPENDENT-SETS}(4)$ to $\#\text{INDEPENDENT-SETS}(3)$, proving that the latter problem is $\#P$ -complete. \square

The set of independent sets in a given hypergraph corresponds in a natural way to the set of satisfying assignments of a certain SAT instance (for background on SAT see, for example, [23, Section 4.2]). A SAT instance is said to be *monotone* if no variable

appears negated. Using notation borrowed from Roth [26], denote by $l\mu$ - k MON the class of all monotone SAT instances where each variable appears in at most l clauses and each clause contains at most k literals. Let G be a hypergraph where all edges have at most m vertices and all vertices appear in at most Δ edges. Let each vertex correspond to a variable (denoted by the same symbol) and for each $e \in E$ form a clause C_e as follows. If $e = \{v_1, \dots, v_d\}$ then let

$$C_e = \bigvee_{i=1}^d \bar{v}_i.$$

Then

$$\bigwedge_{e \in E} C_e \tag{4}$$

is a SAT instance with $|\mathcal{I}(G)|$ satisfying assignments. Replace the variable \bar{v}_i by w_i for each variable v_i . This gives an instance of $\Delta\mu$ - m MON. The above construction can be reversed, showing that the problem of finding the number of independent sets in hypergraphs with maximum degree at most Δ and maximum edge size at most m is equivalent to the counting problem $\#\Delta\mu$ - m MON. Thus Theorem 4.2 shows that the counting problem $\#3\mu$ -2MON is $\#P$ -complete. It follows immediately that $\#3\mu$ -3MON is also $\#P$ -complete. As far as we are aware, the Markov chain $\mathcal{M}(\mathcal{I}(G))$ is the first which can be used to approximately count the number of satisfying assignments of a $\#3\mu$ -3MON instance.

Remark 4.1 It is not difficult to further show that the problem of counting independent sets in 3-regular graphs is $\#P$ -complete. Moreover, it is possible to show that the problem of counting independent sets in hypergraphs with maximum degree three is $\#P$ -complete, even when every edge has size three. This disproves the (admittedly rather unlikely) conjecture that the counting problem $\#3\mu$ -3MON is $\#P$ -complete because of the possible presence of edges of size two. For full details of these proofs see [16, 11].

5 Comparing Markov chains with the same stationary distribution

As remarked in the introduction, there are relatively few methods available for investigating the mixing time of Markov chains. In this paper we have used path coupling. Some alternative approaches have been developed which involve the eigenvalues of the Markov chain (see for example [10, 28]). Let β_1 denote the second eigenvalue of the Markov chain \mathcal{M} . There is a well-known inequality relating the mixing time of a Markov chain and the so called *spectral gap* $1 - \beta_1$ of the chain. The difficulty in applying this inequality lies in obtaining a bound for the spectral gap of \mathcal{M} . The bound on the mixing time obtained in this manner involves a factor related to the stationary distribution π of \mathcal{M} . This often results in a lack of tightness in the bound on the mixing time.

In this section we consider two Markov chains with the same state space and stationary distribution, one of which is known to be rapidly mixing. For the first result we assume that a certain linear relationship exists between the entries of the transition matrices of these chains. We prove that this implies a linear relationship between the

spectral gaps of the chains. In the second result the mixing rate of a Markov chain is used to provide an upper bound for the second eigenvalue of the chain. Combining these results, we show that the mixing time of one of the chains can be bounded in terms of the mixing time of the other using the relationship between the spectral gaps. In certain circumstances this allows us to deduce that the other chain is also rapidly mixing. The bound obtained on the mixing time of the other chain will be larger than the bound for the mixing time of the first chain. This is partly due to the factor related to π incurred through the use of the spectral gap bound, as mentioned above. At the end of the section we apply our method to two Markov chains of independent sets: the simple insert/delete chain described in Section 2 and the new chain $\mathcal{M}(\mathcal{I}(G))$ defined in Section 3.

Our approach is quite simple, but gives similar results to those in [7, 8, 9, 25], which provide alternative methods for relating the spectral gaps of two Markov chains. We consider only Markov chains with the same stationary distribution, but in this case, the conditions of [8, Theorem 2.1] and [9, Lemma 3.3] are met whenever our requirement holds; that is, when the transition matrices of the two chains are linearly related. While our approach may yield worse constants than those produced by [8, Theorem 2.1] or [9, Lemma 3.3], it has the advantage of being extremely simple to apply. In [25] an alternative method of bounding the mixing time of one chain, using the relationship between the spectral gaps of two chains, is given. The method we present here improves the bound on the mixing time by a constant, assuming the same value of a certain parameter.

First we introduce some notation. Let \mathcal{M} be an irreducible Markov chain with transition matrix P and state space Ω . Assume that P is reversible with respect to some probability distribution π such that $\pi(x) > 0$ for all $x \in \Omega$. The eigenvalues $\beta_0, \dots, \beta_{N-1}$ of P are real numbers which satisfy

$$1 = \beta_0 > \beta_1 \geq \dots \geq \beta_{N-1} \geq -1,$$

where $N = |\Omega|$ (see, for example, [8]). For convenience, we can assume that all the eigenvalues are nonnegative by adding a holding probability of $1/2$ to every state of \mathcal{M} . This amounts to replacing P by $(I + P)/2$ and will double the mixing time. For the remainder of the section we assume that all the eigenvalues are nonnegative. The second largest eigenvalue β_1 appears in the following well-known bound on the variation distance of \mathcal{M} at time t with initial state x :

$$d_{\text{TV}}(P^t_x, \pi) \leq \frac{1}{2} \sqrt{\frac{1 - \pi(x)}{\pi(x)}} \beta_1^t.$$

For a proof see, for example, [10, Proposition 3]. It follows that the mixing time $\tau(\varepsilon)$ of \mathcal{M} satisfies

$$\tau(\varepsilon) \leq \frac{\log(\pi^*(2\varepsilon)^{-1})}{1 - \beta_1} \tag{5}$$

where

$$\pi^* = \max \left\{ \sqrt{\frac{1 - \pi(x)}{\pi(x)}} : x \in \Omega \right\}. \tag{6}$$

It follows from (5) that \mathcal{M} is rapidly mixing if both $\log(\pi^*)$ and $(1 - \beta_1)^{-1}$ are bounded above by polynomials in n , where n is a measure of the size of the elements of Ω . This result is of theoretical importance, but of little practical value for a given chain unless an upper bound on the spectral gap of \mathcal{M} can be obtained.

Suppose now that \mathcal{M}_1 and \mathcal{M}_2 are two Markov chains on the same state space Ω which are reversible with respect to the same stationary distribution π . The first result of this section shows how a linear relationship between the entries of the transition matrices of \mathcal{M}_1 and \mathcal{M}_2 gives rise to a linear relationship between the spectral gap of the two chains. *Note added:* Since submitting this paper, we have learnt that this result is implied by results of Caracciolo, Pelissetto and Sokal [5, Theorems A.1, A.3].

Theorem 5.1 *Let \mathcal{M}_1 and \mathcal{M}_2 be Markov chains with state space Ω and stationary distribution π . Let P_i denote the transition matrix of \mathcal{M}_i and let $\beta_1(P_i)$ be the second eigenvalue of P_i , for $i = 1, 2$. Suppose that there exists $\alpha \geq 0$ such that $P_1 \geq \alpha P_2$. Then*

$$(1 - \beta_1(P_1)) \geq \alpha (1 - \beta_1(P_2)). \quad (7)$$

Proof. Define the matrix P_3 by

$$P_3 = \frac{P_1 - \alpha P_2}{1 - \alpha}.$$

Then every entry in P_3 is nonnegative and $P_3 e = e$, where e is the column vector with every entry equal to 1. Therefore P_3 is the transition matrix of a Markov chain on Ω . Moreover P_3 is reversible with respect to π . Let D be the diagonal matrix with (j, j) th entry equal to $\sqrt{\pi_j}$ for $1 \leq j \leq |\Omega|$. Let $R_i = D P_i D^{-1}$ for $1 \leq i \leq 3$. Since P_i is reversible with respect to π the matrix R_i is symmetric for $1 \leq i \leq 3$. Also $\beta_1(R_i) = \beta_1(P_i)$ for $1 \leq i \leq 3$. The usual min-max characterisation of eigenvalues of symmetric matrices states that

$$\beta_1(R_i) = \max \{x^t R_i x : x \cdot e = 0, \|x\|_2 = 1\}$$

for $1 \leq i \leq 3$ (see for example, [6, Theorem 1.9.1]). Since $R_1 = (1 - \alpha)R_3 + \alpha R_2$ it follows that

$$\begin{aligned} \beta_1(R_1) &= \max \{(1 - \alpha)x^t R_3 x + \alpha x^t R_2 x : x \cdot e = 0, \|x\|_2 = 1\} \\ &\leq (1 - \alpha) \max \{x^t R_3 x : x \cdot e = 0, \|x\|_2 = 1\} \\ &\quad + \alpha \max \{x^t R_2 x : x \cdot e = 0, \|x\|_2 = 1\} \\ &= (1 - \alpha)\beta_1(R_3) + \alpha\beta_1(R_2) \\ &\leq 1 - \alpha + \alpha\beta_1(R_2). \end{aligned}$$

This proves the theorem. □

Remark 5.1 Suppose that \mathcal{M}_1 and \mathcal{M}_2 are two Markov chains which satisfy the conditions of Theorem 5.1. Then \mathcal{M}_1 will perform the same transition as \mathcal{M}_2 with probability at least α , irrespective of the current state. This provides some intuition as to why rapid mixing of \mathcal{M}_2 may in certain circumstances imply that \mathcal{M}_1 is also rapidly mixing.

The second result of this section is a straightforward calculation, showing how the second eigenvalue of a Markov chain may be bounded above using its mixing time.

Theorem 5.2 *Let \mathcal{M} be an irreducible Markov chain with mixing time $\tau(\varepsilon)$ and second eigenvalue β_1 . Then $\beta_1 \leq e^{-1/T}$ where $T = \tau(e^{-1})$.*

Proof. Consider a maximal coupling of \mathcal{M} with an arbitrary pair of initial states. After performing T steps of the coupling the probability that the process has not coupled is at most e^{-1} . If we run s independent coupling trials of length T then the probability that we have not coupled after these sT steps is at most e^{-s} . By the Coupling Lemma it follows that

$$d_{\text{TV}}(P^t_x, \pi) \leq \left(e^{-1/T}\right)^t \quad (8)$$

for $t \in \{T, 2T, 3T, \dots\}$ and for all initial states x . We now construct a particular initial distribution μ . Let v be a left eigenvector of β_1 such that $\|v\|_2 = 1$, and let e be the column vector with each entry equal to 1. Now e is a right eigenvector corresponding to eigenvalue β_0 . It follows that $v \cdot e = 0$. Let $\sigma = \min\{\pi(x) : x \in \Omega\}$. Then $\sigma > 0$ as \mathcal{M} is irreducible. Finally let μ be the vector defined by $\mu = \pi - \sigma v$. Then every component of μ is nonnegative and $\mu \cdot e = \pi \cdot e = 1$. Therefore μ is a distribution and

$$d_{\text{TV}}(\mu P^t, \pi) \geq \frac{1}{2} \|\mu P^t - \pi\|_2 = \frac{1}{2} \|(\mu - \pi)P^t\|_2 = \frac{\sigma}{2} \beta_1^t.$$

It follows from (1), (8) that

$$d_{\text{TV}}(\mu P^t, \pi) \leq \left(e^{-1/T}\right)^t$$

for $t \in \{T, 2T, 3T, \dots\}$. Therefore

$$\beta_1 \leq (2/\sigma)^{1/t} e^{-1/T}$$

for these values of t . Letting t tend to infinity along this sequence we obtain $\beta_1 \leq e^{-1/T}$, as required. \square

Whenever two Markov chains $\mathcal{M}_1, \mathcal{M}_2$ satisfy the requirements of Theorem 5.1, the following theorem gives a relationship between the mixing times of the chains.

Theorem 5.3 *Let P_1, P_2 be transition matrices of irreducible Markov chains $\mathcal{M}_1, \mathcal{M}_2$ with state space Ω and stationary distribution π . Assume that all eigenvalues of \mathcal{M}_i are positive, for $i = 1, 2$. Let π^* be as defined in (6) and let $\tau_i(\varepsilon)$ denote the mixing time of \mathcal{M}_i for $i = 1, 2$. Suppose that there exists $\alpha \geq 0$ such that $P_1 \geq \alpha P_2$. Then*

$$\tau_1(\varepsilon) \leq 2\alpha^{-1} \tau_2(e^{-1}) \log(\pi^*(2\varepsilon)^{-1}). \quad (9)$$

Proof. Let $\beta_1(P_i)$ denote the second eigenvalue of P_i for $i = 1, 2$ and let $T = \tau_2(e^{-1})$. Then

$$1 - \beta_1(P_1) \geq \alpha (1 - \beta_1(P_2)) \geq \alpha (1 - e^{-1/T})$$

by Theorems 5.1, 5.2. Using (5), it follows that

$$\tau_1(\varepsilon) \leq \alpha^{-1} \frac{\log(\pi^*(2\varepsilon)^{-1})}{1 - e^{-1/T}}. \quad (10)$$

The inequality

$$e^s \left(1 - \frac{s^2}{r}\right) \leq \left(1 + \frac{s}{r}\right)^r$$

is satisfied for all $r \geq 1$ and $|s| \leq r$ (see for example [22, p. 435]). Taking $r = 2$ and $s = -1/T$ we obtain

$$e^{-1/T} \leq \frac{4T^2 - 4T + 1}{4T^2 - 2}.$$

Therefore

$$\left(1 - e^{-1/T}\right)^{-1} \leq \frac{4T^2 - 2}{4T - 3} \leq 2T.$$

Substituting this into (10) we obtain

$$\tau_1(\varepsilon) \leq 2\alpha^{-1} \tau_2(e^{-1}) \log(\pi^*(2\varepsilon)^{-1}),$$

as required. \square

The following corollary is immediate.

Corollary 5.1 *In the situation described in Theorem 5.3, let n be a measure of the size of the elements of Ω . Suppose that \mathcal{M}_2 is rapidly mixing. If α^{-1} and $\log(\pi^*)$ are each bounded above by a polynomial in n , then \mathcal{M}_1 is also rapidly mixing.*

Remark 5.2 The bound on the mixing time of \mathcal{M}_1 obtained in Theorem 5.3 may be significantly larger than the known bound for $\tau_2(\varepsilon)$. One reason for this is the α^{-1} factor. This factor is large if the probability that \mathcal{M}_1 acts like \mathcal{M}_2 is small (see Remark 5.1). Another factor involved in the bound for $\tau_1(\varepsilon)$ is $\log(\pi^*)$. If $|\Omega|$ is approximately equal to k^n for some k then $\log(\pi^*) \geq n \log(k)/2$. This increases the bound on the mixing time by at least one extra factor of n .

Remark 5.3 Let \mathcal{M}_i be two Markov chains with the same state space Ω and stationary distribution π . Suppose that (7) holds for some $\alpha > 0$. In [25, Proposition 4], Randall and Tetali give the bound

$$\tau_1(\varepsilon) \leq 4\alpha^{-1} \tau_2(\varepsilon) \log((\varepsilon\pi_*)^{-1}) / \log((2\varepsilon)^{-1}),$$

where $\pi_* = \min\{\pi(x) : x \in \Omega\}$. By setting $\varepsilon = e^{-1}$, we can compare this upper bound on $\tau_2(e^{-1})$ with the upper bound given in (9). (Note that $\tau(e^{-1})$ is sometimes taken as the definition of the mixing time of a chain.) It is not difficult to show that the ratio of the Randall–Tetali upper bound with the new upper bound is given by

$$\begin{aligned} \frac{\text{RT}}{\text{DG}} &= \frac{4\alpha^{-1} \tau_2(e^{-1}) \log(e/\pi_*)}{2\alpha^{-1} \tau_2(e^{-1}) \log(\pi^* e/2) \log(e/2)} \\ &= \frac{2(1 + \log(\pi_*^{-1}))}{(1 - \log(2))(\log(\pi^* + 1 - \log(2)))} \\ &\geq \frac{4(1 + \log(\pi_*^{-1}))}{(1 - \log(2))(\log(\pi_*^{-1}) + 2 - 2\log(2))}, \end{aligned}$$

since $\pi_* \leq (\pi^*)^2$. Assuming that $|\Omega|$ grows with n , we see that $\log(\pi_*^{-1})$ also grows with n and so dominates the constant terms. Hence this ratio tends to $4/(1 - \log(2))$ as n tends to infinity. Thus the bound given by (9) is at least 8 times better than the Randall–Tetali bound, for the same value of α , for large values of n .

We now illustrate this method with reference to the insert/delete chain described in Section 2 and the chain $\mathcal{M}(\mathcal{I}(G))$ introduced in Section 3. Both chains have state space $\mathcal{I}(G)$, the set of independent sets of a graph G . The stationary distribution π of both chains is given by $\pi(X) = \lambda^{|X|}/Z$ for some positive parameter λ . We shall not compare the chain $\mathcal{M}(\mathcal{I}(G))$ directly with the insert/delete chain. One reason for this is the fact that a transition of $\mathcal{M}(\mathcal{I}(G))$ may involve two vertices of G while each transition of the insert/delete chain involves only one vertex. Another reason is that we must ensure that all the eigenvalues of the chains are positive. Therefore we consider the two-step insert/delete chain $\mathcal{M}_1(\mathcal{I}(G))$ with a holding probability of $1/2$ at each step. A transition of $\mathcal{M}_1(\mathcal{I}(G))$ from current state X consists of performing the following procedure twice:

with probability $1/2$ do nothing, otherwise

- (i) choose $v \in V$ uniformly at random,
- (ii) let $X' = X \cup \{v\}$ with probability $\lambda/(1 + \lambda)$ otherwise let $X' = X \setminus \{v\}$,
- (iii) if $X' \in \mathcal{I}(G)$ then move to X' , else stay at X .

Similarly, let $\mathcal{M}_2(\mathcal{I}(G))$ be the Markov chain obtained from $\mathcal{M}(\mathcal{I}(G))$ by adding a holding probability of $1/2$ to every state. We prove below that the entries of the transition matrices of these chains satisfy a linear relationship.

Theorem 5.4 *Let $\mathcal{M}_i(\mathcal{I}(G))$ be the Markov chains with state space $\mathcal{I}(G)$ and stationary distribution π , as described above. Let P_i be the transition matrix of $\mathcal{M}_i(\mathcal{I}(G))$ for $i = 1, 2$. Then $P_1 \geq \alpha P_2$, where*

$$\alpha = \min \left\{ \frac{1}{2}, \frac{4}{(1 + \lambda)n} \right\}.$$

Proof. We need only verify the inequality $P_1 \geq \alpha P_2$ for nonzero entries of P_2 . Now $\mathcal{M}_1(\mathcal{I}(G))$ has a holding probability of $1/2$ at every state. Therefore

$$P_1(X, X) \geq \frac{1}{2} \geq \alpha \geq \alpha P_2(X, X)$$

for all $X \in \mathcal{I}(G)$. Suppose next that $H(X, Y) = 1$. We can assume without loss of generality that $|Y| = |X| + 1$, so $P_2(X, Y) = \lambda/(2n(1 + \lambda))$. The probability that $\mathcal{M}_1(\mathcal{I}(G))$ makes the same transition is given by the probability that the chain performs

the correct insertion in one step and does nothing in the other step. Since the order that the steps are performed is immaterial it follows that

$$P_1(X, Y) \geq \frac{\lambda}{2n(1+\lambda)} = P_2(X, Y) \geq \alpha P_2(X, Y)$$

for these pairs (X, Y) . Finally suppose that $Y = (X \cup \{v\}) \setminus \{w\}$ for some edge $\{v, w\}$. Then $P_2(X, Y) = \lambda/(8n(1+\lambda))$ and

$$P_1(X, Y) = \frac{\lambda}{2n^2(1+\lambda)^2} = \frac{4}{n(1+\lambda)} P_2(X, Y) \geq \alpha P_2(X, Y).$$

This proves the theorem. □

Since \mathcal{M}_i satisfy the requirements of Theorem 5.1, we can apply Theorem 5.3 to relate their mixing times. This implies that the insert/delete chain is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$, as described below.

Theorem 5.5 *Let G be a graph with maximum degree Δ and let λ be a positive parameter. The insert/delete chain with state space $\mathcal{I}(G)$ is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$.*

Proof. Consider the Markov chains $\mathcal{M}_i(\mathcal{I}(G))$ defined above, for $i = 1, 2$. Let π denote their stationary distribution and let α be as defined in Theorem 5.4. Let $\tau_i(\varepsilon)$ denote the mixing time of \mathcal{M}_i , for $i = 1, 2$. By Theorems 5.3 and 5.4, the mixing times are related as follows:

$$\tau_1(\varepsilon) < 2\alpha^{-1}\tau_2(e^{-1}) \log(\pi^*(2\varepsilon)^{-1}). \quad (11)$$

By definition, the Markov chain $\mathcal{M}_2(\mathcal{I}(G))$ is obtained by adding a holding probability of $1/2$ to each state of the chain $\mathcal{M}(\mathcal{I}(G))$ described in Section 3. Hence, by Theorem 3.1, the chain $\mathcal{M}_2(\mathcal{I}(G))$ is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$. The mixing time $\tau_2(\varepsilon)$ bounded above by twice the bound given in Theorem 3.1. In order to apply Corollary 5.1, we must show that both $\log(\pi^*)$ and α^{-1} are bounded above by a polynomial in n .

Now $\pi(X) = \lambda^{|X|}/Z$, where λ is a positive parameter. Therefore the quantity π^* , defined in (6), is given by

$$\pi^* = \begin{cases} \sqrt{|Z|\lambda^{-n} - 1} & \text{if } \lambda \leq 1, \\ \sqrt{|Z| - 1} & \text{if } \lambda > 1. \end{cases}$$

In both cases

$$\log(\pi^*) = \frac{n}{2}(\log(\lambda^*) + 1) \quad (12)$$

where $\lambda^* = \max\{\lambda, \lambda^{-1}\}$. This quantity is linear in n . Moreover the constant α defined in Theorem 5.4 satisfies

$$\alpha^{-1} \leq \max\{2, (1+\lambda)n/4\} \quad (13)$$

which is either constant or linear in n . Thus, by Corollary 5.1, the Markov chain \mathcal{M}_1 is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$.

By definition, the transitions of $\mathcal{M}_1(\mathcal{I}(G))$ consist of two steps of the simple insert/delete chain, with a holding probability of $1/2$ added at each step. The simple insert/delete chain (without the holding probabilities) has the same mixing time as $\mathcal{M}_1(\mathcal{I}(G))$, since the factor 2 increase incurred by adding a holding probability of $1/2$ is cancelled by taking two steps per transition. Thus the insert/delete chain is rapidly mixing for $\lambda \leq 2/(\Delta - 2)$ with mixing time $\tau_1(\varepsilon)$. \square

Let us compare this result with Theorem 2.2. We may conclude that the insert/delete chain is rapidly mixing for a wider range of λ than was previously known. In particular the insert/delete chain can be used to approximately count independent sets of graphs with maximum degree at most four. However, extending the range of λ for which rapid mixing of the insert/delete chain is known using this method comes at a cost. Ignoring constant factors, the upper bound for the mixing time of the insert/delete chain provided by Theorem 2.2 is $O(n \log(n))$ for $\lambda < 1/(\Delta - 1)$ and $O(n^2)$ for $\lambda = 1/(\Delta - 1)$. Using Theorem 3.1, it follows that the upper bound given by (11) is $O(n^3 \log(n))$ when $\lambda < 2/(\Delta - 2)$, and $O(n^4)$ when $\lambda = 2/(\Delta - 2)$. A factor of n is introduced by each of $\log(\pi^*)$ and α^{-1} , as can be seen from (12) and (13) respectively. This illustrates the warning given in Remark 5.2.

This technique has also been applied to the problem of counting graph colourings. In [4], a Markov chain was defined with state space $\Omega_k(G)$, the set of all proper k -colourings of the given graph G . Using a computer-assisted proof, this chain was shown to be rapidly mixing for $k = 5$ when G has maximum degree 3. The result was extended to show rapid mixing for $k = 7$ when G is a triangle-free 4-regular graph. Previously, other Markov chains for graph colourings were only known to be rapidly mixing for $k \geq 2\Delta$, including the simple Markov chain discovered independently by Jerrum [17] and Salas and Sokal [27]. Using the comparison technique described in this section, we can conclude that the Jerrum/Salas–Sokal chain is also rapidly mixing for $k = 5$ when G has maximum degree 3, and for $k = 7$ when G is a triangle-free 4-regular graph.

Remark 5.4 Let P_{LV} be the transition matrix of the Luby–Vigoda chain and let P_{DG} be the transition matrix of the new chain, as introduced in Section 3. It is not difficult to show that there exists a constant α , independent of n , such that $P_{LV} \geq \alpha P_{DG}$. Therefore, applying the comparison technique described in this section, we can conclude that the Luby–Vigoda chain is also rapidly mixing for $\lambda \leq 2/(\Delta - 2)$, and that the mixing time of the Luby–Vigoda chain is at most $O(n)$ times greater than the mixing time of the new chain in the range $1/(\Delta - 3) \leq \lambda \leq 2/(\Delta - 2)$. However, this $O(n)$ factor is almost certainly just an artefact of the analysis.

6 Acknowledgements

The authors would like to thank Mark Jerrum, Alan Sokal and David Wilson for their helpful comments.

References

- [1] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, in A. Dold and B. Eckmann, eds., *Séminaire de Probabilités XVII 1981/1982*, vol. 986 of *Springer-Verlag Lecture Notes in Mathematics*, Springer-Verlag, New York, 1983 pp. 243–297.
- [2] R. Bubley and M. Dyer, *Graph orientations with no sink and an approximation for a hard case of #SAT*, in *8th Annual Symposium on Discrete Algorithms*, ACM–SIAM, New York–Philadelphia, 1996 pp. 248–257.
- [3] R. Bubley and M. Dyer, *Path coupling: A technique for proving rapid mixing in Markov chains*, in *38th Annual Symposium on Foundations of Computer Science*, IEEE, Los Alimitos, 1997 pp. 223–231.
- [4] R. Bubley, M. Dyer, C. Greenhill, and M. Jerrum, *On approximately counting colourings of small degree graphs*, *SIAM Journal on Computing*, 29 (1999), pp. 387–400.
- [5] S. Caracciolo, A. Pelissetto, and A. Sokal, *Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints*, *Journal of Statistical Physics*, 60 (1990), pp. 1–53.
- [6] F. Chatelin, *Eigenvalues of matrices*, John Wiley and Sons, Chichester, (1993).
- [7] F. R. K. Chung and R. L. Graham, *Random walks on generating sets for finite groups*, *Electronic Journal of Combinatorics*, (1997), pp. 14 (electronic).
- [8] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, *The Annals of Applied Probability*, 3 (1993), pp. 696–730.
- [9] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, *The Annals of Applied Probability*, 6 (1996), pp. 695–750.
- [10] P. Diaconis and D. Strook, *Geometric bounds for eigenvalues of Markov chains*, *The Annals of Applied Probability*, 1 (1991), pp. 36–61.
- [11] M. Dyer and C. Greenhill, *Some #P-completeness proofs for colourings and independent sets*, Tech. Rep. 97.47, School of Computer Studies, University of Leeds, (1997).
- [12] M. Dyer and C. Greenhill, *A more rapidly mixing Markov chain for graph colourings*, *Random Structures and Algorithms*, 13 (1998), pp. 285–317.
- [13] 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*, IEEE, Los Alimitos, 1999 (To appear).
- [14] J. A. Fill, *An interruptible algorithm for perfect sampling via Markov chains*, *The Annals of Applied Probability*, (1998), pp. 131–162.

- [15] M. R. Garey and D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-completeness*, W. H. Freeman and Company, San Francisco, (1979).
- [16] C. Greenhill, *The complexity of counting colourings and independent sets in sparse graphs and hypergraphs*, Computational Complexity, (To appear).
- [17] M. Jerrum, *A very simple algorithm for estimating the number of k -colorings of a low-degree graph*, Random Structures and Algorithms, 7 (1995), pp. 157–165.
- [18] M. Jerrum and A. Sinclair, *The Markov chain Monte Carlo method: an approach to approximate counting and integration*, in D. Hochbaum, ed., *Approximation Algorithms for NP-Hard Problems*, PWS Publishing, Boston, 1996 pp. 482–520.
- [19] M. Luby, D. Randall, and A. Sinclair, *Markov chain algorithms for planar lattice structures (extended abstract)*, in *36th Annual Symposium on Foundations of Computer Science*, IEEE, Los Alimitos, 1995 pp. 150–159.
- [20] M. Luby and E. Vigoda, *Fast convergence of the Glauber dynamics for sampling independent sets: Part I*, Random Structures and Algorithms, (To appear).
- [21] M. Luby and E. Vigoda, *Approximately counting up to four*, in *Twenty-Ninth Annual Symposium on Theory of Computing*, ACM, New York, 1997 pp. 682–687.
- [22] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, (1995).
- [23] C. H. Papadimitriou, *Computational Complexity*, Addison–Wesley, Reading, MA, (1994).
- [24] J. G. Propp and D. B. Wilson, *Exact sampling from coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, 9 (1996), pp. 223–252.
- [25] D. Randall and P. Tetali, *Analysing Glauber dynamics by comparison of Markov chains*, in *Third Latin American Symposium on Theoretical Informatics*, vol. 1380 of *Lecture Notes in Computer Science*, Springer, Campinas, Brazil, 1998 pp. 292–304.
- [26] D. Roth, *On the hardness of approximate reasoning*, Artificial Intelligence, 82 (1996), pp. 273–302.
- [27] J. Salas and A. D. Sokal, *Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem*, Journal of Statistical Physics, 86 (Feb. 1997) 3–4, pp. 551–579.
- [28] A. Sinclair, *Improved bounds for mixing rates of Markov chains and multicommodity flow*, Combinatorics, Probability and Computing, 1 (1992), pp. 351–370.
- [29] S. Vadhan, *The complexity of counting*, (1995), Undergraduate thesis, Harvard University.
- [30] S. P. Vadhan, *The complexity of counting in sparse, regular, and planar graphs*, Preprint (Available from <http://www-math.mit.edu/~salil/>), (May 1997).