

A counterexample to rapid mixing of the Ge-Štefankovič process*

Leslie Ann Goldberg[†] Mark Jerrum[‡]

Abstract

Ge and Štefankovič have recently introduced a Markov chain which, if rapidly mixing, would provide an efficient procedure for sampling independent sets in a bipartite graph. Such a procedure would be a breakthrough because it would give an efficient randomised algorithm for approximately counting independent sets in a bipartite graph, which would in turn imply the existence of efficient approximation algorithms for a number of significant counting problems whose computational complexity is so far unresolved. Their Markov chain is based on a novel two-variable graph polynomial which, when specialised to a bipartite graph, and evaluated at the point $(\frac{1}{2}, 1)$, gives the number of independent sets in the graph. The Markov chain is promising, in the sense that it overcomes the most obvious barrier to rapid mixing. However, we show here, by exhibiting a sequence of counterexamples, that its mixing time is exponential in the size of the input when the input is chosen from a particular infinite family of bipartite graphs.

Keywords: Glauber dynamics; Independent sets in graphs; Markov chains; Mixing time; Randomised algorithms.

AMS MSC 2010: 60J10; 05C31; 05C69; 68Q17.

Submitted to ECP on October 18, 2011, final version accepted on December 14, 2011.

Supersedes arXiv:1109.5242v1.

1 Overview

In this note, we study the mixing time of a Markov chain, proposed by Ge and Štefankovič, for sampling independent sets in a bipartite graph. The Markov chain is important because, if it were rapidly mixing, it would give an efficient randomised algorithm for approximately counting independent sets in a bipartite graph, which would have consequences in the field of computational complexity.

We start by describing the complexity-theoretic background. Consider the following basic computational problem: Given a bipartite graph, compute the number of independent sets in the graph. It has long been known that this problem, which is denoted #BIS, is computationally difficult, in the sense that it is #P-complete [11]. Informally, that means that #BIS is as difficult as a host of other computational problems, such as computing the permanent of a 0,1-matrix or counting the number of satisfying assignments to a Boolean formula. There is unlikely to be an efficient algorithm for exactly solving the problem.

*Supported by EPSRC Research Grant EP/I011528/1 “Computational Counting”.

[†]Department of Computer Science, University of Liverpool, Ashton Building, Liverpool L69 3BX, UK.

E-mail: l.a.goldberg@liverpool.ac.uk

[‡]School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 NS, UK.

E-mail: m.jerrum@qmul.ac.uk

However, it is still possible that there is an efficient algorithm for *approximately* counting the independent sets in a bipartite graph. This is a fascinating open problem which has received quite a bit of attention. The standard notion of efficient approximation, in the context of counting problems, is the “fully polynomial approximation scheme” or FPRAS. Roughly speaking, an FPRAS is a polynomial-time randomised algorithm that produces an estimate that is close in relative error to the true solution with high probability. (See [10, Defn 11.2] for a precise definition.)

It is well-known (see, for example, [7, §3.2]) that any efficient algorithm for uniformly sampling independent sets in a bipartite graph could easily be turned into an FPRAS for #BIS. It turns out that this would also have consequences for the approximability of other computational counting problems. Dyer et al. [3] showed that a number of counting problems are equivalent to #BIS under approximation-preserving reducibility, and further #BIS-equivalent problems have been presented in subsequent work [1, 6]. Since no FPRAS has been found for any of the counting problems in this equivalence class, it is becoming standard to progress under the assumption that none exists.

So finding an FPRAS for #BIS at this stage would be a significant development. Not only would it imply the existence of an FPRAS for several natural counting problems — such as counting downsets in a partial order, or evaluating the partition function of the ferromagnetic Ising model with local fields — but it would also resolve the complexity of approximating #BIS in the opposite direction to the one many people expect.

The most fruitful approach to designing efficient approximation algorithms for counting problems has been Markov chain Monte Carlo (MCMC). A direct application of MCMC to #BIS would work as follows. Given a bipartite graph G with n vertices, consider the Markov chain whose state space, Ω , is the set of all independent sets in G , and whose transition probabilities $P(\cdot, \cdot)$ are as follows, where \oplus denotes symmetric difference and $H(I) = \{I' \in \Omega \mid |I \oplus I'| = 1\}$:

$$P(I, I') = \begin{cases} 1 - H(I)/2n, & \text{if } I' = I; \\ 1/2n, & \text{if } I' \in H(I); \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that this Markov chain has the uniform distribution on independent sets as its unique stationary distribution. So, simulating the Markov chain for sufficiently many steps would enable us to sample independent sets nearly uniformly. As we mentioned above, from there it is a short step to estimating the number of independent sets [7, §3.2].

To obtain an FPRAS from this approach, one requires that the Markov chain on independent sets is rapidly mixing, i.e., that it is close to the stationary distribution after a number of steps that is polynomial in n . Unfortunately, it is clear that the proposed Markov chain does not have this property. Consider the complete bipartite graph with equal numbers of vertices in the left and right blocks of the bipartition. There are $2^{n/2} - 1$ independent sets that have non-empty intersection with the left block, and the same number with non-empty intersection with the right. Any sequence of transitions which starts in a left-oriented independent set and ends in a right-oriented one must necessarily pass through the empty independent set. Informally, the empty independent set presents an obstruction to rapid mixing by forming a constriction in the state space. This intuition can be made rigorous by noting that the “conductance” of the Markov chain is exponentially small, which implies exponential (in n) mixing time [2, Claim 2.3]. In fact, it is not even necessary to have a dense graph in order to obtain such a constriction: degree 6 will do [2, Thm 2.1]; see also Mossel et al. [9] for sharper results.

Ge and Štefankovič [5] have recently introduced an intriguing graph polynomial $R'_2(G; \lambda, \mu)$, in two indeterminates λ and μ , that is associated with a bipartite graph G . At the point $(\lambda, \mu) = (\frac{1}{2}, 1)$ it counts independent set in G ; specifically, the number of independent sets in G is given by $2^{n-m} R'_2(G; \frac{1}{2}, 1)$, where n is the number of vertices, and m the number of edges in G [5, Thm 4]. This polynomial inspires them to propose a new Markov chain [5, Defn 6] that potentially could be used to sample independent sets from a bipartite graph and hence provide an approximation algorithm for #BIS. The Markov chain, which is described below, is very different from the one discussed earlier. In particular, its states are subsets of the edge set of G rather than subsets of the vertex set. Thus, sampling an independent set of G is a two-stage procedure: (a) sample an edge subset A of G from the appropriate distribution, and then (b) sample an independent set from a distribution conditioned on A . Details will be given below.

The encouraging aspect of this new Markov chain, which we call the Ge-Štefankovič Process, or GS Process for short, is that it is immune to the obvious counterexamples, such as the complete bipartite graph. Unfortunately, with a certain amount of effort it is possible to find a counterexample to rapid mixing. In §2 we describe the GS Process, and in §3 construct a sequence of graphs on which its mixing time is exponential (in the number of vertices of the graph). The precise lower bound on mixing time is presented in Theorem 3.2. Although this counterexample rules out their Markov chain as an approach to constructing a general FPRAS for #BIS, we may still hope that it provides an efficient algorithm for some restricted class of graphs. For example, [5, Theorem 7] shows that it provides an efficient algorithm on trees.

2 The Ge-Štefankovič Process

Before stating our result, we need to formalise what we mean by mixing, rapid or otherwise. Let (X_t) be an ergodic Markov chain with state space Ω , distribution p_t at time t , and unique stationary distribution π . Let $x_0 \in \Omega$ be the initial state of the chain, so that p_0 assigns unit mass to state x_0 . Define the *mixing time* $\tau(x_0)$ with initial state $x_0 \in \Omega$, as the first time t at which $\frac{1}{2} \|p_t - \pi\|_1 \leq e^{-1}$, i.e., at which the distance between the t -step and stationary distributions is at most e^{-1} in total variation; then define the *mixing time* τ as the maximum of $\tau(x_0)$ over all choices of initial state x_0 .

Suppose $G = (U \cup V, E)$ is a bipartite graph, where U, V are disjoint sets forming the vertex bipartition, and E is the edge set. We are interested in two probability spaces, (Ω, π_Ω) and (Σ, π_Σ) , where $\Omega = 2^E$ and $\Sigma = 2^U$. We construct the probability distributions $\pi_\Omega : \Omega \rightarrow [0, 1]$ and $\pi_\Sigma : \Sigma \rightarrow [0, 1]$ with the help of a certain consistency relation χ on $\Sigma \times \Omega$, which is defined as follows. For a pair $(I, A) \in \Sigma \times \Omega$, consider the subgraph of $(U \cup V, A)$ induced by the vertex set $I \cup V$. We say that the relation $\chi(I, A)$ holds iff every vertex of V has even degree in this subgraph. Start with the probability space of consistent pairs $\{(I, A) \in \Sigma \times \Omega \mid \chi(I, A)\}$ with the uniform distribution. Then π_Ω (respectively π_Σ) is the induced marginal distribution on Ω (respectively Σ). We call π_Σ the *marginal BIS distribution* on Σ . The terminology is justified by the following observation.

Lemma 2.1. *With the definitions given above, π_Σ is the distribution induced on U by a uniform random independent set in G . Furthermore, π_Ω is the distribution denoted “RWM($\frac{1}{2}, 1$)” in [5].*

Proof. These facts can be read off from [5, Lemma 10]. The translation of the notation is as follows: $u \in \mathbb{F}_2^{|U|}$ is the characteristic vector of I , B is the graph $(U \cup V, A)$ viewed as a $|U| \times |V|$ bipartite adjacency matrix over \mathbb{F}_2 , and $u^T B = 0$ is equivalent to the condition $\chi(I, A)$. In [5], the distribution RWM($\frac{1}{2}, 1$) is defined in terms of the rank of the matrix B , but it is clear that it is the same as the distribution denoted π_Ω here. \square

The GS-process is an ergodic “single bond flip” Markov chain on state space Ω which has stationary distribution π_Ω . The exact definition of this Markov chain is not important to us, as our counterexample applies to any Markov chain on state space Ω with stationary distribution π_Ω that does not change too many edges in one step. In order to formalise this last requirement, say that a Markov chain with transition probabilities $P : \Omega^2 \rightarrow [0, 1]$, is d -cautious if

$$P(A, A') > 0 \implies |A \oplus A'| \leq d, \quad \text{for all } A, A' \in \Omega.$$

The GS Process is a 1-cautious Markov chain. Our negative result applies to all d -cautious chains, where d depends at most linearly on the number of vertices of G .

3 A counterexample to rapid mixing

The following lemma (taken from [2, Claim 2.3]) packages the conductance argument in a convenient way for us to obtain explicit lower bounds on mixing time.

Lemma 3.1. *Let (X_t) be a Markov chain with state space Ω , transition matrix P and stationary distribution π . Let $\{S, T\}$ be a partition of Ω such that $\pi(S) \leq \frac{1}{2}$, and $C \subset \Omega$ be a set of states that form a “barrier” in the sense that $P(s, t) = 0$ whenever $s \in S \setminus C$ and $t \in T \setminus C$. Then the mixing time of the Markov chain is at least $\pi(S)/8\pi(C)$.*

Let n, m be positive integers such that $(3/2)^m \leq 2^n - 1 < (3/2)^{m+1}$. Note that for every n there is a unique m satisfying the inequalities, and that m depends linearly on n , asymptotically. The counterexample graph (actually sequence of graphs indexed by n) $G_n = (U' \cup V \cup U'', E)$ has vertex set $U' \cup V \cup U''$ where $|U'| = n$ and $|V| = |U''| = m$. The edge set is $E = U' \times V \cup M$, where M is a perfect matching of the vertices in V and U'' . Thus, (a) G_n has bipartition (U, V) where $U = U' \cup U''$, (b) U', V and U'' are all independent sets, (c) the edges between U' and V form a complete graph, and (d) the edges between V and U'' form a matching.

Partition Σ as $\Sigma = \Sigma_0 \cup \Sigma_1$, where $\Sigma_0 = \{I \in \Sigma \mid I \cap U' = \emptyset\}$ and $\Sigma_1 = \Sigma \setminus \Sigma_0$. Observe there are 3^m independent sets in G_n that exclude all vertices in U' , and $(2^n - 1)2^m$ that include some vertex. Since π_Σ is the marginal distribution of independent sets in G_n (see Lemma 2.1),

$$\pi_\Sigma(\Sigma_0) = \frac{3^m}{3^m + (2^n - 1)2^m} = \alpha,$$

where $\frac{2}{5} < \alpha \leq \frac{1}{2}$, by choice of n, m . So $\Sigma_0 \cup \Sigma_1$ is a nearly balanced partition of the state-space Σ . Also it is easy to check that the cut defined by this partition is a witness to the conductance of the “single site flip” Markov chain of §1 being exponentially small in n . This implies that the mixing time of the single site flip Markov chain is exponential in n (which, of course, was never in doubt). Next we identify a partition $\Omega_0 \cup \Omega_1$ that mirrors the partition $\Sigma_0 \cup \Sigma_1$, and itself witnesses exponentially small conductance of the GS Process.

Define the *weight* $w(A)$ of $A \in \Omega$ to be $w(A) = |A \cap M|$. Partition Ω as $\Omega = \Omega_0 \cup \Omega_1$, where $\Omega_0 = \{A \in \Omega \mid w(A) \leq \frac{5}{12}m\}$ and $\Omega_1 = \Omega \setminus \Omega_0$. We aim to show that the weights of states in Ω are concentrated around $\frac{1}{3}m$ and $\frac{1}{2}m$, and there are exponentially few states near the boundary of Ω_0 and Ω_1 . With a view to applying Lemma 3.1, define a “barrier set” (of states) by

$$C = \{A \in \Omega \mid \frac{9}{24}m \leq w(A) \leq \frac{11}{24}m\}.$$

It is not clear how to sample a state A from the distribution (Ω, π_Ω) directly, so instead we sample a state I from (Σ, π_Σ) and then sample u.a.r. a state A consistent with I , i.e., satisfying $\chi(I, A)$. This amounts to the same thing. We have a good handle on (Σ, π_Σ) , since we know it is the marginal BIS distribution (Lemma 2.1).

Suppose we start with a state I sampled from (Σ, π_Σ) , conditional on $I \in \Sigma_0$. The set $I \cap U''$ is determined by a Bernoulli process with success probability $\frac{1}{3}$. (For each edge e in M there are three possibilities for the restriction of the independent set I to e , and only one of them includes a vertex from U'' . These choices are independent for each $e \in M$.) When we come to select a random consistent edge set A , we must exclude all edges in M that are incident to a vertex in $I \cap U''$. The other edges in M are free to be included or excluded. So the set of edges $A \cap M$ is determined by a Bernoulli process with success probability $\frac{1}{3}$. Thus $\mathbb{E}(w(A)) = \frac{1}{3}m$ and, by a Chernoff bound, $\Pr(w(A) \geq \frac{9}{24}m)$ is exponentially small in m . Specifically,

$$\Pr(A \in C) \leq \Pr(w(S) \geq \frac{9}{24}m) \leq \exp(-m/576) \tag{3.1}$$

by [8, Thm 4.4(2)], setting $\delta = \frac{1}{8}$ and $\mu = \frac{1}{3}m$.

Now suppose I is sampled from (Σ, π_Σ) , conditional on $I \in \Sigma_1$. Now select a uniform random A , conditional on the event $\chi(I, A)$. We argue that the probability that a given edge $e = (v, u)$ of M is included in A is $\frac{1}{2}$, independent of all the other edges of M . Suppose $v \in V$ and $(v, u) \in M$. Imagine we are deciding which edges incident to v are to be included in A . First we decide whether to include the edge (v, u) itself. In selecting the remaining edges for A from the n available, we just have to make sure that the parity of $A \cap (\{v\} \times (I \cap U'))$ is odd, if $(v, u) \in A$ and $u \in I$, and even otherwise. Since $I \cap U' \neq \emptyset$, the number of ways to do this is 2^{n-1} , independent of whether we included edge e in the first place. It follows that the set of edges $A \cap M$ is determined by a Bernoulli process with success probability $\frac{1}{2}$. Thus $\mathbb{E}(w(A)) = \frac{1}{2}m$ and, by Chernoff, $\Pr(w(A) \leq \frac{11}{24}m)$ is exponentially small in n . Specifically,

$$\Pr(A \in C) \leq \Pr(w(S) \leq \frac{11}{24}m) \leq \exp(-m/576) \tag{3.2}$$

by [8, Thm 4.5(2)], setting $\delta = \frac{1}{12}$ and $\mu = \frac{1}{2}m$.

We see now that the partition $\Omega_0 \cup \Omega_1 = \Omega$ is balanced, since $\pi_\Omega(\Omega_0) = \alpha \pm o(1)$ and $\frac{2}{5} < \alpha \leq \frac{1}{2}$. Moreover, from (3.1) and (3.2), $\Pr(\frac{9}{24} \leq w(A) \leq \frac{11}{24})$ is exponentially small when A is selected from the distribution (Ω, π_Ω) ; specifically, $\pi_\Omega(C) \leq \exp(-m/576)$. Thus the cut (Ω_0, Ω_1) is witness to the conductance of the single bond flip MC being exponentially small. Suppose $d \leq m/12$. Observe that no d -cautious chain can make a transition from $\Omega_0 \setminus C$ to $\Omega_1 \setminus C$. Applying Lemma 3.1, we therefore obtain.

Theorem 3.2. *Suppose that n, m, G_n, Ω and π_Ω are as above, and that $d \leq m/12$. Any ergodic Markov chain on state space Ω with stationary distribution π_Ω that is d -cautious has mixing time $\Omega(\exp(m/576))$. In particular, the GS Process, which is 1-cautious, has mixing time exponential in the number of vertices in G_n .*

It is also natural to consider a ‘‘Swendsen-Wang-style’’ Markov chain for sampling from (Σ, π_Σ) . Let $I \in \Sigma$ be the current state. Choose A u.a.r. from the set $\{A \in \Omega \mid \chi(I, A)\}$. Then choose I' u.a.r. from the set $\{I' \in \Sigma \mid \chi(I', A)\}$. The new state is I' . We can think of this process as a Markov chain on state space $\Sigma \cup \Omega$ with stationary distribution $\frac{1}{2}\pi_\Sigma$ on Σ and $\frac{1}{2}\pi_\Omega$ on Ω . (Assume a continuous time process to avoid the obvious periodicity.) It follows from the earlier analysis that the cut $(\Sigma_0 \cup \Omega_0, \Sigma_1 \cup \Omega_1)$ witnesses exponentially small conductance. To see this, we calculate the probability in stationarity of observing a transition from $\Sigma_0 \cup \Omega_0$ to $\Sigma_1 \cup \Omega_1$. There are two possibilities: a transition from Σ_0 to Ω_1 , or one from Ω_0 to Σ_1 . The probability of the former, we have seen, is $\frac{1}{2}\pi_\Sigma(\Sigma_0)$ times a quantity that is exponentially small in n . The latter is, by time reversibility, the same as observing, in stationarity, a transition from Σ_1 to Ω_0 . This probability is again exponentially small in n . Hence the conductance is exponentially small so the mixing time of the Swendsen-Wang-style Markov chain is exponential in n .

We can also look a little closer, to see what is going on in more detail. Sample a state ω at random from (Σ, π_Σ) , conditioned on the event $\omega \in \Sigma_0$, and apply a “half-step” of the SW-like process to obtain a state $A \in \Omega$. We know that $A \cap M$ is described by a Bernoulli process with success probability $\frac{1}{3}$. Moreover, it is easy to see the remaining edges of A are Bernoulli with success probability $\frac{1}{2}$. Now consider the transition from A to I' . As in [5], view the set $I' \cap U'$ as a n -vector u' over \mathbb{F}_2 . Each of the vertices in V that is *not* incident to an edge of $A \cap M$ generates a linear equation, with constant term zero, constraining u' . These $\frac{2}{3}m \approx 1.1397n$ random linear equations constrain just n variables; so with high probability the only solution is to set all n variables to 0. (Equivalently, a random $n \times \frac{2}{3}m$ matrix over \mathbb{F}_2 has rank n with high probability.) In other words, $I' \cap U' = \emptyset$, except with exponentially small probability, and we find ourselves back in Σ_0 again.

References

- [1] Prasad Chebolu, Leslie Ann Goldberg, and Russell Martin. The complexity of approximately counting stable matchings. In *Proceedings of the 13th international conference on Approximation, and 14 the International conference on Randomization, and combinatorial optimization: algorithms and techniques*, APPROX/RANDOM'10, pages 81–94, Berlin, Heidelberg, 2010. Springer-Verlag.
- [2] Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM J. Comput.*, 31(5):1527–1541 (electronic), 2002. MR-1936657
- [3] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004. Approximation algorithms. MR-2044886
- [4] Martin Dyer, Leslie Ann Goldberg, Mark Jerrum, and Russell Martin. Markov chain comparison. *Probab. Surv.*, 3:89–111 (electronic), 2006. MR-2216963
- [5] Qi Ge and Daniel Stefankovic. A graph polynomial for independent sets of bipartite graphs. In Kamal Lodaya and Meena Mahajan, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2010)*, volume 8 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 240–250, Dagstuhl, Germany, 2010.
- [6] Leslie Ann Goldberg and Mark Jerrum. The complexity of ferromagnetic Ising with local fields. *Combin. Probab. Comput.*, 16(1):43–61, 2007. MR-2286511
- [7] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003. MR-1960003
- [8] Michael Mitzenmacher and Eli Upfal. *Probability and Computing*. Cambridge University Press, Cambridge, 2005. Randomized algorithms and probabilistic analysis. MR-2144605
- [9] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probab. Theory Related Fields*, 143(3-4):401–439, 2009. MR-2475668
- [10] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995. MR-1344451
- [11] J. Scott Provan and Michael O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Comput.*, 12(4):777–788, 1983. MR-0721012