Markov chain Monte Carlo algorithm for the mean field random-cluster model

Antonio Blanca
University of California-Berkeley
(Advisor: Alistair Sinclair)
ablanca@cs.berkeley.edu

Submitted to the ACM Student Research Competition Grand Finals 04/2013

Abstract

The random-cluster model has emerged as a unifying framework for studying random graphs, physical spin systems, and electrical networks [6]. Given a base graph $G = (V, E)$, the random-cluster model consists of a probability distribution π over subsets of E, where $p \in [0, 1]$ and $q > 0$ are real parameters determining the bond and cluster weights respectively. For $q \geq 1$, there is a value of $p = p_c$ at which the random-cluster model undergoes a phase transition.

The heat-bath dynamics is a local Markov chain used to sample from π. For $G = K_n$, the mean field case, we show that the heat-bath dynamics mixes in $O(n^2 \log n)$ steps when $q > 1$ and $p < \min\{p_c, \frac{2}{n}\}$. Our result provides the first upper bounds for the mixing time for non-integer values of q, and it is tight up to a logarithmic factor. For some specific cases where an upper bound was previously known, we improve such bounds by at least a $O(n^2)$ factor. Our result also provides the first polynomial bound for the mixing time of the Swendsen-Wang dynamics when $q \geq 3$ and $\frac{1/3}{n} < p < \frac{2}{n}$.

1 Problem and Motivation

The *random-cluster model* was created by Fortuin and Kasteleyn in the late 1960's as a unifying framework for studying random graphs, physical spin systems, and electrical networks [6]. Let $G = (V, E)$ be a graph, $p \in [0, 1]$, and $q > 0$. The configurations of the random-cluster model are all subgraphs $H = (V, E_0)$ of G, which we identify with their edge set $E_0 \subseteq E$. The probability of a configuration E_0 is:

$$\pi(E_0 \subseteq E) = \frac{p^{|E_0|}(1-p)^{|E|-|E_0|}q^{c(E_0)}}{Z_{\mathrm{RC}}}$$

where $c(E_0)$ denotes the number of connected components of $H = (V, E_0)$, and Z_{RC} is a normalizing constant called the *partition function*.

The special case when $q = 1$ has been widely studied in the framework of random graphs [1]. In particular when G is the complete graph with n vertices (denoted K_n), the random-cluster model reduces to the standard Erdős-Rényi $G(n, p)$ model where a configuration is obtained by adding each edge with probability p independently. When $q < 1$, configurations with fewer connected components ("clusters") are favored whereas the opposite happens when $q > 1$.

The random-cluster model is a generalization of the Ising ($q = 2$) and Potts models ($q \geq 3$ with $q \in \mathbb{Z}$), two fundamental models for ferromagnetism in statistical physics [11, 20]. Much of the physical theory of the Ising and Potts models can be understood in the context of the random-cluster model. In the late 1980's, interest in the random-cluster model resurged [5, 17], and it became a central tool for modeling and understanding the ferromagnet and its phase transition [8].

In addition, because the random-cluster model is a model for random graphs which takes into account connectivity properties, it is also useful in the study of electrical and other networks [8].

1.1 The heat-bath dynamics

The heat-bath dynamics is a local Markov chain over the space of random-cluster configurations $\Omega = \{E_0 : E_0 \subseteq E\}$. Given a configuration $E_0 \subseteq E$, the heat-bath dynamics performs the following random step:

1. Choose an edge $e \in E$ uniformly at random.
2. With probability p_e the new configuration is $E_0 \cup \{e\}$, where p_e is the conditional probability of $E_0 \cup \{e\}$ determined by the state of the other edges:

$$p_e = \frac{\pi(E_0 \cup \{e\})}{\pi(E_0 \cup \{e\}) + \pi(E_0 \setminus \{e\})}.$$

Otherwise, the new configuration is $E_0 \setminus \{e\}$:

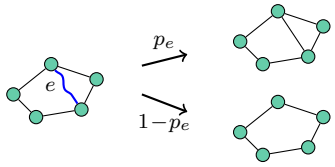


Figure 1: Heat-bath step.

The stationary distribution of the heat-bath dynamics is the random-cluster measure π (see, e.g., [8]). Therefore, the heat-bath dynamics provides a standard Markov chain Monte Carlo (MCMC) algorithm to sample from π by repeating the steps above until the Markov chain is statistically close to its stationary distribution. The time it takes for a Markov chain to reach such a steady state is called the *mixing time*.

To rigorously establish the running time of this MCMC algorithm, we study the mixing time of the heat-bath dynamics. In particular, we focus on the case when $G = K_n$, also known as the *mean field* case. An important motivation for studying the convergence of the heat-bath dynamics in the mean field is the *universality conjecture*, which suggests that the mixing behavior for K_n should extend to a larger class of graphs [13] providing a more general MCMC algorithm.

The heat-bath dynamics is of interest in its own right as a model of the evolution of a physical system by local moves. Its mixing time is related to the speed at which the system approaches a steady state.

2 Background and Related Work

2.1 Markov chain convergence

Let P be the transition matrix of a Markov chain with state space Ω and stationary distribution μ. The *mixing time* measures the time until the chain is close to its stationary distribution starting from a worst possible configuration in Ω :

$$t_{mix} = \max_{z \in \Omega} \min_t \left\{ ||P^t(z, \cdot) - \mu(\cdot)||_{\mathrm{TV}} \leq 1/4 \right\}$$

where $|| \cdot ||_{\mathrm{TV}}$ is the total variation distance[1].

A *coupling* of a Markov chain specifies the joint evolution of a pair of realizations (X_t, Y_t) of the chain, such that the processes $\{X_t\}$ and $\{Y_t\}$, viewed in isolation, are faithful copies of the Markov chain, and if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

For any pair of states $x, y \in \Omega$, the *coupling time* from starting states x and y, denoted T_{xy}, is the first time the chains $\{X_t\}$ and $\{Y_t\}$ meet:

$$T_{xy} = \min_t \{X_t = Y_t | X_0 = x, Y_0 = y\}.$$

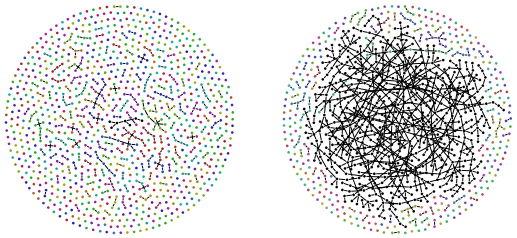
The overall coupling time T is the maximum of T_{xy} over all possible starting states. If at time t the probability that the processes $\{X_t\}$ and $\{Y_t\}$ have not met is sufficiently small, then t bounds the mixing time. This is captured in the following standard inequality:

$$t_{mix} \leq \min_t \{\Pr[T > t] \leq 1/4\}. \tag{1}$$

2.2 Previous Work

When $q \geq 1$, the random-cluster model exhibits a phase transition which marks the arrival of a giant component in the graph. More precisely, there is a threshold value p_c of p such that a configuration will have a connected component of size $O(n)$ with high probability only when $p > p_c$. When $p < p_c$, the largest component will have size $O(\log n)$ with high probability (see Fig. 2).

[1]For two probability distributions μ and ν supported on a set Ω, $||\mu - \nu||_{TV} = \max_{S \subset \Omega} |\mu(S) - \nu(S)|$.



(a) $p = \frac{1.5}{n}$ (b) $p = \frac{2.5}{n}$

Figure 2: Samples from the mean field random-cluster measure with $n = 1000$, $q = 2$, and p below and above p_c. When $q = 2$, $p_c = \frac{2}{n}$.

In the mean field, the phase transition is well understood thanks to work by Bollobás, Grimmett and Janson [2] and Luczak and Łuczak [14].

The mixing behavior of the heat-bath dynamics is closely linked to its phase transition. The best known bounds for the mixing time come from an interesting connection unraveled recently by Ullrich [18, 19] between the heat-bath dynamics and a non-local process called the Swendsen-Wang (SW) dynamics [17].

The following bounds for the mixing time of the heat-bath dynamics are obtained by applying Ullrich's results to previous work done for the SW dynamics in the mean field. Due to the indirect nature of this approach, a $O(n^4 \log n)$ factor is lost.

As a consequence of work by Gore and Jerrum [7], the heat-bath dynamics is known to converge exponentially slowly when $q \geq 3$, $q \in \mathbb{Z}$ and $p = p_c$. Huber's result [9] for the same range of q guarantees that the mixing time is bounded by $O(n^4 \log^2 n)$ and $O(qn^5 \log n)$ when $p < \frac{1/3}{n}$ and $p > \frac{2q \log(3qn)}{n}$ respectively.

When $q = 2$, the heat-bath dynamics is known to mix in $O(n^{4.5} \log n)$ steps at all non-critical temperatures ($p \neq p_c$) as a consequence of work by Cooper, Dyer, Frieze and Rue [4]. Long, Nachmias, Ning, and Peres' results [13] improve upon this result for $q = 2$ guaranteeing that the heat-bath dynamics mixes in $O(n^4 \log n)$ steps in the sub-critical case ($p < p_c$), $O(n^{4.25} \log n)$ steps at criticality ($p = p_c$), and $O(n^4 \log^2 n)$ steps in the super-critical case ($p > p_c$).

There is a trivial $\Omega(n^2)$ lower bound for the mixing time of the heat-bath dynamics for all p and q [12]. Extensive simulations suggest that for $q = 2$ the mixing time is much closer to this lower bound than what the known upper bounds suggest. In fact, we conjecture that the mixing time is $\tilde{O}(n^2)$ for all values of p.

3 Approach and Uniqueness

3.1 Results

Our result is stated in the following theorem.

Theorem 1. *Let $q > 1$, and $p < \min\left\{p_c, \frac{2}{n}\right\}$. The heat-bath dynamics in the mean field random-cluster model mixes in $O(n^2 \log n)$ steps.*

For the important special when $q = 2$, our result improves the best known upper bound due to Long et al. [13] by a $O(n^2)$ factor for $p < p_c$. For $q \geq 3$ with $q \in \mathbb{Z}$, our result improves the best known upper bound due to Huber [9] by a $O(n^2 \log n)$ factor for $p < \frac{1/3}{n}$. Moreover, our bounds apply to a much larger regime of the parameters p and q, providing the first polynomial upper bounds for non-integer values of q. In particular, our bound applies for $q \in (1, 2]$ throughout the sub-critical regime ($p < p_c$). Our result also implies the first polynomial bound for the mixing time of the Swendsen-Wang dynamics for $\frac{1/3}{n} < p < \frac{2}{n}$ when $q \geq 3$ with $q \in \mathbb{Z}$.

Our work is the first to analyze the mixing time of the heat-bath dynamics directly. We believe that in order to obtain tight upper bounds (as those in Theorem 1) for other regimes of the parameters, a direct approach is also required.

3.2 Methods

Let $\{X_t\}$ and $\{Y_t\}$ be two copies of the heat-bath Markov chain. We couple (X_t, Y_t) in each heat-bath step by using the same random edge $e \in E$ for both copies of the chain and the same uniform random number $r \in [0, 1]$ to compare to p_e and decide whether to add or remove e. This

is a valid coupling since $\{X_t\}$ and $\{Y_t\}$ will be faithful copies of the chain. For $q > 1$, it is easy to check that if $Y_t \subseteq X_t$, then $Y_{t+1} \subseteq X_{t+1}$. A coupling with this property is called a *monotone coupling*.

For any $X_t, Y_t \in \Omega$ such that $Y_t \subseteq X_t$, define the metric $d(X_t, Y_t) = |X_t| - |Y_t|$, and let $\Delta_t = d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)$ be the one-step change in distance. Ideally, we would like to show that $d(X_t, Y_t)$ is decreasing on expectation. That is,

$$\mathrm{E}[\Delta_t | X_t, Y_t] \leq -\epsilon d(X_t, Y_t) \tag{2}$$

for some $\epsilon \in (0, 1)$. Inductively, (2) implies that

$$\mathrm{E}[d(X_t, Y_t) | X_0, Y_0] \leq (1 - \epsilon)^t \operatorname{diam}(\Omega),$$

where $\operatorname{diam}(\Omega)$ is the maximum possible distance $d(\cdot, \cdot)$ between any pair of configurations. By Markov's inequality,

$$\begin{aligned} \Pr[X_t \neq Y_t | X_0, Y_0] &= \Pr[d(X_t, Y_t) \geq 1 | X_0, Y_0] \\ &\leq \mathrm{E}[d(X_t, Y_t) | X_0, Y_0] \\ &\leq (1 - \epsilon)^t \operatorname{diam}(\Omega). \end{aligned}$$

Therefore $\Pr[X_t \neq Y_t | X_0, Y_0] \to 0$ when $t \approx 2\epsilon^{-1} \ln \operatorname{diam}(\Omega)$, and by (1) the mixing time is $O(\epsilon^{-1} \log n)$. Therefore, if we could establish (2) with $\epsilon = \Omega(n^{-2})$ we would have our desired bound of $O(n^2 \log n)$ on the mixing time.

3.3 Proof Overview

A *bridge* is an edge whose removal increases the number of connected components in the graph. Let $D(Y_t)$ be the set of edges in $E \setminus Y_t$ that would be bridges if they were added to Y_t. Define $D(X_t)$ similarly, and let $D_t = |D(Y_t)| - |D(X_t)|$ and $\mu = p\left(1 - q^{-1}\right)$. We are able to bound the expected one-step change in distance:

$$\mathrm{E}[\Delta_t | X_t, Y_t] \leq \frac{-(|X_t| - |Y_t|) + \mu\,(D_t + n)}{\binom{n}{2}}. \tag{3}$$

We might now hope to use (3) to prove (2) for some suitable ϵ. However, it is fairly simple to construct valid pairs (X_t, Y_t) where the right hand side of (3) is positive. Therefore, the first moment method described in Section 3.2 is not sufficient. Inspired by a more refined coupling technique developed by Hayes and Vigoda [10], we are able to overcome this difficulty.

Take $X_0 = K_n$, $Y_0 = \emptyset$, $Z_0 \sim \pi$ and couple these three realizations of the heat-bath Markov chain. Notice that X_0 and Y_0 are the two extremal configurations of the coupling. Therefore, by monotonicity $T = \min_t\{X_t = Y_t\}$ is an upper bound for the coupling time, and $Y_t \subseteq Z_t \subseteq X_t$ for all t (see Fig. 3). Then,

$$T \leq \max\{T_{top}, T_{bottom}\}$$

where $T_{top} = \min_t\{X_t = Z_t\}$ and $T_{bottom} = \min_t\{Z_t = Y_t\}$. We focus then on bounding T_{top} and T_{bottom}. Notice that $Z_t \sim \pi$ for all t, so in a sense we are *coupling with the stationary distribution*, as in [10], but the monotonicity of our coupling simplifies the arguments (see Fig. 3).

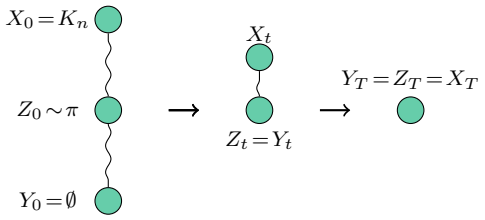


Figure 3: Coupling with the stationary distribution. In this case $t = T_{bottom}$ and $T = T_{top}$.

Bounding T_{top} and T_{bottom} is now an easier task since we only need to consider pairs of states (X_t, Y_t) where either X_t or Y_t is from the stationary distribution. The following lemma due to Bollobás, Grimmett and Janson [2] establishes structural properties of the stationary distribution that we crucially use to bound (3).

Lemma 2 ([2]). *If $p < p_c$ and $H = (V, E) \sim \pi$ with $|V| = n$. Then, quite surely (q.s.) [2], the largest component of H has size $O(\log n)$. Moreover, $|E| \leq \frac{p}{2q} n^2 + o(n)$ q.s..*

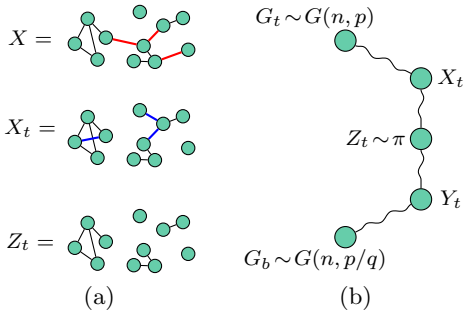
[2] We say that an event $\mathcal{A}$ happens "quite surely" if $\Pr[\mathcal{A}] \geq 1 - O(n^{-a})$ for any $a \in \mathbb{N}$. In fact, Lemma 2 is a refinement of the lemma presented in [2].

To bound T_{bottom} we use Lemma 2 to obtain a bound for (3) as in (2) with $\epsilon = \Omega(n^{-2})$. We omit the details due to lack of space.

Bounding T_{top} requires a more delicate argument. For any pair of states (X_t, Z_t) with $Z_t \sim \pi$ and $Z_t \subseteq X_t$, it is sufficient to bound (3) for that $X \subseteq E$ which maximizes D_t such that $|X| = |X_t|$ and $Z_t \subseteq X$. If $m = |X_t \setminus Z_t|$, this X is the result of merging the $m + 1$ largest components of Z_t (see Fig. 4a). This key observation allows us to use the extra structure of X to bound D_t.

To get a good lower bound for $|X|$, we run a "burn-in" phase until every edge has been sampled by the coupling. By coupon collecting, the time until all edges have been sampled is $O(n^2 \log n)$ q.s.; therefore, the burn-in phase will not change the order of the coupling time.

For any edge $e \in E$, p_e is equal (asymptotically) to either p or p/q. Therefore, after the burn-in phase, Z_t contains a subgraph distributed as $G(n, p/q)$ and a supergraph distributed as $G(n, p)$ by monotonicity (see Fig. 4b). Then, using binomial concentration we can bound the number of edges in X.



(a) (b)

Figure 4: In (a), X maximizes D_t for (X_t, Z_t). In (b) is shown the status of the coupling after the "burn-in" phase.

The upper bound for $|Y_t|$ comes directly from Lemma 2. Putting together all these bounds we can obtain a bound for (3) as in (2) with $\epsilon = \Omega(n^{-2})$. By the arguments at the beginning of Section 2.1 (and the similar bound for T_{bottom}), this implies the desired $O(n^2 \log n)$ bound on the mixing time. Again, the details of the proofs are omitted due to lack of space.

4 Contributions

The random-cluster model is a fundamental tool in statistical physics for modeling and understanding the ferromagnet and its phase transition. It is also an important model of random graphs and networks that takes into account their connectivity properties. Therefore, sampling efficiently from the random-cluster measure is a problem of high interest for the statistical physics and applied probability communities.

We showed that a Markov chain Monte Carlo algorithm using the heat-bath dynamics runs in $O(n^2 \log n)$ time over a large range of the parameters. For some of the cases where an upper bound for the running time was known, we improve such results by a $O(n^2)$ factor at least.

Although monotonicity, "burn-in" phases and coupling with the stationary distribution techniques have been used separately in the past to bound the coupling time of several Markov chains, we combine these techniques in an interesting way to obtain our results. We believe our methods could help to establish the convergence of other important Markov chains.

Our analysis of the heat-bath dynamics is rigorous, and our work provides tight theoretical bounds for the running time of a Markov chain Monte Carlo sampling algorithm for the random-cluster measure. The heat-bath algorithm is easy to implement, and it is likely to be used by practitioners to obtain high-quality samples from the random-cluster measure given the theoretical guarantees we have shown.

In addition, the heat-bath dynamics models the evolution of a physical system by local moves. Its mixing time is of interest to physicist for understanding the speed at which the system approaches a steady state. We hope our bounds for the mixing time help characterize the behavior of this physical system.

Our work is a compelling example of how computer science can effectively contribute to advances in other sciences by providing efficient tools and computational insights into important problems.

References

[1] B. Bollobás. Random graphs. *Academic Press,* London, 1985.

[2] B. Bollobás, G. Grimmett, and S. Janson . The Random-Cluster Model On The Complete Graph. *Probability Theory and Related Fields* **104**: 283-317, 1996.

[3] C. Borgs, J. T. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. Vu. Torpid mixing of some mcmc algorithms in statistical physics. *FOCS* **2**: 218–229, 1999.

[4] C. Cooper, M. Dyer, A. Frieze, and R. Rue. Mixing properties of the swendsen-wang process on the complete graph and narrow grids. *Journal of Mathematical Physics* **41**: 1499–1527, 2000.

[5] Edwards, R. G. and Sokal, A. D. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *The Physical Review D* **38**: 2009–2012, 1988.

[6] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. i. introduction and relation to other models. *Physica* **57**: 536–564, 1972.

[7] V. Gore and M. Jerrum. The swendsen-wang process does not always mix rapidly. *STOC*: 674–681, 1996.

[8] G. R. Grimmett. The random-cluster model. *A Series of Comprehensive Studies in Mathematics,* Springer, Berlin, 2000.

[9] M. Huber. A bounding chain for Swendsen-Wang. *Random Structures & Algorithms* **22**: 43-59, 2003.

[10] T. Hayes, and E. Vigoda. Coupling with the Stationary Distribution and Improved Sampling for Colorings and Independent Sets. *Annals of Applied Probability* **16(4)**: 1397-1318, 2006.

[11] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Phys* **31**: 253–258, 1925.

[12] D.A. Levin, Y. Peres, and E.L. Wilmer. Markov Chains and Mixing times. *AMS*, Providence, 2009.

[13] Y. Long, A. Nachmias, W. Ning, and Y. Peres. A power law of order 1/4 for critical mean-field Swendsen-Wang dynamics. *To appear,* 2011.

[14] M. Luczak and T. Łuczak The phase transition in the cluster-scaled model of a random graph *Random Structures & Algorithms* **28(2)**: 215-246, 2005.

[15] B. Pittel On tree census and the giant component in sparse random graphs. *Random Structures & Algorithms* **1(3)**: 311-342, 1990.

[16] R. B. Potts Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society* **48 (01)**: 106–109, 1952.

[17] R. H. Swendsen and J.S. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical Reviews Letters* **58**: 86–88, 1987.

[18] M. Ullrich. Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions. *Submitted*, 2012.

[19] M. Ullrich. Swendsen-Wang is faster than single-bond dynamics. *Submitted*, 2012.

[20] Fa-Yueh Wu. The Potts model. *Reviews of Modern Physics* **54(1)**: 235–268, 1982.