Rapid Mixing Swendsen-Wang Sampler for Stochastic Partitioned Attractive Models

Sejun Park, Yunhun Jang, Andreas Galanis, Jinwoo Shin, Daniel Štefankovič, Eric Vigoda

Motivation: Torpid Mixing of Gibbs Sampler

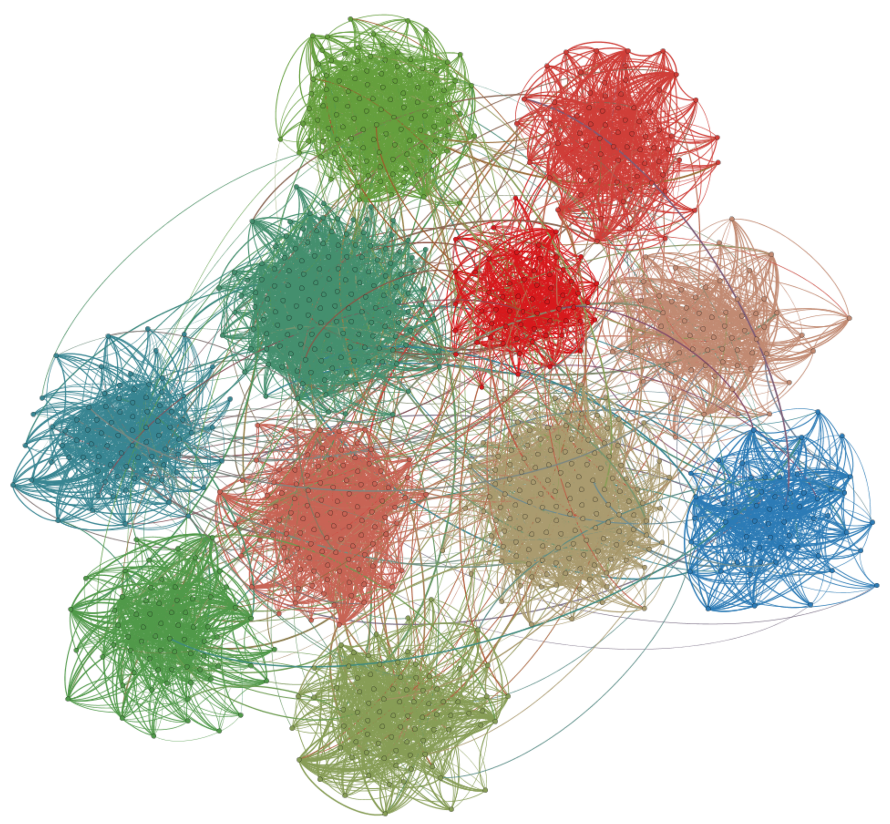
- Gibbs sampler is one of the most popular Markov chains used for learning and inference problems in graphical models. These tasks are computationally intractable in general, and Gibbs sampler often suffers from slow mixing.
- In this research, we study Swendsen-Wang dynamics (SW) which is a more sophisticated Markov chain for attractive Ising models designed to overcome bottlenecks of Gibbs sampler.
- Recently, Guo et al. proved $O(|V|^{10})$ mixing of SW for Ising models on arbitrary graphs. Rapid mixing of SW has been studied for complete graph (Peres et al., Vigoda et al.), grid graph (Ullrich) which have fully symmetric graphical structures.

Ising Model and Stochastic Partitioned Model

- Given a graph (V, E), Ising model, equivalently pairwise binary model, is a joint distribution on $x \in \{-1, 1\}^V$ defined as

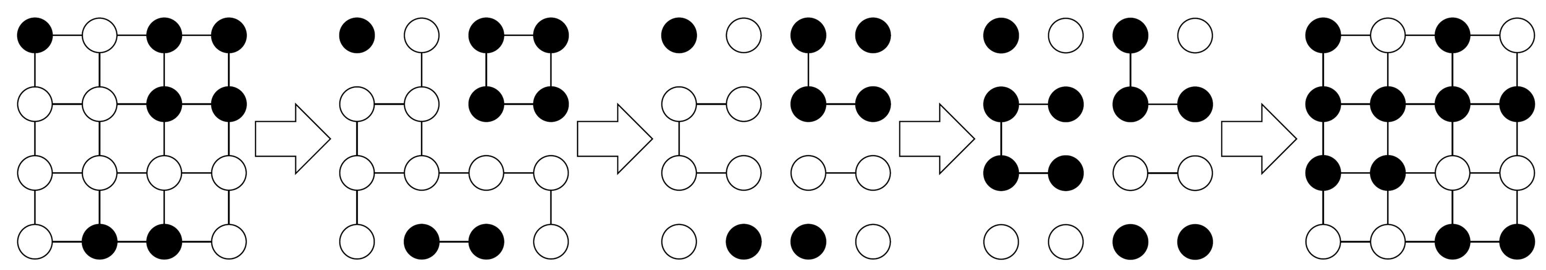
$$P(x) \propto \exp\left(\sum_{v \in V} \gamma_v x_v + \sum_{(u,v) \in E} \beta_{uv} x_u x_v\right).$$

- Stochastic partitioned model is a random graph model such that given vertices and disjoint communities of vertices $S_1, \dots, S_r$, any edge between communities S_i and S_j exists with probability p_{ij}.
- Social graphs, Erdős-Rényi random graphs and r-partite graphs are examples of stochastic partitioned models.



Swendsen-Wang Dynamics

- Swendsen-Wang dynamics is a Markov chain running on attractive Ising model (i.e., $\beta_{uv} \geq 0$) which converges to the distribution of Ising model. The transition from X^t to X^{t+1} is as follows:



1. Delete edges between vertices of different spins in X^t.
2. For each remaining edge (u, v), delete it with probability $q(\beta_{uv})$.
3. For each connected component C, assign a spin 1 to vertices in C with probability $q(\gamma_C)$ and -1 with probability $1 - q(\gamma_C)$.
4. Denote X^{t+1} as a resulting state.

Main Result: Rapid Mixing of Swendsen-Wang

- Our main result is proving the rapid mixing of Swendsen-Wang dynamics for Ising models on stochastic partitioned graphs while Gibbs sampler suffers from exponential running time.

Theorem 1 *The mixing time of Swendsen-Wang chain on a stochastic partitioned model is $O(\log |V|)$ a.a.s. if $\gamma \geq 0$ and either (a) or (b) holds:*

(a) *For all i and for all $u, v \in S_i$, $p_{ii}, \beta_{uv} = \Omega(1)$.*

(b) *For all $i \neq j$ and for all $u \in S_i, v \in S_j$, $p_{ij}, \beta_{uv} = \Omega(1)$.*

- Theorem 1 states that Swendsen-Wang dynamics rapidly mixes under mild conditions $(\beta, p = \Omega(1))$.
- Ising models on social graphs, restricted Boltzmann machine, deep Boltzmann machine are application of Theorem 1.

Theorem 2 *The mixing time of Swendsen-Wang chain on a complete bipartite graph is $O(\log |V|)$ if $\beta = k/|V|$ where $k \neq 2$ is any constant.*

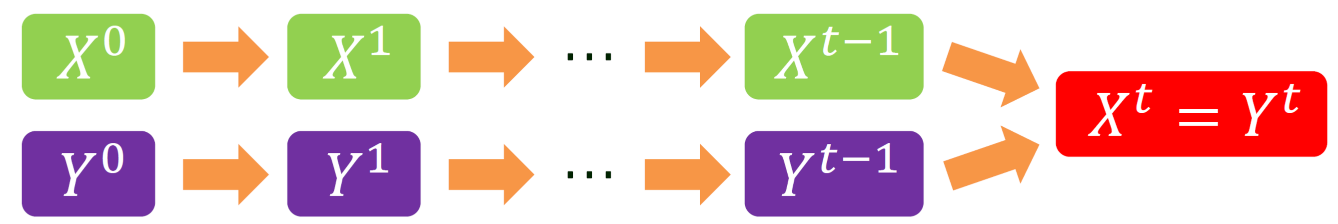
- Theorem 2 states that Swendsen-Wang dynamics on a complete bipartite graph rapidly mixes in the high temperature region.
- Theorem 2 is theoretically challenging as only fully symmetric graphs were analyzed so far.

Proof Sketch

- We provide a proof for Theorem 1 (b), i.e., the practical setting.

Lemma 1 *The mixing time is bounded above by*

$$\arg\min_t \left(\sup_{X^0, Y^0} P(X^t \neq Y^t) \leq \frac{1}{4}\right).$$



- From Lemma 1, we address to bound the probability $P(X^t \neq Y^t)$.

Lemma 2 *For any X^0, the larger spin class S^t of X^t is connected after step 2 of Swendsen-Wang dynamics. for $t = O(\log |V|)$ a.a.s. Also,*

$$E[|S^t|] \geq \frac{1}{2}(|V| + |S^{t-1}|).$$

- From Lemma 2 and Markov's inequality, X^t consists of a single spin, i.e., $S^t = V$, for some $t = O(\log |V|)$ a.a.s.
- Same argument holds for Y^t. Therefore, $P(X^t = Y^t) = 1 - o(1)$ for some $t = O(\log |V|)$ which directly bounds the mixing time.

Experiments

- We compare Swendsen-Wang dynamics and Gibbs sampler by learning parameters of Ising models using contrastive divergence.
- Learning tasks are performed for attractive Ising models on two real world social graphs and synthetic graphs while parameters are randomly chosen from $\gamma_v \sim$ Unif$(-1, 1)$ and $\beta_{uv} \sim$ Unif$(0, x)$.
- **Facebook Graph:** 4039 vertices and 88234 edges