Dynamic Inference in Probabilistic Graphical Models

Weiming Feng
State Key Laboratory for Novel Software Technology, Nanjing University, China
fengwm@smail.nju.edu.cn

Kun He
Shenzhen University, China
Shenzhen Institute of Computing Sciences, China
hekun.threebody@foxmail.com

Xiaoming Sun
State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China
sunxiaoming@ict.ac.cn

Yitong Yin
State Key Laboratory for Novel Software Technology, Nanjing University, China
yinyt@nju.edu.cn

Abstract

Probabilistic graphical models, such as Markov random fields (MRFs), are useful for describing high-dimensional distributions in terms of local dependence structures. The probabilistic inference is a fundamental problem related to graphical models, and sampling is a main approach for the problem. In this paper, we study probabilistic inference problems when the graphical model itself is changing dynamically with time. Such dynamic inference problems arise naturally in today's application, e.g. multivariate time-series data analysis and practical learning procedures.

We give a dynamic algorithm for sampling-based probabilistic inferences in MRFs, where each dynamic update can change the underlying graph and all parameters of the MRF simultaneously, as long as the total amount of changes is bounded. More precisely, suppose that the MRF has n variables and polylogarithmic-bounded maximum degree, and $N(n)$ independent samples are sufficient for the inference for a polynomial function $N(\cdot)$. Our algorithm dynamically maintains an answer to the inference problem using $\widetilde{O}(nN(n))$ space cost, and $\widetilde{O}(N(n)+n)$ incremental time cost upon each update to the MRF, as long as the Dobrushin-Shlosman condition is satisfied by the MRFs. This well-known condition has long been used for guaranteeing the efficiency of Markov chain Monte Carlo (MCMC) sampling in the traditional static setting. Compared to the static case, which requires $\Omega(nN(n))$ time cost for redrawing all $N(n)$ samples whenever the MRF changes, our dynamic algorithm gives a $\widetilde{\Omega}(\min\{n, N(n)\})$-factor speedup. Our approach relies on a novel dynamic sampling technique, which transforms local Markov chains (a.k.a. single-site dynamics) to dynamic sampling algorithms, and an "algorithmic Lipschitz" condition that we establish for sampling from graphical models, namely, when the MRF changes by a small difference, samples can be modified to reflect the new distribution, with cost proportional to the difference on MRF.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Computing methodologies → Machine learning

Keywords and phrases Dynamic inference, probabilistic graphical model, Gibbs sampling, Markov random filed

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.25

Related Version A full version of the paper is available at `https://arxiv.org/abs/1904.11807`.

12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 25; pp. 25:1–25:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding Weiming Feng and Yitong Yin are supported by the National Key R&D Program of China 2018YFB1003202 and the National Natural Science Foundation of China under Grant Nos. 61722207 and 61672275. Kun He and Xiaoming Sun are supported by the National Natural Science Foundation of China Grants No. 61832003, 61433014 and K.C.Wong Education Foundation.

1 Introduction

The probabilistic graphical models provide a rich language for describing high-dimensional distributions in terms of the dependence structures between random variables. The *Markov random filed* (MRF) is a basic graphical model that encodes pairwise interactions of complex systems. Given a graph $G = (V, E)$, each vertex $v \in V$ is associated with a function $\phi_v : Q \to \mathbb{R}$, called the *vertex potential*, on a finite domain $Q = [q]$ of q *spin states*, and each edge $e \in E$ is associated with a symmetric function $\phi_e : Q^2 \to \mathbb{R}$, called the *edge potential*, which describes a pairwise interaction. Together, these induce a probability distribution μ over all configurations $\sigma \in Q^V$:

$$\mu(\sigma) \propto \exp(H(\sigma)) = \exp\Big(\sum_{v\in V}\phi_v(\sigma_v) + \sum_{e=\{u,v\}\in E}\phi_e(\sigma_u,\sigma_v)\Big).$$

This distribution μ is known as the Gibbs distribution and $H(\sigma)$ is the *Hamiltonian.* It arises naturally from various physical models, statistics or learning problems, and combinatorial problems in computer science [29, 25].

The *probabilistic inference* is one of the most fundamental computational problems in graphical model. Some basic inference problems ask to calculate the marginal distribution, conditional distribution, or maximum-a-posteriori probabilities of one or several random variables [37]. Sampling is perhaps the most widely used approach for probabilistic inference. Given a graphical model, independent samples are drawn from the Gibbs distribution and certain statistics are computed using the samples to give estimates for the inferred quantity. For most typical inference problems, such statistics are easy to compute once the samples are given, for instance, for estimating the marginal distribution on a variable subset S, the statistics is the frequency of each configuration in Q^S among the samples, thus the cost for inference is dominated by the cost for generating random samples [24, 35].

The classic probabilistic inference assumes a static setting, where the input graphical model is fixed. In today's application, dynamically changing graphical models naturally arise in many scenarios. In various practical algorithms for learning graphical models, e.g. the contrastive divergence algorithm for learning the restricted Boltzmann machine [22] and the iterative proportional fitting algorithm for maximum likelihood estimation of graphical models [37], the optimal model $\mathcal{I}^*$ is obtained by updating the parameters of the graphical model iteratively (usually by gradient descent), which generates a sequence of graphical models $\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_M$, with the goal that $\mathcal{I}_M$ is a good approximation of $\mathcal{I}^*$. Also in the study of the multivariate time-series data, the dynamic Gaussian graphical models [5], multiregression dynamic model [32], dynamic graphical model [14], and dynamic chain graph models [2], are all dynamically changing graphical models and have been used in a variety of applications. Meanwhile, with the advent of Big Data, scalable machine learning systems need to deal with continuously evolving graphical models (see e.g. [33] and [34]).

The theoretical studies of probabilistic inference in dynamically changing graphical models are lacking. In the aforementioned scenarios in practice, it is common that a sequence of graphical models is presented with time, where any two consecutive graphical models can differ from each other in all potentials but by a small total amount. Recomputing the inference problem from scratch at every time when the graphical model is changed, can give the correct solution, but is very wasteful. A fundamental question is whether probabilistic inference can be solved dynamically and efficiently.

In this paper, we study the problem of probabilistic inference in an MRF when the MRF itself is changing dynamically with time. At each time, the whole graphical model, including all vertices and edges as well as their potentials, are subject to changes. Such *non-local* updates are very general and cover all applications mentioned above. The problem of *dynamic inference* then asks to maintain a correct answer to the inference in a dynamically changing MRF with low incremental cost proportional to the amount of changes made to the graphical model at each time.

1.1 Our results

We give a dynamic algorithm for sampling-based probabilistic inferences. Given an MRF instance with n vertices, suppose that $N(n)$ independent samples are sufficient to give an approximate solution to the inference problem, where $N : \mathbb{N}^+ \to \mathbb{N}^+$ is a polynomial function. We give dynamic algorithms for general inference problems on dynamically changing MRF.

Suppose that the current MRF has n vertices and polylogarithmic-bounded maximum degree, and each update to the MRF may change the underlying graph and/or all vertex/edge potentials, as long as the total amount of changes is bounded. Our algorithm maintains an approximate solution to the inference with $\widetilde{O}(nN(n))$ space cost, and with $\widetilde{O}(N(n) + n)$ incremental time cost upon each update, assuming that the MRFs satisfy the Dobrushin-Shlosman condition [8, 9, 7]. The condition has been widely used to imply the efficiency of Markov chain Monte Carlo (MCMC) sampling (e.g. see [19, 12]). Compared to the static algorithm, which requires $\Omega(nN(n))$ time for redrawing all $N(n)$ samples each time, our dynamic algorithm significantly improves the time cost with an $\widetilde{\Omega}(\min\{n, N(n)\})$-factor speedup.

On specific models, the Dobrushin-Shlosman condition has been established in the literature, which directly gives us following efficient dynamic inference algorithms, with $\widetilde{O}\left(nN(n)\right)$ space cost and $\widetilde{O}\left(N(n) + n\right)$ time cost per update, on graphs with n vertices and maximum degree $\Delta = O(1)$:

- for Ising model with temperature β satisfying $\mathrm{e}^{-2|\beta|} > 1 - \frac{2}{\Delta+1}$, which is close to the uniqueness threshold $\mathrm{e}^{-2|\beta_c|} = 1 - \frac{2}{\Delta}$, beyond which the static versions of sampling or marginal inference problem for anti-ferromagnetic Ising model is intractable [17, 16];
- for hardcore model with fugacity $\lambda < \frac{2}{\Delta-2}$, which matches the best bound known for sampling algorithm with near-linear running time on general graphs with bounded maximum degree [36, 28, 13];
- for proper q-coloring with $q > 2\Delta$, which matches the best bound known for sampling algorithm with near-linear running time on general graphs with bounded maximum degree [23].

Our dynamic inference algorithm is based on a dynamic sampling algorithm, which efficiently maintains $N(n)$ independent samples for the current MRF while the MRF is subject to changes. More specifically, we give a dynamic version of the *Gibbs sampling* algorithm, a local Markov chain for sampling from the Gibbs distribution that has been studied extensively. Our techniques are based on: (1) couplings for dynamic instances of graphical models; and (2) dynamic data structures for representing single-site Markov chains so that the couplings can be realized algorithmically in sub-linear time. Both these techniques are of independent interest, and can be naturally extended to more general settings with multi-body interactions.

Our results show that on dynamically changing graphical models, sampling-based probabilistic inferences can be solved significantly faster than rerunning the static algorithm at each time. This has practical significance in speeding up the iterative procedures for learning graphical models.

1.2 Related work

The problem of dynamic sampling from graphical models was introduced very recently in [14]. There, a dynamic sampling algorithm was given for graphical models with soft constraints, and can only deal with local updates that change a single vertex or edge at each time. The regimes for such dynamic sampling algorithm to be efficient are much more restrictive than the conditions for the rapid mixing of Markov chains. Our algorithm greatly improves the regimes for efficient dynamic sampling for the Ising and hardcore models in [14], and for the first time, can handle non-local updates that change all vertex/edge potentials simultaneously. Besides, the dynamic/online sampling from log-concave distributions was also studied in [31, 26].

Another related topic is the dynamic graph problems, which ask to maintain a solution (e.g. spanners [15, 30, 38] or shortest paths [3, 21, 20]) while the input graph is dynamically changing. More recently, important progress has been made on dynamically maintaining structures that are related to graph random walks, such as spectral sparsifier [11, 1] or effective resistances [10, 18]. Instead of one particular solution, dynamic inference problems ask to maintain an estimate of a statistics, such statistics comes from an exponential-sized probability space described by a dynamically changing graphical model.

1.3 Organization of the paper

In Section 2, we formally introduce the dynamic inference problem. In Section 3, we formally state the main results. Preliminaries are given in Section 4. In Section 5, we outline our dynamic inference algorithm. In Section 6, we present the algorithms for dynamic Gibbs sampling. The conclusion is given in Section 7. The analyses of the dynamic sampling algorithms and the proof of the main theorem on dynamic inference are provided in the full version of the paper.

2 Dynamic inference problem

2.1 Markov random fields

An instance of *(pairwise) Markov random field (MRF)* is specified by a tuple $\mathcal{I} = (V, E, Q, \Phi)$, where $G = (V, E)$ is an undirected simple graph; Q is a domain of $q = |Q|$ *spin states*, for some finite $q > 1$; and $\Phi = (\phi_a)_{a \in V \cup E}$ associates each $v \in V$ a *vertex potential* $\phi_v : Q \to \mathbb{R}$ and each $e \in E$ an *edge potential* $\phi_e : Q^2 \to \mathbb{R}$, where ϕ_e is symmetric.

A *configuration* $\sigma \in Q^V$ maps each vertex $v \in V$ to a spin state in Q, so that each vertex can be interpreted as a variable. And the *Hamiltonian* of a configuration $\sigma \in Q^V$ is defined as:

$$H(\sigma) \triangleq \sum_{v \in V} \phi_v(\sigma_v) + \sum_{e=\{u,v\} \in E} \phi_e(\sigma_u, \sigma_v).$$

This defines the *Gibbs distribution* $\mu_{\mathcal{I}}$, which is a probability distribution over Q^V such that

$$\forall \sigma \in Q^V, \quad \mu_{\mathcal{I}}(\sigma) = \frac{1}{Z} \exp(H(\sigma)),$$

where the normalizing factor $Z \triangleq \sum_{\sigma \in Q^V} \exp(H(\sigma))$ is called the *partition function*.

The Gibbs measure $\mu(\sigma)$ can be 0 as the functions ϕ_v, ϕ_e can take the value $-\infty$. A configuration σ is called *feasible* if $\mu(\sigma) > 0$. To trivialize the problem of constructing a feasible configuration, we further assume the following natural condition for the MRF instances considered in this paper:[1]

$$\forall\, v \in V,\ \forall \sigma \in Q^{\Gamma_G(v)}: \quad \sum_{c \in Q} \exp\left(\phi_v(c) + \sum_{u \in \Gamma_v} \phi_{uv}(\sigma_u, c)\right) > 0. \tag{1}$$

where $\Gamma_G(v) \triangleq \{u \in V \mid \{u,v\} \in E\}$ denotes the neighborhood of v in graph $G = (V,E)$.

Some well studied typical MRFs include:

- *Ising model*: The domain of each spin is $Q = \{-1,+1\}$. Each edge $e \in E$ is associated with a *temperature* $\beta_e \in \mathbb{R}$; and each vertex $v \in V$ is associated with a *local field* $h_v \in \mathbb{R}$. For each configuration $\sigma \in \{-1,+1\}^V$, $\mu_{\mathcal{I}}(\sigma) \propto \exp\left(\sum_{\{u,v\} \in E} \beta_e \sigma_u \sigma_v + \sum_{v \in V} h_v \sigma_v\right)$.
- *Hardcore model*: The domain is $Q = \{0,1\}$. Each configuration $\sigma \in Q^V$ indicates an independent set in $G = (V,E)$, and $\mu_{\mathcal{I}}(\sigma) \propto \lambda^{\|\sigma\|}$, where $\lambda > 0$ is the *fugacity* parameter.
- *proper q-coloring*: uniform distribution over all proper q-colorings of $G = (V,E)$.

2.2 Probabilistic inference and sampling

In graphical models, the task of probabilistic inference is to derive the probabilities regarding one or more random variables of the model. Abstractly, this is described by a function $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$ that maps each graphical model $\mathcal{I} \in \mathfrak{M}$ to a target K-dimensional probability vector, where $\mathfrak{M}$ is the class of graphical models containing the random variables we are interested in and the K-dimensional vector describes the probabilities we want to derive. Given $\boldsymbol{\theta}(\cdot)$ and an MRF instance $\mathcal{I} \in \mathfrak{M}$, the inference problem asks to estimate the probability vector $\boldsymbol{\theta}(\mathcal{I})$.

Here are some fundamental inference problems [37] for MRF instances. Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance and $A, B \subseteq V$ two disjoint sets where $A \uplus B \subseteq V$.

- *Marginal inference*: estimate the marginal distribution $\mu_{A,\mathcal{I}}(\cdot)$ of the variables in A, where

$$\forall \sigma_A \in Q^A, \quad \mu_{A,\mathcal{I}}(\sigma_A) \triangleq \sum_{\tau \in Q^{V \setminus A}} \mu_{\mathcal{I}}(\sigma_A, \tau).$$

- *Posterior inference*: given any $\tau_B \in Q^B$, estimate the posterior distribution $\mu_{A,\mathcal{I}}(\cdot \mid \tau_B)$ for the variables in A, where

$$\forall \sigma_A \in Q^A, \quad \mu_{A,\mathcal{I}}(\sigma_A \mid \tau_B) \triangleq \frac{\mu_{A \cup B,\mathcal{I}}(\sigma_A, \tau_B)}{\mu_{B,\mathcal{I}}(\tau_B)}.$$

- *Maximum-a-posteriori (MAP) inference*: find the maximum-a-posteriori (MAP) probabilities $P^*_{A,\mathcal{I}}(\cdot)$ for the configurations over A, where

$$\forall \sigma_A \in Q^A, \quad P^*_{A,\mathcal{I}}(\sigma_A) \triangleq \max_{\tau_B \in Q^B} \mu_{A \cup B,\mathcal{I}}(\sigma_A, \tau_B).$$

[1] This condition guarantees that the marginal probabilities are always well-defined, and the problem of constructing a feasible configuration σ, where $\mu_{\mathcal{I}}(\sigma) > 0$, is trivial. The condition holds for all MRFs with soft constraints, or with hard constraints where there is a permissive spin, e.g. the hardcore model. For MRFs with truly repulsive hard constraints such as proper q-coloring, the condition may translate to the condition $q \geq \Delta + 1$ where Δ is the maximum degree of graph G, which is necessary for the irreducibility of local Markov chains for q-colorings.

All these fundamental inference problems can be described abstractly by a function $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$. For instances, for marginal inference, $\mathfrak{M}$ contains all MRF instances where A is a subset of the vertices, $K = |Q|^{|A|}$, and $\boldsymbol{\theta}(\mathcal{I}) = (\mu_{A,\mathcal{I}}(\sigma_A))_{\sigma_A \in Q^A}$; and for posterior or MAP inference, $\mathfrak{M}$ contains all MRF instances where $A \uplus B$ is a subset of the vertices, $K = |Q|^{|A|}$ and $\boldsymbol{\theta}(\mathcal{I}) = (\mu_{A,\mathcal{I}}(\sigma_A \mid \tau_B))_{\sigma_A \in Q^A}$ (for posterior inference) or $\boldsymbol{\theta}(\mathcal{I}) = (P^*_{A,\mathcal{I}}(\sigma_A))_{\sigma_A \in Q^A}$ (for MAP inference).

One canonical approach for probabilistic inference is by sampling: sufficiently many independent samples are drawn (approximately) from the Gibbs distribution of the MRF instance and an estimate of the target probabilities is calculated from these samples. Given a probabilistic inference problem $\boldsymbol{\theta}(\cdot)$, we use $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ to denote an estimating function that approximates $\boldsymbol{\theta}(\mathcal{I})$ using independent samples drawn approximately from $\mu_{\mathcal{I}}$. For the aforementioned problems of marginal, posterior and MAP inferences, such estimating function $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ simply counts the frequency of the samples that satisfy certain properties.

The sampling cost of an estimator is captured in two aspects: the number of samples it uses and the accuracy of each individual sample it requires.

▶ **Definition 1** ((N, ϵ)-estimator for $\boldsymbol{\theta}$)**.** *Let $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$ be a probabilistic inference problem and $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ an estimating function for $\boldsymbol{\theta}(\cdot)$ that for each instance $\mathcal{I} = (V, E, Q, \Phi) \in \mathfrak{M}$, maps samples in Q^V to an estimate of $\boldsymbol{\theta}(\mathcal{I})$. Let $N : \mathbb{N}^+ \to \mathbb{N}^+$ and $\epsilon : \mathbb{N}^+ \to (0,1)$. For any instance $\mathcal{I} = (V, E, Q, \Phi) \in \mathfrak{M}$ where $n = |V|$, the random variable $\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N(n))})$ is said to be an* (N, ϵ)-estimator *for $\boldsymbol{\theta}(\mathcal{I})$ if $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N(n))} \in Q^V$ are $N(n)$ independent samples drawn approximately from $\mu_{\mathcal{I}}$ such that $d_{\mathrm{TV}}\left(\boldsymbol{X}^{(j)}, \mu_{\mathcal{I}}\right) \leq \epsilon(n)$ for all $1 \leq j \leq N(n)$.*

In Definition 1, an estimator is viewed as a black-box algorithm specified by two functions N and ϵ. Usually, the estimator is more accurate if more independent samples are drawn and each sample provides a higher level of accuracy. Thus, one can choose some large N and small ϵ to achieve a desired quality of estimate.

2.3 Dynamic inference problem

We consider the inference problem where the input graphical model is changed dynamically: at each step, the current MRF instance $\mathcal{I} = (V, E, Q, \Phi)$ is updated to a new instance $\mathcal{I}' = (V', E', Q, \Phi')$. We consider general update operations for MRFs that can change both the **underlying graph** and **all edge/vertex potentials** simultaneously, where the update request is made by a *non-adaptive adversary* independently of the randomness used by the inference algorithm. Such updates are general enough and cover many applications, e.g. analyses of time series network data [5, 32, 14, 2], and learning algorithms for graphical models [22, 37].

The difference between the original and the updated instances is measured as follows.

▶ **Definition 2** (difference between MRF instances)**.** *The difference between two MRF instances $\mathcal{I} = (V, E, Q, \Phi)$ and $\mathcal{I}' = (V', E', Q, \Phi')$, where $\Phi = (\phi_a)_{a \in V \cup E}$ and $\Phi' = (\phi'_a)_{a \in V' \cup E'}$, is defined as*

$$d(\mathcal{I}, \mathcal{I}') \triangleq \sum_{v \in V \cap V'} \|\phi_v - \phi'_v\|_1 + \sum_{e \in E \cap E'} \|\phi_e - \phi'_e\|_1 + |V \oplus V'| + |E \oplus E'|, \tag{2}$$

where $A \oplus B = (A \setminus B) \cup (B \setminus A)$ stands for the symmetric difference between two sets A and B, $\|\phi_v - \phi'_v\|_1 \triangleq \sum_{c \in Q} |\phi_v(c) - \phi'_v(c)|$, and $\|\phi_e - \phi'_e\|_1 \triangleq \sum_{c,c' \in Q} |\phi_e(c,c') - \phi'_e(c,c')|$.

Given a probability vector specified by the function $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$, the *dynamic inference problem* asks to maintain an estimator $\hat{\boldsymbol{\theta}}(\mathcal{I})$ of $\boldsymbol{\theta}(\mathcal{I})$ for the current MRF instance $\mathcal{I} = (V, E, Q, \Phi) \in \mathfrak{M}$, with a data structure, such that when $\mathcal{I}$ is updated to $\mathcal{I}' = (V', E', Q, \Phi') \in \mathfrak{M}$, the algorithm updates $\hat{\boldsymbol{\theta}}(\mathcal{I})$ to an estimator $\hat{\boldsymbol{\theta}}(\mathcal{I}')$ of the new vector $\boldsymbol{\theta}(\mathcal{I}')$, or equivalently, outputs the difference between the estimators $\hat{\boldsymbol{\theta}}(\mathcal{I})$ and $\hat{\boldsymbol{\theta}}(\mathcal{I}')$.

It is desirable to have the dynamic inference algorithm which maintains an (N, ϵ)-estimator for $\boldsymbol{\theta}(\mathcal{I})$ for the current instance $\mathcal{I}$. However, the dynamic algorithm cannot be efficient if $N(n)$ and $\epsilon(n)$ change drastically with n (so that significantly more samples or substantially more accurate samples may be needed when a new vertex is added), or if recalculating the estimating function $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ itself is expensive. We introduce a notion of *dynamical efficiency* for the estimators that are suitable for dynamic inference.

▶ **Definition 3** (dynamical efficiency). *Let $N : \mathbb{N}^+ \to \mathbb{N}^+$ and $\epsilon : \mathbb{N}^+ \to (0, 1)$. Let $\mathcal{E}(\cdot)$ be an estimating function for some K-dimensional probability vector of MRF instances. An tuple $(N, \epsilon, \mathcal{E})$ is said to be* dynamically efficient *if it satisfies:*

- ***(bounded difference)*** *there exist constants $C_1, C_2 > 0$ such that for any $n \in \mathbb{N}^+$,*

$$|N(n+1) - N(n)| \leq \frac{C_1 \cdot N(n)}{n} \quad \text{and} \quad |\epsilon(n+1) - \epsilon(n)| \leq \frac{C_2 \cdot \epsilon(n)}{n};$$

- ***(small incremental cost)*** *there is a deterministic algorithm that maintains $\mathcal{E}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)})$ using $(mn + K) \cdot \mathrm{polylog}(mn)$ bits where $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)} \in Q^V$ and $n = |V|$, such that when $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)} \in Q^V$ are updated to $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(m')} \in Q^{V'}$, where $n' = |V'|$, the algorithm updates $\mathcal{E}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)})$ to $\mathcal{E}(\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(m')})$ within time cost $\mathcal{D} \cdot \mathrm{polylog}(mm'nn') + O(m + m')$, where $\mathcal{D}$ is the size of the difference between two sample sequences defined as:*

$$\mathcal{D} \triangleq \sum_{i \leq \max\{m, m'\}} \sum_{v \in V \cup V'} \mathbf{1}\left[\boldsymbol{X}^{(i)}(v) \neq \boldsymbol{Y}^{(i)}(v)\right], \tag{3}$$

where an unassigned $\boldsymbol{X}^{(i)}(v)$ or $\boldsymbol{Y}^{(i)}(v)$ is not equal to any assigned spin.

The dynamic efficiency basically asks $N(\cdot), \epsilon(\cdot)$, and $\mathcal{E}(\cdot)$ to have some sort of "Lipschitz" properties. To satisfy the bounded difference condition, $N(n)$ and $1/\epsilon(n)$ are necessarily polynomially bounded, and they can be any constant, polylogarithmic, or polynomial functions, or multiplications of such functions. The condition with small incremental cost also holds very commonly. In particular, it is satisfied by the estimating functions for all the aforementioned problems for the marginal, posterior and MAP inferences as long as the sets of variables have sizes $|A|, |B| = O(\log n)$. We remark that the $O(\log n)$ upper bound is somehow necessary for the efficiency of inference, because otherwise the dimension of $\boldsymbol{\theta}(\mathcal{I})$ itself (which is at least $q^{|A|}$) becomes super-polynomial in n.

3 Main results

Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance, where $G = (V, E)$. Let $\Gamma_G(v)$ denote the neighborhood of v in G. For any vertex $v \in V$ and any configuration $\sigma \in Q^{\Gamma_G(v)}$, we use $\mu^{\sigma}_{v,\mathcal{I}}(\cdot) = \mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ to denote the marginal distribution on v conditional on σ:

$$\forall c \in Q : \quad \mu^{\sigma}_{v,\mathcal{I}}(c) = \mu_{v,\mathcal{I}}(c \mid \sigma) \triangleq \frac{\exp\left(\phi_v(c) + \sum_{u \in \Gamma_G(v)} \phi_{uv}(\sigma_u, c)\right)}{\sum_{a \in Q} \exp\left(\phi_v(a) + \sum_{u \in \Gamma_G(v)} \phi_{uv}(\sigma_u, a)\right)}.$$

Due to the assumption in (1), the marginal distribution is always well-defined. The following condition is the *Dobrushin-Shlosman condition* [8, 9, 7, 19, 12].

▶ **Condition 4** (Dobrushin-Shlosman condition). *Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance with Gibbs distribution $\mu = \mu_{\mathcal{I}}$. Let $A_{\mathcal{I}} \in \mathbb{R}_{\geq 0}^{V \times V}$ be the* influence matrix *which is defined as*

$$A_{\mathcal{I}}(u,v) \triangleq \begin{cases} \max_{(\sigma,\tau)\in B_{u,v}} d_{\mathrm{TV}}\left(\mu_v^{\sigma}, \mu_v^{\tau}\right), & \{u,v\} \in E, \\ 0 & \{u,v\} \notin E, \end{cases}$$

where the maximum is taken over the set $B_{u,v}$ of all $(\sigma,\tau) \in Q^{\Gamma_G(v)} \times Q^{\Gamma_G(v)}$ that differ only at u, and $d_{\mathrm{TV}}\left(\mu_v^{\sigma}, \mu_v^{\tau}\right) \triangleq \frac{1}{2}\sum_{c\in Q} |\mu_v^{\sigma}(c) - \mu_v^{\tau}(c)|$ is the total variation distance between μ_v^{σ} and μ_v^{τ}. An MRF instance $\mathcal{I}$ is said to satisfy the Dobrushin-Shlosman condition *if there is a constant $\delta > 0$ such that*

$$\max_{u \in V} \sum_{v \in V} A_{\mathcal{I}}(u,v) \leq 1 - \delta.$$

Our main theorem assumes the following setup: Let $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$ be a probabilistic inference problem that maps each MRF instance in $\mathfrak{M}$ to a K-dimensional probability vector, and let $\mathcal{E}_{\boldsymbol{\theta}}$ be its estimating function. Let $N : \mathbb{N}^+ \to \mathbb{N}^+$ and $\epsilon : \mathbb{N}^+ \to (0,1)$. We use $\mathcal{I} = (V, E, Q, \Phi) \in \mathfrak{M}$, where $n = |V|$, to denote the current instance and $\mathcal{I}' = (V', E', Q, \Phi') \in \mathfrak{M}$, where $n' = |V'|$, to denote the updated instance.

▶ **Theorem 5** (dynamic inference algorithm). *Assume that $(N, \epsilon, \mathcal{E}_{\boldsymbol{\theta}})$ is dynamically efficient, both $\mathcal{I}$ and $\mathcal{I}'$ satisfy the Dobrushin-Shlosman condition, and $d(\mathcal{I}, \mathcal{I}') \leq L = o(n)$.*

There is an algorithm that maintains an (N, ϵ)-estimator $\hat{\boldsymbol{\theta}}(\mathcal{I})$ of the probability vector $\boldsymbol{\theta}(\mathcal{I})$ for the current MRF instance $\mathcal{I}$, using $\widetilde{O}\left(nN(n) + K\right)$ bits, such that when $\mathcal{I}$ is updated to $\mathcal{I}'$, the algorithm updates $\hat{\boldsymbol{\theta}}(\mathcal{I})$ to an (N, ϵ)-estimator $\hat{\boldsymbol{\theta}}(\mathcal{I}')$ of $\boldsymbol{\theta}(\mathcal{I}')$ for the new instance $\mathcal{I}'$, within expected time cost

$$\widetilde{O}\left(\Delta^2 L N(n) + \Delta n\right),$$

where $\widetilde{O}(\cdot)$ hides a polylog(n) *factor, $\Delta = \max\{\Delta_G, \Delta_{G'}\}$, where Δ_G and $\Delta_{G'}$ denote the maximum degree of $G = (V, E)$ and $G' = (V', E')$ respectively.*

Note that the extra $O(\Delta n)$ cost is necessary for editing the current MRF instance $\mathcal{I}$ to $\mathcal{I}'$.

Typically, the difference between two MRF instances $\mathcal{I}, \mathcal{I}'$ is small[2], and the underlying graphs are sparse [6] , that is, $L, \Delta \leq \mathrm{polylog}(n)$. In such cases, our algorithm updates the estimator within time cost $\widetilde{O}(N(n) + n)$, which significantly outperforms static sampling-based inference algorithms that require time cost $\Omega(n'N(n')) = \Omega(nN(n))$ for redrawing all $N(n')$ independent samples.

Dynamic sampling. The core of our dynamic inference algorithm is a dynamic sampling algorithm: Assuming the Dobrushin-Shlosman condition, the algorithm can maintain a sequence of $N(n)$ independent samples $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N(n))} \in Q^V$ that are $\epsilon(n)$-close to $\mu_{\mathcal{I}}$ in total variation distance, and when $\mathcal{I}$ is updated to $\mathcal{I}'$ with difference $d(\mathcal{I}, \mathcal{I}') \leq L = o(n)$, the algorithm can update the maintained samples to $N(n')$ independent samples $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(N(n'))} \in Q^{V'}$ that are $\epsilon(n')$-close to $\mu_{\mathcal{I}'}$ in total variation distance, using a time cost $\widetilde{O}\left(\Delta^2 L N(n) + \Delta n\right)$ in expectation. This shows an "algorithmic Lipschitz" condition

[2] In multivariate time-series data analysis, the MRF instances of two sequential times are similar. In the iterative algorithms for learning graphical models, the difference between two sequential MRF instances generated by gradient descent are bounded to prevent oscillations. Specifically, the difference is very small when the iterative algorithm approaches to the convergence state [22, 37].

holds for sampling from Gibbs distributions: when the MRF changes insignificantly, a population of samples can be modified to reflect the new distribution, with cost proportional to the difference on MRF. We show that such property is guaranteed by the Dobrushin-Shlosman condition. This dynamic sampling algorithm is formally described in Theorem 9 and is of independent interest [14].

Applications on specific models. On specific models, we have the following results, where $\delta > 0$ is an arbitrary constant.

Table 1 Dynamic inference for specific models.

model	regime	space cost	time cost for each update
Ising	$\mathrm{e}^{-2\lvert\beta\rvert} \geq 1 - \frac{2-\delta}{\Delta+1}$	$\widetilde{O}\left(nN(n) + K\right)$	$\widetilde{O}\left(\Delta^2 LN(n) + \Delta n\right)$
hardcore	$\lambda \leq \frac{2-\delta}{\Delta-2}$	$\widetilde{O}\left(nN(n) + K\right)$	$\widetilde{O}\left(\Delta^3 LN(n) + \Delta n\right)$
q-coloring	$q \geq (2+\delta)\Delta$	$\widetilde{O}\left(nN(n) + K\right)$	$\widetilde{O}\left(\Delta^2 LN(n) + \Delta n\right)$

The results for Ising model and q-coloring are corollaries of Theorem 5. The regime for hardcore model is better than the Dobrushin-Shlosman condition (which is $\lambda \leq \frac{1-\delta}{\Delta-1}$), because we use the coupling introduced by Vigoda [36] to analyze the algorithm.

4 Preliminaries

Total variation distance and coupling. Let μ and ν be two distributions over Ω. The *total variation distance* between μ and ν is defined as

$$d_{\mathrm{TV}}(\mu, \nu) \triangleq \frac{1}{2} \sum_{x \in \Omega} \lvert \mu(x) - \nu(x) \rvert .$$

A *coupling* of μ and ν is a joint distribution $(X, Y) \in \Omega \times \Omega$ such that marginal distribution of X is μ and the marginal distribution of Y is ν. The following coupling lemma is well-known.

▶ **Proposition 6** (coupling lemma). *For any coupling (X, Y) of μ and ν, it holds that*

$$\Pr[X \neq Y] \geq d_{\mathrm{TV}}(\mu, \nu) .$$

Furthermore, there is an optimal coupling *that achieves equality.*

Local neighborhood. Let $G = (V, E)$ be a graph. For any vertex $v \in V$, let $\Gamma_G(v) \triangleq \{u \in V \mid \{u, v\} \in E\}$ denote the neighborhood of v, and $\Gamma_G^+(v) \triangleq \Gamma_G(v) \cup \{v\}$ the inclusive neighborhood of v. We simply write $\Gamma_v = \Gamma(v) = \Gamma_G(v)$ and $\Gamma_v^+ = \Gamma^+(v) = \Gamma_G^+(v)$ for short when G is clear in the context. We use $\Delta = \Delta_G \triangleq \max_{v \in V} \lvert \Gamma_v \rvert$ to denote the maximum degree of graph G.

A notion of **local neighborhood for MRF** is frequently used. Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance. For $v \in V$, we denote by $\mathcal{I}_v \triangleq \mathcal{I}[\Gamma_v^+]$ the restriction of $\mathcal{I}$ on the inclusive neighborhood Γ_v^+ of v, i.e. $\mathcal{I}_v = (\Gamma_v^+, E_v, Q, \Phi_v)$, where $E_v = \{\{u, v\} \in E\}$ and $\Phi_v = (\phi_a)_{a \in \Gamma_v^+ \cup E_v}$.

Gibbs sampling. The *Gibbs sampling* (a.k.a. *heat-bath*, *Glauber dynamics*), is a classic Markov chain for sampling from Gibbs distributions. Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance and $\mu = \mu_{\mathcal{I}}$ its Gibbs distribution. The chain of Gibbs sampling (Algorithm 1) is on the space $\Omega \triangleq Q^V$, and has the stationary distribution $\mu_{\mathcal{I}}$ [27, Chapter 3].

Algorithm 1 Gibbs sampling.

```
Initialization: an initial state $\boldsymbol{X}_0 \in \Omega$ (not necessarily feasible);
for $t = 1, 2, \ldots, T$ do
    pick $v_t \in V$ uniformly at random;
    draw a random value $c \in Q$ from the marginal distribution $\mu_{v_t}(\cdot \mid X_{t-1}(\Gamma_{v_t}))$;
    $X_t(v_t) \leftarrow c$ and $X_t(u) \leftarrow X_{t-1}(u)$ for all $u \in V \setminus \{v_t\}$;
```

Marginal distributions. Here $\mu_v(\cdot \mid \sigma(\Gamma_v)) = \mu_{v,\mathcal{I}}(\cdot \mid \sigma(\Gamma_v))$ denotes the marginal distribution at $v \in V$ conditioning on $\sigma(\Gamma_v) \in Q^{\Gamma_v}$, which is computed as:

$$\forall c \in Q: \quad \mu_v(c \mid \sigma(\Gamma_v)) = \frac{\phi_v(c) \prod_{u \in \Gamma_v} \phi_{uv}(\sigma_u, c)}{\sum_{c' \in Q} \phi_v(c') \prod_{u \in \Gamma_v} \phi_{uv}(\sigma_u, c')}. \tag{4}$$

Due to the assumption (1), this marginal distribution is always well defined, and its computation uses only the information of $\mathcal{I}_v$.

Coupling for mixing time. Consider a chain $(\boldsymbol{X}_t)_{t=0}^{\infty}$ on space Ω with stationary distribution $\mu_{\mathcal{I}}$ for MRF instance $\mathcal{I}$. The *mixing rate* is defined as: for $\epsilon > 0$,

$$\tau_{\mathsf{mix}}(\mathcal{I}, \epsilon) \triangleq \max_{\boldsymbol{X}_0} \min \left\{ t \mid d_{\mathrm{TV}}\left(\boldsymbol{X}_t, \mu_{\mathcal{I}}\right) \le \epsilon \right\},$$

where $d_{\mathrm{TV}}\left(\boldsymbol{X}_t, \mu_{\mathcal{I}}\right)$ denotes the *total variation distance* between $\mu_{\mathcal{I}}$ and the distribution of $\boldsymbol{X}_t$.

A coupling of a Markov chain is a joint process $(\boldsymbol{X}_t, \boldsymbol{Y}_t)_{t \ge 0}$ such that $(\boldsymbol{X}_t)_{t \ge 0}$ and $(\boldsymbol{Y}_t)_{t \ge 0}$ marginally follow the same transition rule as the original chain. Consider the following type of couplings.

▶ **Definition 7** (one-step optimal coupling for Gibbs sampling)**.** *A coupling $(\boldsymbol{X}_t, \boldsymbol{Y}_t)_{t \ge 0}$ of Gibbs sampling on an MRF instance $\mathcal{I} = (V, E, Q, \Phi)$ is a* one-step optimal coupling *if it is constructed as follows: For $t = 1, 2, \ldots$,*

1. *pick the same random $v_t \in V$, and let $(X_t(u), Y_t(u)) \leftarrow (X_{t-1}(u), Y_{t-1}(u))$ for all $u \ne v_t$;*
2. *sample $(X_t(v_t), Y_t(v_t))$ from an optimal coupling $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_{v_t}}(\cdot,\cdot)$ of the marginal distributions $\mu_{v_t}(\cdot \mid \sigma)$ and $\mu_{v_t}(\cdot \mid \tau)$ where $\sigma = X_{t-1}(\Gamma_{v_t})$ and $\tau = Y_{t-1}(\Gamma_{v_t})$.*

The coupling $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_{v_t}}(\cdot,\cdot)$ is an *optimal coupling* of $\mu_{v_t}(\cdot \mid \sigma)$ and $\mu_{v_t}(\cdot \mid \tau)$ that attains the maximum $\Pr[\boldsymbol{x} = \boldsymbol{y}]$ for all couplings $(\boldsymbol{x}, \boldsymbol{y})$ of $\boldsymbol{x} \sim \mu_{v_t}(\cdot \mid \sigma)$ and $\boldsymbol{y} \sim \mu_{v_t}(\cdot \mid \tau)$. The coupling $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_{v_t}}(\cdot,\cdot)$ is determined by the local information $\mathcal{I}_v$ and $\sigma, \tau \in Q^{\deg(v)}$.

With such a coupling, we can establish the following relation between the Dobrushin-Shlosman condition and the rapid mixing of the Gibbs sampling [8, 9, 7, 4, 19, 12].

▶ **Proposition 8** ([4, 19])**.** *Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance with $n = |V|$, and $\Omega = Q^V$ the state space. Let $H(\sigma, \tau) \triangleq |\{v \in V \mid \sigma_v \ne \tau_v\}|$ denote the Hamming distance between $\sigma \in \Omega$ and $\tau \in \Omega$. If $\mathcal{I}$ satisfies the Dobrushin-Shlosman condition (Condition 4) with constant $\delta > 0$, then the one-step optimal coupling $(\boldsymbol{X}_t, \boldsymbol{Y}_t)_{t \ge 0}$ for Gibbs sampling (Definition 7) satisfies*

$$\forall \sigma, \tau \in \Omega: \quad \mathbb{E}\left[H(\boldsymbol{X}_t, \boldsymbol{Y}_t) \mid \boldsymbol{X}_{t-1} = \sigma \wedge \boldsymbol{Y}_{t-1} = \tau \right] \le \left(1 - \frac{\delta}{n}\right) \cdot H(\sigma, \tau),$$

and hence the mixing rate of Gibbs sampling on $\mathcal{I}$ is bounded as $\tau_{\mathsf{mix}}(\mathcal{I}, \epsilon) \le \left\lceil \frac{n}{\delta} \log \frac{n}{\epsilon} \right\rceil$.

5 Outlines of algorithm

Let $\boldsymbol{\theta} : \mathfrak{M} \to \mathbb{R}^K$ be a probabilistic inference problem that maps each MRF instance in $\mathfrak{M}$ to a K-dimensional probability vector, and let $\mathcal{E}_{\boldsymbol{\theta}}$ be its estimating function. Le $\mathcal{I} = (V, E, Q, \Phi) \in \mathfrak{M}$ be the current instance, where $n = |V|$. Our dynamic inference algorithm maintains a sequence of $N(n)$ independent samples $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N(n))} \in Q^V$ which are $\epsilon(n)$-close to the Gibbs distribution $\mu_{\mathcal{I}}$ in total variation distance and an (N, ϵ)-estimator $\hat{\boldsymbol{\theta}}(\mathcal{I})$ of $\boldsymbol{\theta}(\mathcal{I})$ such that

$$\hat{\boldsymbol{\theta}}(\mathcal{I}) = \mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \ldots, \boldsymbol{X}^{(N(n))}).$$

Upon an update request that modifies $\mathcal{I}$ to a new instance $\mathcal{I}' = (V', E', Q, \Phi') \in \mathfrak{M}$, where $n' = |V'|$, our algorithm does the followings:

- *Update the sample sequence.* Update $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N(n))}$ to a new sequence of $N(n')$ independent samples $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(N(n'))} \in Q^{V'}$ that are $\epsilon(n')$-close to $\mu_{\mathcal{I}'}$ in total variation distance, and output the difference between two sample sequences.
- *Update the estimator.* Given the difference between the two sample sequences, update $\hat{\boldsymbol{\theta}}(\mathcal{I})$ to $\hat{\boldsymbol{\theta}}(\mathcal{I}') = \mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(N(n'))})$ by accessing the oracle in Definition 3.

Obviously, the updated estimator $\hat{\boldsymbol{\theta}}(\mathcal{I}')$ is an (N, ϵ)-estimator for $\boldsymbol{\theta}(\mathcal{I}')$.

Our main technical contribution is to give an algorithm that dynamically maintains a sequence of $N(n)$ independent samples for $\mu_{\mathcal{I}}$, while $\mathcal{I}$ itself is dynamically changing. The dynamic sampling problem was recently introduced in [14]. The dynamical sampling algorithm given there only handles update of a single vertex or edge and works only for graphical models with soft constraints.

In contrast, our dynamic sampling algorithm maintains a sequence of $N(n)$ independent samples for $\mu_{\mathcal{I}}$ within total variation distance $\epsilon(n)$, while the entire specification of the graphical model $\mathcal{I}$ is subject to dynamic update (to a new $\mathcal{I}'$ with difference $d(\mathcal{I}, \mathcal{I}') \leq L = o(n)$). Specifically, the algorithm updates the sample sequence within expected time $O(\Delta^2 N(n) L \log^3 n + \Delta n)$. Note that the extra $O(\Delta n)$ cost is necessary for just editing the current MRF instance $\mathcal{I}$ to $\mathcal{I}'$ because a single update may change all the vertex and edge potentials simultaneously. This incremental time cost dominates the time cost of the dynamic inference algorithm, and is efficient for maintaining $N(n)$ independent samples, especially when $N(n)$ is sufficiently large, e.g. $N(n) = \Omega(n/L)$, in which case the average incremental cost for updating each sample is $O(\Delta^2 L \log^3 n + \Delta n / N(n)) = O(\Delta^2 L \log^3 n)$.

We illustrate the main idea by explaining how to maintain one sample. The idea is to represent the trace of the Markov chain for generating the sample by a dynamic data structure, and when the MRF instance is changed, this trace is modified to that of the new chain for generating the sample for the updated instance. This is achieved by both a set of efficient dynamic data structures and the coupling between the two Markov chains.

Specifically, let $(\boldsymbol{X}_t)_{t=0}^T$ be the Gibbs sampler chain for distribution $\mu_{\mathcal{I}}$. When the chain is rapidly mixing, starting from an arbitrary initial configuration $\boldsymbol{X}_0 \in Q^V$, after suitably many steps $\boldsymbol{X} = \boldsymbol{X}_T$ is an accurate enough sample for $\mu_{\mathcal{I}}$. At each step, $\boldsymbol{X}_{t-1}$ and $\boldsymbol{X}_t$ may differ only at a vertex v_t which is picked from V uniformly and independently at random. The evolution of the chain is fully captured by the initial state $\boldsymbol{X}_0$ and the sequence of pairs $\langle v_t, X_t(v_t) \rangle$, from $t = 1$ to $t = T$, which is called the *execution log* of the chain in the paper.

Now suppose that the current instance $\mathcal{I}$ is updated to $\mathcal{I}'$. We construct such a coupling between the original chain $(\boldsymbol{X}_t)_{t=0}^T$ and the new chain $(\boldsymbol{Y}_t)_{t=0}^T$, such that $(\boldsymbol{Y}_t)_{t=0}^T$ is a faithful Gibbs sampling chain for the updated instance $\mathcal{I}'$ given that $(\boldsymbol{X}_t)_{t=0}^T$ is a faithful chain for $\mathcal{I}$, and the difference between the two chains is small, in the sense that they have almost the same execution logs except for about $O(TL/n)$ steps, where L is the difference between $\mathcal{I}$ and $\mathcal{I}'$.

To simplify the exposition of such coupling, for now we restrict ourselves to the cases where the update to the instance $\mathcal{I}$ does not change the set of variables. Without loss of generality, we only consider the following two basic update operations that modifies $\mathcal{I}$ to $\mathcal{I}'$.

- *Graph update.* The update only adds or deletes some edges, while all vertex potentials and the potentials of unaffected edges are not changed.
- *Hamiltonian update.* The update changes (possibly all) potentials of vertices and edges, while the underlying graph remains unchanged.

The general update of graphical model can be obtained by combining these two basic operations.

Then the new chain $(\boldsymbol{Y}_t)_{t=0}^{T}$ can be coupled with $(\boldsymbol{X}_t)_{t=0}^{T}$ by using the same initial configuration $\boldsymbol{Y}_0 = \boldsymbol{X}_0$ and the same sequence $v_1, v_2, \ldots, v_T \in V$ of randomly picked vertices. And for $t = 1, 2, \ldots, T$, the transition $\langle\, v_t, Y_t(v_t) \,\rangle$ of the new chain can be generated using the same vertex v_t as in the original $(\boldsymbol{X}_t)_{t=0}^{T}$ chain, and a random $Y_t(v_t)$ generated according to a coupling of the marginal distributions of $X_t(v_t)$ and $Y_t(v_t)$, conditioning respectively on the current states of the neighborhood of v_t in $(\boldsymbol{X}_t)_{t=0}^{T}$ and $(\boldsymbol{Y}_t)_{t=0}^{T}$. Note that these two marginal distributions must be identical unless **(I)** $\boldsymbol{X}_{t-1}$ and $\boldsymbol{Y}_{t-1}$ differ from each other over the neighborhood of v_t or **(II)** the v_t itself is incident to where the models $\mathcal{I}$ and $\mathcal{I}'$ differ. The event **(II)** occurs rarely due to the following reasons.

- For graph update, the event **(II)** occurs only if v_t is incident to an updated edge. Since only L edges are updated, the event occurs in at most $O(TL/n)$ steps in expectation.
- For Hamiltonian update, all the potentials of vertices and edges can be changed, thus $\mathcal{I}, \mathcal{I}'$ may differ everywhere. The key observation is that, as the total difference between the current and updated potentials is bounded by L, we can apply a filter to first select all candidate steps where the coupling may actually fail due to the difference between $\mathcal{I}$ and $\mathcal{I}'$, which can be as small as $O(TL/n)$, and the actual coupling between $(\boldsymbol{X}_t)_{t=0}^{\infty}$ and $(\boldsymbol{Y}_t)_{t=0}^{\infty}$ is constructed with such prior.

Finally, when $\mathcal{I}$ and $\mathcal{I}'$ both satisfy the Dobrushin-Shlosman condition, the percolation of disagreements between $(\boldsymbol{X}_t)_{t=0}^{T}$ and $(\boldsymbol{Y}_t)_{t=0}^{T}$ is bounded, and we show that the two chains are almost always identically coupled as $\langle\, v_t, X_t(v_t) \,\rangle = \langle\, v_t, Y_t(v_t) \,\rangle$, with exceptions at only $O(TL/n)$ steps. The original chain $(\boldsymbol{X}_t)_{t=0}^{T}$ can then be updated to the new chain $(\boldsymbol{Y}_t)_{t=0}^{T}$ by only editing these $O(TL/n)$ local transitions $\langle\, v_t, Y_t(v_t) \,\rangle$ which are different from $\langle\, v_t, X_t(v_t) \,\rangle$. This is aided by the dynamic data structure for the execution log of the chain, which is of independent interest.

6 Dynamic Gibbs sampling

In this section, we give the dynamic sampling algorithm that updates the sample sequences.

In the following theorem, we use $\mathcal{I} = (V, E, Q, \Phi)$, where $n = |V|$, to denote the current MRF instance and $\mathcal{I}' = (V', E', Q, \Phi')$, where $n' = |V'|$, to denote the updated MRF instance. And define

$$\begin{aligned} d_{\mathsf{graph}}(\mathcal{I}, \mathcal{I}') &\triangleq |V \oplus V'| + |E \oplus E'| \\ d_{\mathsf{Hamil}}(\mathcal{I}, \mathcal{I}') &\triangleq \sum_{v \in V \cap V'} \|\phi_v - \phi'_v\|_1 + \sum_{e \in E \cap E'} \|\phi_e - \phi'_e\|_1 . \end{aligned}$$

Note that $d(\mathcal{I}, \mathcal{I}') = d_{\mathsf{graph}}(\mathcal{I}, \mathcal{I}') + d_{\mathsf{Hamil}}(\mathcal{I}, \mathcal{I}')$, where $d(\mathcal{I}, \mathcal{I}')$ is defined in (2).

▶ **Theorem 9** (dynamic sampling algorithm). *Let $N : \mathbb{N}^+ \to \mathbb{N}^+$ and $\epsilon : \mathbb{N}^+ \to (0,1)$ be two functions satisfying the bounded difference condition in Definition 3. Assume that $\mathcal{I}$ and $\mathcal{I}'$ both satisfy Dobrushin-Shlosman condition, $d_{\mathit{graph}}(\mathcal{I},\mathcal{I}') \le L_{\mathsf{graph}} = o(n)$ and $d_{\mathit{Hamil}}(\mathcal{I},\mathcal{I}') \le L_{\mathsf{Hamil}}$.*

There is an algorithm that maintains a sequence of $N(n)$ independent samples $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N(n))} \in Q^V$ where $d_{\mathrm{TV}}\left(\mu_{\mathcal{I}}, \boldsymbol{X}^{(i)}\right) \le \epsilon(n)$ for all $1 \le i \le N(n)$, using $O\left(nN(n)\log n\right)$ memory words, each of $O(\log n)$ bits, such that when $\mathcal{I}$ is updated to $\mathcal{I}'$, the algorithm updates the sequence to $N(n')$ independent samples $\boldsymbol{Y}^{(1)},\dots,\boldsymbol{Y}^{(N(n'))} \in Q^{V'}$ where $d_{\mathrm{TV}}\left(\mu_{\mathcal{I}'}, \boldsymbol{Y}^{(i)}\right) \le \epsilon(n')$ for all $1 \le i \le N(n')$, within expected time cost

$$O\left(\Delta^2(L_{\mathsf{graph}} + L_{\mathsf{Hamil}})N(n)\log^3 n + \Delta n\right), \tag{5}$$

where $\Delta = \max\{\Delta_G, \Delta_{G'}\}$, and $\Delta_G, \Delta_{G'}$ denote the maximum degree of $G = (V,E)$ and $G' = (V',E')$.

Our algorithm is based on the Gibbs sampling algorithm. Let $N : \mathbb{N}^+ \to \mathbb{N}^+$ and $\epsilon : \mathbb{N}^+ \to (0,1)$ be two functions in Theorem 9. We first give the *single-sample dynamic Gibbs sampling algorithm* (Algorithm 2) that maintains a single sample $\boldsymbol{X} \in Q^V$ for the current MRF instance $\mathcal{I} = (V,E,Q,\Phi)$ where $n = |V|$ such that $d_{\mathrm{TV}}\left(\boldsymbol{X}, \mu_{\mathcal{I}}\right) \le \epsilon(n)$. We then use this algorithm to obtain the *multi-sample dynamic Gibbs sampling algorithm* that maintains $N(n)$ independent samples for the current instance.

Given the error function $\epsilon : \mathbb{N}^+ \to (0,1)$, suppose that $T(\mathcal{I})$ is an easy-to-compute integer-valued function that upper bounds the mixing time on instance $\mathcal{I}$, such that

$$T(\mathcal{I}) \ge \tau_{\mathsf{mix}}(\mathcal{I}, \epsilon(n)), \tag{6}$$

where $\tau_{\mathsf{mix}}(\mathcal{I}, \epsilon(n))$ denotes the mixing rate for the Gibbs sampling chain $(\boldsymbol{X}_t)_{t\ge 0}$ on instance $\mathcal{I}$. By Proposition 8, if the Dobrushin-Shlosman condition is satisfied, we can set

$$T(\mathcal{I}) = \left\lceil \frac{n}{\delta} \log \frac{n}{\epsilon(n)} \right\rceil. \tag{7}$$

Our algorithm for single-sample dynamic Gibbs sampling maintains a random process $(\boldsymbol{X}_t)_{t=0}^T$, which is a Gibbs sampling chain on instance $\mathcal{I}$ of length $T = T(\mathcal{I})$, where $T(\mathcal{I})$ satisfies (6). Clearly $\boldsymbol{X}_T$ is a sample for $\mu_{\mathcal{I}}$ with $d_{\mathrm{TV}}\left(\boldsymbol{X}_T, \mu_{\mathcal{I}}\right) \le \epsilon(n)$.

When the current instance $\mathcal{I}$ is updated to a new instance $\mathcal{I}' = (V',E',Q,\Phi')$ where $n' = |V'|$, the original process $(\boldsymbol{X}_t)_{t=0}^T$ is transformed to a new process $(\boldsymbol{Y}_t)_{t=0}^{T'}$ such that the following holds as an invariant: $(\boldsymbol{Y}_t)_{t=0}^{T'}$ is a Gibbs sampling chain on $\mathcal{I}'$ with $T' = T(\mathcal{I}')$. Hence $\boldsymbol{Y}_T$ is a sample for the new instance $\mathcal{I}'$ with $d_{\mathrm{TV}}\left(\boldsymbol{Y}_T, \mu_{\mathcal{I}'}\right) \le \epsilon(n')$. This is achieved through the following two steps:

1. We construct couplings between $(\boldsymbol{X}_t)_{t=0}^T$ and $(\boldsymbol{Y}_t)_{t=0}^{T'}$, so that the new process $(\boldsymbol{Y}_t)_{t=0}^{T'}$ for $\mathcal{I}'$ can be obtained by making small changes to the original process $(\boldsymbol{X}_t)_{t=0}^T$ for $\mathcal{I}$.
2. We give a data structure which represents $(\boldsymbol{X}_t)_{t=0}^T$ incrementally and supports various updates and queries to $(\boldsymbol{X}_t)_{t=0}^T$ so that the above coupling can be generated efficiently.

The data structure is provided in the full version. In the following, we give the couplings.

6.1 Coupling for dynamic instances

The Gibbs sampling chain $(\boldsymbol{X}_t)_{t=0}^T$ can be uniquely and fully recovered from: the initial state $\boldsymbol{X}_0 \in Q^V$, and the pairs $\langle v_t, X_t(v_t)\rangle_{t=1}^T$ that record the transitions. We call $\langle v_t, X_t(v_t)\rangle_{t=1}^T$ the *execution-log* for the chain $(\boldsymbol{X}_t)_{t=0}^T$, and denote it with

$$\textsf{Exe-Log}(\mathcal{I}, T) \triangleq \langle v_t, X_t(v_t)\rangle_{t=1}^T.$$

The following invariants are assumed for the random execution-log with an initial state.

▶ **Condition 10** (invariants for Exe-Log). *Fixed an initial state $\boldsymbol{X}_0 \in Q^V$, the followings hold for the random execution-log* $\textsf{Exe-Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ *for the Gibbs sampling chain $(\boldsymbol{X}_t)_{t=0}^{T}$ on instance $\mathcal{I} = (V, E, Q, \Phi)$:*

- *$T = T(\mathcal{I})$ where $T(\mathcal{I})$ satisfies (6);*
- *the random process $(\boldsymbol{X}_t)_{t=0}^{T}$ uniquely recovered from the transitions $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ and the initial state $\boldsymbol{X}_0$, is identically distributed as the Gibbs sampling (Algorithm 1) on instance $\mathcal{I}$ starting from initial state $\boldsymbol{X}_0$ with v_t as the vertex picked at the t-th step.*

Such invariants guarantee that $\boldsymbol{X}_T$ provides a sample for $\mu_{\mathcal{I}}$ with $d_{\mathrm{TV}}(\boldsymbol{X}_T, \mu_{\mathcal{I}}) \le \epsilon(|V|)$.

Suppose the current instance $\mathcal{I}$ is updated to a new instance $\mathcal{I}'$. We construct couplings between the execution-log $\textsf{Exe-Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ with initial state $\boldsymbol{X}_0 \in Q^V$ for $\mathcal{I}$ and the execution-log $\textsf{Exe-Log}(\mathcal{I}', T') = \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T'}$ with initial state $\boldsymbol{Y}_0 \in Q^{V'}$ for $\mathcal{I}'$. Our goal is as follows: assuming Condition 10 for $\boldsymbol{X}_0$ and $\textsf{Exe-Log}(\mathcal{I}, T)$, the same condition should hold invariantly for $\boldsymbol{Y}_0$ and $\textsf{Exe-Log}(\mathcal{I}', T')$.

Unlike traditional coupling of Markov chains for the analysis of mixing time, where the two chains start from arbitrarily distinct initial states but proceed by the same transition rule, here the two chains $(\boldsymbol{X}_t)_{t=0}^{T}$ and $(\boldsymbol{Y}_t)_{t=0}^{T}$ start from similar states but have to obey different transition rules due to differences between instances $\mathcal{I}$ and $\mathcal{I}'$.

Due to the technical reason, we divide the update from $\mathcal{I} = (V, E, Q, \Phi)$ to $\mathcal{I}' = (V', E', Q, \Phi')$ into two steps: we first update $\mathcal{I} = (V, E, Q, \Phi)$ to

$$\mathcal{I}_{\textsf{mid}} = (V, E, Q, \Phi^{\textsf{mid}}), \tag{8}$$

where the potentials $\Phi^{\textsf{mid}} = (\phi_a^{\textsf{mid}})_{a \in V \cup E}$ in the middle instance $\mathcal{I}_{\textsf{mid}}$ are defined as

$$\forall a \in V \cup E, \quad \phi_a^{\textsf{mid}} \triangleq \begin{cases} \phi'_a & \text{if } a \in V' \cup E' \\ \phi_a & \text{if } a \notin V' \cup E'; \end{cases}$$

then we update $\mathcal{I}_{\textsf{mid}} = (V, E, Q, \Phi^{\textsf{mid}})$ to $\mathcal{I}' = (V', E', Q, \Phi')$. In other words, the update from $\mathcal{I}$ to $\mathcal{I}_{\textsf{mid}}$ is only caused by updating the potentials of vertices and edges, while the underlying graph remains unchanged; and the update from $\mathcal{I}_{\textsf{mid}}$ to $\mathcal{I}'$ is only caused by updating the underlying graph, i.e. adding vertices, deleting vertices, adding edges and deleting edges.

The dynamic Gibbs sampling algorithm can be outlined as follows.

- $\textsf{UpdateHamiltonian}$: update $\boldsymbol{X}_0$ and $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ to a new initial state $\boldsymbol{Z}_0$ and a new execution log $\textsf{Exe-Log}(\mathcal{I}_{\textsf{mid}}, T) = \langle u_t, Z_t(u_t)\rangle_{t=1}^{T}$ such that the random process $(\boldsymbol{Z}_t)_{t=0}^{T}$ is the Gibbs sampling on instance $\mathcal{I}_{\textsf{mid}}$.
- $\textsf{UpdateGraph}$: update $\boldsymbol{Z}_0$ and $\langle u_t, Z_t(u_t)\rangle_{t=1}^{T}$ to a new initial state $\boldsymbol{Y}_0$ and a new execution log $\textsf{Exe-Log}(\mathcal{I}', T) = \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T}$ such that the random process $(\boldsymbol{Y}_t)_{t=0}^{T}$ is the Gibbs sampling on instance $\mathcal{I}'$.
- $\textsf{LengthFix}$: change the length of the execution log $\langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T}$ from T to T', where $T' = T(\mathcal{I}')$ and $T(\mathcal{I}')$ satisfies (6).

The dynamic Gibbs sampling algorithm is given in Algorithm 2.

The subroutine $\textsf{LengthFix}$ is given in Algorithm 3. The subroutine $\textsf{UpdateGraph}$ is provided in the full version of the paper. In the following, we give the subroutines $\textsf{UpdateHamiltonian}$.

We consider the update of changing potentials of vertices and edges. The update do not change the underlying graph. Let $\mathcal{I} = (V, E, Q, \Phi)$ be the current MRF instance. Let $\boldsymbol{X}_0$ and $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ be the current initial state and execution log such that the random process $(\boldsymbol{X}_t)_{t=0}^{T}$ is the Gibbs sampling on instance $\mathcal{I}$. Upon such an update, the new instance becomes

Algorithm 2 Dynamic Gibbs sampling.

```
Data    : $\boldsymbol{X}_0 \in Q^V$ and $\textsf{Exe-Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ for current $\mathcal{I} = (V, E, Q, \Phi)$.
Update: an update that modifies $\mathcal{I}$ to $\mathcal{I}' = (V', E', Q, \Phi')$.
compute $T' = T(\mathcal{I}')$ satisfying (6) and construct $\mathcal{I}_{\textsf{mid}} = (V', E', Q, \Phi^{\textsf{mid}})$ as in (8);
$\left(\boldsymbol{Z}_0, \langle u_t, Z_t(u_t)\rangle_{t=1}^{T}\right) \leftarrow \textsf{UpdateHamiltonian}\left(\mathcal{I}, \mathcal{I}_{\textsf{mid}}, \boldsymbol{X}_0, \langle v_t, X_t(v_t)\rangle_{t=1}^{T}\right)$;
// update the potentials: $\mathcal{I} \to \mathcal{I}_{\textsf{mid}}$
$\left(\boldsymbol{Y}_0, \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T}\right) \leftarrow \textsf{UpdateGraph}\left(\mathcal{I}_{\textsf{mid}}, \mathcal{I}', \boldsymbol{Z}_0, \langle u_t, Z_t(u_t)\rangle_{t=1}^{T}\right)$;
// update the underlying graph: $\mathcal{I}_{\textsf{mid}} \to \mathcal{I}'$
$\left(\boldsymbol{Y}_0, \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T'}\right) \leftarrow \textsf{LengthFix}\left(\mathcal{I}', \boldsymbol{Y}_0, \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T}, T'\right)$, where $T' = T(\mathcal{I}')$ ;
// change the length of the execution log from $T$ to $T' = T(\mathcal{I}')$
update the data to $\boldsymbol{Y}_0$ and $\textsf{Exe-Log}(\mathcal{I}', T') = \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T'}$;
```

Algorithm 3 $\textsf{LengthFix}\left(\mathcal{I}, \boldsymbol{X}_0, \langle v_t, X_t(v_t)\rangle_{t=1}^{T}, T'\right)$.

```
Data    : $\boldsymbol{X}_0 \in Q^V$ and $\textsf{Exe-Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ for current $\mathcal{I} = (V, E, Q, \Phi)$.
Input   : the new length $T' > 0$.
if $T' < T$ then
    truncate $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ to $\langle v_t, X_t(v_t)\rangle_{t=1}^{T'}$;
else
    extend $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ to $\langle v_t, X_t(v_t)\rangle_{t=1}^{T'}$ by simulating the Gibbs sampling chain
     on $\mathcal{I}$ for $T - T'$ more steps;
update the data to $\boldsymbol{X}_0$ and $\textsf{Exe-Log}(\mathcal{I}, T') = \langle v_t, X_t(v_t)\rangle_{t=1}^{T'}$
```

$\mathcal{I}' = (V, E, Q, \Phi')$. The algorithm $\textsf{UpdateHamiltonian}(\mathcal{I}, \mathcal{I}', \boldsymbol{X}_0, \langle v_t, X_t(v_t)\rangle_{t=1}^{T})$ updates the data to $\boldsymbol{Y}_0$ and $\langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T}$ such that the random process $(\boldsymbol{Y}_t)_{t=0}^{T}$ is the Gibbs sampling on instance $\mathcal{I}'$.

We transform the pair of $\boldsymbol{X}_0 \in Q^V$ and $\langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ to a new pair of $\boldsymbol{Y}_0 \in Q^V$ and $\langle v_t, Y_t(v_t)\rangle_{t=1}^{T}$ for $\mathcal{I}'$. This is achieved as follows: the vertex sequence $(v_t)_{t=1}^{T}$ is identically coupled and the chain $(\boldsymbol{X}_t)_{t=0}^{T}$ is transformed to $(\boldsymbol{Y}_t)_{t=0}^{T}$ by the following one-step local coupling between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

▶ **Definition 11** (one-step local coupling for Hamiltonian update). *The two chains $(\boldsymbol{X}_t)_{t=0}^{\infty}$ on instance $\mathcal{I} = (V, E, Q, \Phi)$ and $(\boldsymbol{Y}_t)_{t=0}^{\infty}$ on instance $\mathcal{I}' = (V, E, Q, \Phi')$ are coupled as:*

- *Initially $\boldsymbol{X}_0 = \boldsymbol{Y}_0 \in Q^V$;*
- *for $t = 1, 2, \ldots$, the two chains $\boldsymbol{X}$ and $\boldsymbol{Y}$ jointly do:*
  1. *pick the same $v_t \in V$, and let $(X_t(u), Y_t(u)) \leftarrow (X_{t-1}(u), Y_{t-1}(u))$ for all $u \in V \setminus \{v_t\}$;*
  2. *sample $(X_t(v_t), Y_t(v_t))$ from a coupling $D^{\sigma,\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(\cdot,\cdot)$ of the marginal distributions $\mu_{v_t,\mathcal{I}}(\cdot \mid \sigma)$ and $\mu_{v_t,\mathcal{I}'}(\cdot \mid \tau)$ with $\sigma = X_{t-1}(\Gamma_G(v_t))$ and $\tau = Y_{t-1}(\Gamma_G(v_t))$, where $G = (V, E)$.*

The local coupling $D^{\sigma,\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot,\cdot)$ for Hamiltonian update is specified as follows.

▶ **Definition 12** (local coupling $\boldsymbol{D^{\sigma,\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot,\cdot)}$ for Hamiltonian update). *Let $v \in V$ be vertex and $\sigma, \tau \in Q^{\Gamma_G(v)}$ two configurations, where $G = (V, E)$. We say a random pair $(c, c') \in Q^2$ is drawn from the coupling $D^{\sigma,\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot,\cdot)$ if (c, c') is generated by the following two steps:*

- ***sampling step:*** *sample* $(c, c') \in Q^2$ *jointly from an optimal coupling* $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_v}$ *of the marginal distributions* $\mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ *and* $\mu_{v,\mathcal{I}}(\cdot \mid \tau)$*, such that* $c \sim \mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ *and* $c' \sim \mu_{v,\mathcal{I}}(\cdot \mid \tau)$*;*
- ***resampling step:*** *flip a coin independently with the probability of HEADS being*

$$p^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(c') \triangleq \begin{cases} 0 & \text{if } \mu_{v,\mathcal{I}}(c' \mid \tau) \le \mu_{v,\mathcal{I}'}(c' \mid \tau), \\ \frac{\mu_{v,\mathcal{I}}(c'|\tau)-\mu_{v,\mathcal{I}'}(c'|\tau)}{\mu_{v,\mathcal{I}}(c'|\tau)} & \text{otherwise ;} \end{cases} \tag{9}$$

if the outcome of coin flipping is HEADS, resample c' *from the distribution* $\nu^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}$ *independently, where the distribution* $\nu^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}$ *is defined as*

$$\forall b \in Q: \quad \nu^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(b) \triangleq \frac{\max\{0, \mu_{v,\mathcal{I}'}(b \mid \tau) - \mu_{v,\mathcal{I}}(b \mid \tau)\}}{\sum_{x\in Q} \max\{0, \mu_{v,\mathcal{I}}(x \mid \tau) - \mu_{v,\mathcal{I}'}(x \mid \tau)\}}. \tag{10}$$

▶ **Lemma 13.** $D^{\sigma,\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot,\cdot)$ *in Definition 12 is a valid coupling between* $\mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ *and* $\mu_{v,\mathcal{I}'}(\cdot \mid \tau)$*.*

By Lemma 13, the resulting $(\boldsymbol{Y}_t)_{t=0}^T$ is a faithful copy of the Gibbs sampling on instance $\mathcal{I}'$, assuming that $(\boldsymbol{X}_t)_{t=0}^T$ is such a chain on instance $\mathcal{I}$.

Next we give an upper bound for the probability $p^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot)$ defined in (9).

▶ **Lemma 14.** *For any two instances* $\mathcal{I} = (V, E, Q, \Phi)$ *and* $\mathcal{I}' = (V, E, Q, \Phi')$ *of MRF model, and any* $v \in V, c \in Q$ *and* $\sigma \in Q^{\Gamma_G(v)}$*, it holds that*

$$p^{\tau}_{\mathcal{I}_v,\mathcal{I}'_v}(c) \le 2\left(\|\phi_v - \phi'_v\|_1 + \sum_{e=\{u,v\}\in E} \|\phi_e - \phi'_e\|_1\right), \tag{11}$$

where $\|\phi_v - \phi'_v\|_1 = \sum_{c\in Q} |\phi_v(c) - \phi'_v(c)|$ *and* $\|\phi_e - \phi'_e\|_1 = \sum_{c,c'\in Q} |\phi_e(c,c') - \phi'_e(c,c')|$*.*

By Lemma 14, for each vertex $v \in V$, we define an upper bound of the probability $p^{\cdot}_{\mathcal{I}_v,\mathcal{I}'_v}(\cdot)$ as

$$p^{\mathsf{up}}_v \triangleq \min\left\{2\left(\|\phi_v - \phi'_v\|_1 + \sum_{e=\{u,v\}\in E} \|\phi_e - \phi'_e\|_1\right), 1\right\}. \tag{12}$$

With p^{up}_v, we can implement the one-step local coupling in Definition 11 as follows. We first sample each $v_i \in V$ for $1 \le i \le T$ uniformly and independently. For each vertex $v \in V$, let $T_v \triangleq \{1 \le t \le T \mid v_t = v\}$ be the set of all the steps that pick the vertex v. We select each $t \in T_v$ independently with probability p^{up}_v to construct a random subset $\mathcal{P}_v \subseteq T_v$, and let $\mathcal{P} \triangleq \bigcup_{v\in V} \mathcal{P}_v$. We then couple the two chains $(\boldsymbol{X}_t)_{t=0}^T$ and $(\boldsymbol{Y}_t)_{t=0}^T$. First set $\boldsymbol{X}_0 = \boldsymbol{Y}_0$. For each $1 \le t \le T$, we set $(X_t(u), Y_t(u)) \leftarrow (X_{t-1}(u), Y_{t-1}(u))$ for all $u \in V \setminus \{v_t\}$; then generate the random pair $(X_t(v_t), Y_t(v_t))$ by the following procedure.

- **sampling step:** Let $\sigma = X_{t-1}(\Gamma_G(v_t))$ and $\tau = Y_{t-1}(\Gamma_G(v_t))$. We draw a random pair $(c, c') \in Q^2$ from the optimal coupling $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_v}$ of the marginal distributions $\mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ and $\mu_{v,\mathcal{I}}(\cdot \mid \tau)$ such that $c \sim \mu_{v,\mathcal{I}}(\cdot \mid \sigma)$ and $c' \sim \mu_{v,\mathcal{I}}(\cdot \mid \tau)$;
- **resampling step:** If $t \notin \mathcal{P}$, set $X_t(v_t) = c$ and $Y_t(v_t) = c'$. Otherwise, set $X_t(v_t) = c$ and

$$Y_t(v_t) = \begin{cases} b \sim \nu^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}} & \text{with probability } p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c')/p^{\mathsf{up}}_{v_t} \\ c' & \text{with probability } 1 - p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c')/p^{\mathsf{up}}_{v_t}. \end{cases} \tag{13}$$

Note that $p^{\mathsf{up}}_{v_t} > 0$ if $t \in \mathcal{P}$. By Lemma 14, it must hold that $p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c') \le p^{\mathsf{up}}_{v_t}$. Hence, the probability $p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c')/p^{\mathsf{up}}_{v_t}$ is valid. Note that the probability that $Y_t(v_t)$ is set as b is

$$\Pr[Y_t(v_t) \text{ is set as } b] = \Pr\left[t \in \mathcal{P}\right] \cdot \frac{p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c')}{p^{\mathsf{up}}_{v_t}} = p^{\mathsf{up}}_{v_t} \cdot \frac{p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c')}{p^{\mathsf{up}}_{v_t}} = p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(c').$$

Hence, our implementation perfectly simulates the coupling in Definition 11.

Let $\mathcal{D}_t$ denote the *set of disagreements* between $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$. Formally,

$$\mathcal{D}_t \triangleq \{v \in V \mid X_t(v) \neq Y_t(v)\}.$$

Note that if $v_t \notin \Gamma_G(\mathcal{D}_{t-1})$, the random pair (c, c') drawn from the coupling $D^{\sigma,\tau}_{\mathsf{opt},\mathcal{I}_v}$ must satisfy $c = c'$. Thus it is easy to make the following observation for the $(\boldsymbol{X}_t)_{t=0}^{T}$ and $(\boldsymbol{Y}_t)_{t=0}^{T}$ coupled as above.

▶ **Observation 15.** *For any integer $t \in [1, T]$, if $v_t \notin \Gamma^+_G(\mathcal{D}_{t-1})$ and $t \notin \mathcal{P}$, then $X_t(v_t) = Y_t(v_t)$ and $\mathcal{D}_t = \mathcal{D}_{t-1}$.*

With this observation, the new $\boldsymbol{Y}_0$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}', T) = \langle v_t, Y_t(v_t)\rangle_{t=1}^{T}$ can be generated from $\boldsymbol{X}_0$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ as Algorithm 4.

Observation 15 says that the nontrivial coupling between $X_t(v_t)$ and $Y_t(v_t)$ is only needed when $v_t \in \Gamma^+_G(\mathcal{D}_{t-1})$ or $t \in \mathcal{P}$, which occurs rarely as long as $\mathcal{D}_{t-1}$ and $\mathcal{P}$ are small. This is a key to ensure the small incremental time cost of Algorithm 4. The following lemma bounds the expected cost for UpdateHamiltonian.

▶ **Lemma 16** (cost of the coupling for UpdateHamiltonian). *Let $\mathcal{I} = (V, E, Q, \Phi)$ be the current MRF instance and $\mathcal{I}' = (V, E, Q, \Phi')$ the updated instance. Assume that $\mathcal{I}$ satisfies Dobrushin-Shlosman condition (Condition 4) with constant $\delta > 0$, and $d_{\mathsf{Hamil}}(\mathcal{I}, \mathcal{I}') = \sum_{v\in V} \|\phi_v - \phi'_v\|_1 + \sum_{e\in E} \|\phi_e - \phi'_e\|_1 \le L$. It holds that $\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\left[t \in \mathcal{P} \vee v_t \in \Gamma^+_G(\mathcal{D}_{t-1})\right]\right] = O\left(\frac{\Delta T L}{n\delta}\right)$, where $n = |V|$, Δ is the maximum degree of graph $G = (V, E)$.*

6.2 Dynamic Gibbs sampling algorithm

The couplings constructed in Section 6.1 can be implemented as the algorithm for dynamic Gibbs sampling. Recall $d_{\mathsf{graph}}(\cdot,\cdot)$ and $d_{\mathsf{Hamil}}(\cdot,\cdot)$ are defined in (2).

▶ **Lemma 17** (single-sample dynamic Gibbs sampling algorithm). *Let $\epsilon : \mathbb{N}^+ \to (0, 1)$ be an error function. Let $\mathcal{I} = (V, E, Q, \Phi)$ be an MRF instance with $n = |V|$ and $\mathcal{I}' = (V', E', Q, \Phi')$ the updated instance with $n' = |V'|$. Denote $T = T(\mathcal{I})$, $T' = T(\mathcal{I}')$ and $T_{\max} = \max\{T, T'\}$. Assume $d_{\mathsf{graph}}(\mathcal{I}, \mathcal{I}') \le L_{\mathsf{graph}} = o(n)$, $d_{\mathsf{Hamil}}(\mathcal{I}, \mathcal{I}') \le L_{\mathsf{Hamil}}$, and $T, T' \in \Omega(n \log n)$. The single-sample dynamic Gibbs sampling algorithm (Algorithm 2) does the followings:*

- *(**space cost**) The algorithm maintains an explicit copy of a sample $\boldsymbol{X} \in Q^V$ for the current instance $\mathcal{I}$, and also a data structure using $O(T)$ memory words, each of $O(\log T)$ bits, for representing an initial state $\boldsymbol{X}_0 \in Q^V$ and an execution-log $\mathsf{Exe\text{-}Log}(\mathcal{I}, T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ for the Gibbs sampling $(\boldsymbol{X}_t)_{t=0}^{T}$ on $\mathcal{I}$ generating sample $\boldsymbol{X} = \boldsymbol{X}_T$.*
- *(**correctness**) Assuming that Condition 10 holds for $\boldsymbol{X}_0$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}, T)$ for the Gibbs sampling on $\mathcal{I}$, upon each update that modifies $\mathcal{I}$ to $\mathcal{I}'$, the algorithm updates $\boldsymbol{X}$ to an explicit copy of a sample $\boldsymbol{Y} \in Q^{V'}$ for the new instance $\mathcal{I}'$, and correspondingly updates the $\boldsymbol{X}_0$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}, T)$ represented by the data structure to a $\boldsymbol{Y}_0 \in Q^{V'}$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}', T') = \langle v'_t, Y_t(v'_t)\rangle_{t=1}^{T'}$ for the Gibbs sampling $(\boldsymbol{Y}_t)_{t=0}^{T'}$ on $\mathcal{I}'$ generating the new sample $\boldsymbol{Y} = \boldsymbol{Y}_{T'}$, where $\boldsymbol{Y}_0$ and $\mathsf{Exe\text{-}Log}(\mathcal{I}', T')$ satisfy Condition 10 for the Gibbs sampling on $\mathcal{I}'$, therefore,*

$$d_{\mathrm{TV}}\left(\boldsymbol{Y}, \mu_{\mathcal{I}'}\right) \le \epsilon(n').$$

Algorithm 4 UpdateHamiltonian $\left(\mathcal{I},\mathcal{I}',\boldsymbol{X}_0,\langle v_t, X_t(v_t)\rangle_{t=1}^{T}\right)$.

```
Data    : $\boldsymbol{X}_0 \in Q^V$ and Exe-Log$(\mathcal{I},T) = \langle v_t, X_t(v_t)\rangle_{t=1}^{T}$ for $\mathcal{I} = (V,E,Q,\Phi)$.
Update: an update that modifies $\mathcal{I}$ to $\mathcal{I}' = (V,E,Q,\Phi')$.
$t_0 \leftarrow 0$, $\mathcal{D} \leftarrow \varnothing$, and construct a $\boldsymbol{Y}_0 \leftarrow \boldsymbol{X}_0$;
for each $v \in V$, construct a random subset $\mathcal{P}_v \subseteq T_v \triangleq \{1 \le t \le T \mid v_t = v\}$ such that
  each element in $T_v$ is selected independently with probability $p^{\mathsf{up}}_v$ defined in (12);
construct the set $\mathcal{P} \leftarrow \bigcup_{v \in V} \mathcal{P}_v$;
while $\exists t_0 < t \le T$ such that $v_t \in \Gamma^{+}_{G}(\mathcal{D})$ or $t \in \mathcal{P}$ do
    find the smallest $t > t_0$ such that $v_t \in \Gamma^{+}_{G}(\mathcal{D})$ or $t \in \mathcal{P}$;
    for all $t_0 < i < t$, let $Y_i(v_i) = X_i(v_i)$;
    sample $Y_t(v_t) \in Q$ conditioning on $X_t(v_t)$ according to the optimal coupling
      between $\mu_{v_t,\mathcal{I}}(\cdot \mid X_{t-1}(\Gamma_G(v_t)))$ and $\mu_{v_t,\mathcal{I}}(\cdot \mid Y_{t-1}(\Gamma_G(v_t)))$;
    if $t \in \mathcal{P}$ then
        with probability $p^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}(Y_t(v_t))/p^{\mathsf{up}}_{v_t}$ where $\tau = Y_{t-1}(\Gamma_G(v_t))$ do
            resample $Y_t(v_t) \sim \nu^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}$, where $\nu^{\tau}_{\mathcal{I}_{v_t},\mathcal{I}'_{v_t}}$ is defined in (10) ;
    if $X_t(v_t) \neq Y_t(v_t)$ then $\mathcal{D} \leftarrow \mathcal{D} \cup \{v_t\}$ else $\mathcal{D} \leftarrow \mathcal{D} \setminus \{v_t\}$;
    $t_0 \leftarrow t$;
for all remaining $t_0 < i \le T$: let $Y_i(v_i) = X_i(v_i)$;
update the data to $\boldsymbol{Y}_0$ and Exe-Log$(\mathcal{I}',T) = \langle v_t, Y_t(v_t)\rangle_{t=1}^{T}$;
```

- *(**time cost**) Assuming Condition 10 for $\boldsymbol{X}_0$ and* Exe-Log$(\mathcal{I},T)$ *for the Gibbs sampling on $\mathcal{I}$, the expected time complexity for resolving an update is*

$$O\left(\Delta n + \Delta\left(|T - T'| + \left(\Delta \log n + \frac{T_{\max}}{n}\right)(L_{\mathsf{Hamil}} + L_{\mathsf{graph}})\right)\log^2 T_{\max}\right),$$

  where $\Delta = \max\{\Delta_G, \Delta_{G'}\}$, Δ_G,$\Delta_{G'}$ denote the maximum degrees of $G = (V,E)$ and $G' = (V',E')$.

We remark that the $O(\Delta n)$ in time cost is necessary because the update from $\mathcal{I}$ to $\mathcal{I}'$ may change all the potentials of vertices and edges. One can reduce the $O(\Delta n)$ from the time cost if we further restrict that one update can only change constant number of vertices, edges, and potentials.

One can extend Algorithm 2 to an *Multi-sample dynamic Gibbs sampling algorithm* that maintains multiple independent random samples for the current MRF instance. By Lemma 17, it is easy to prove that the Multi-sample algorithm is correct and efficient. Thus Theorem 9 follows immediately. The detail of the Multi-sample dynamic Gibbs sampling algorithm and the proof of Theorem 9 are provided in the full version of the paper.

7 Conclusion

In this paper we study probabilistic inference problem in a graphical model when the model itself is changing dynamically with time. We study the non-local updates so that two consecutive graphical models may differ everywhere as long as the total amount of their difference is bounded. This general setting covers many typical applications. We give a sampling-based dynamic inference algorithm that maintains an inference solution efficiently against the dynamic inputs. The algorithm significantly improves the time cost compared to the static sampling-based inference algorithm.

Our algorithm generically reduces the dynamic inference to dynamic sampling problem. Our main technical contribution is a dynamic Gibbs sampling algorithm that maintains random samples for graphical models dynamically changed by non-local updates. Such technique is extendable to all single-site dynamics. This gives us a systematic approach for transforming classic MCMC samplers on static inputs to the sampling and inference algorithms in a dynamic setting. Our dynamic algorithms are efficient as long as the one-step optimal coupling exhibits a step-wise decay, a key property that has been widely used in supporting efficient MCMC sampling in the classic static setting and captured by the Dobrushin-Shlosman condition.

Our result is the first one that shows the possibility of efficient probabilistic inference in dynamically changing graphical models (especially when the graphical models are changed by non-local updates). Our dynamic inference algorithm has potentials in speeding up the iterative algorithms for learning graphical models, which deserves more theoretical and experimental research. In this paper, we focus on discrete graphical models and sampling-based inference algorithms. Important future directions include considering more general distributions and the dynamic algorithms based on other inference techniques.

References

1 Ittai Abraham, David Durfee, Ioannis Koutis, Sebastian Krinninger, and Richard Peng. On fully dynamic graph sparsifiers. In *FOCS*, 2016.

2 Osvaldo Anacleto, Catriona Queen, et al. Dynamic chain graph models for time series network data. *Bayesian Anal.*, 12(2):491–509, 2017.

3 Aaron Bernstein and Shiri Chechik. Deterministic decremental single source shortest paths: beyond the $o(mn)$ bound. In *STOC*, 2016.

4 Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *FOCS*, 1997.

5 Carlos M. Carvalho and Mike West. Dynamic matrix-variate graphical models. *Bayesian Anal.*, 2(1):69–97, 2007.

6 Christopher De Sa, Kunle Olukotun, and Christopher Ré. Ensuring rapid mixing and low bias for asynchronous Gibbs sampling. In *ICML*, 2016.

7 RL Dobrushin and SB Shlosman. Completely analytical interactions: constructive description. *J. Statist. Phys.*, 46(5-6):983–1014, 1987.

8 Roland L Dobrushin and Senya B Shlosman. Completely analytical Gibbs fields. In *Statistical Physics and Dynamical Systems*, pages 371–403. Springer, 1985.

9 Roland Lvovich Dobrushin and Senya B Shlosman. Constructive criterion for the uniqueness of Gibbs field. In *Statistical Physics and Dynamical Systems*, pages 347–370. Springer, 1985.

10 David Durfee, Yu Gao, Gramoz Goranci, and Richard Peng. Fully dynamic effective resistances. *arXiv preprint*, 2018. `arXiv:1804.04038`.

11 David Durfee, Yu Gao, Gramoz Goranci, and Richard Peng. Fully dynamic spectral vertex sparsifiers and applications. In *STOC*, 2019.

12 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Dobrushin conditions and systematic scan. *Combin. Probab. Comput.*, 17(6):761–779, 2008.

13 Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.

14 Weiming Feng, Nisheeth K Vishnoi, and Yitong Yin. Dynamic sampling from graphical models. In *STOC*, 2019.

15 Sebastian Forster and Gramoz Goranci. Dynamic low-stretch trees via dynamic low-diameter decompositions. In *STOC*, pages 377–388, 2019.

16 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6):50, 2015.

17 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combin. Probab. Comput.*, 25(04):500–559, 2016.

18 Gramoz Goranci, Monika Henzinger, and Pan Peng. Dynamic Effective Resistances and Approximate Schur Complement on Separable Graphs. In *ESA*, volume 112, 2018.

19 Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *FOCS*, 2006.

20 Monika Henzinger, Sebastian Krinninger, and Danupon Nanongkai. Decremental single-source shortest paths on undirected graphs in near-linear total update time. In *FOCS*, 2014.

21 Monika Henzinger, Sebastian Krinninger, and Danupon Nanongkai. Dynamic approximate all-pairs shortest paths: Breaking the $O(mn)$ barrier and derandomization. *SIAM J. Comput.*, 45(3):947–1006, 2016.

22 Geoffrey E Hinton. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade*, pages 599–619. Springer, 2012.

23 Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.

24 Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.*, 43:169–188, 1986.

25 Daphne Koller, Nir Friedman, and Francis Bach. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

26 Holden Lee, Oren Mangoubi, and Nisheeth Vishnoi. Online sampling from log-concave distributions. In *NIPS*, 2019.

27 David A Levin and Yuval Peres. *Markov chains and mixing times.* American Mathematical Soc., 2017.

28 Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures & Algorithms*, 15(3-4):229–241, 1999.

29 Marc Mezard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

30 Danupon Nanongkai, Thatchaphol Saranurak, and Christian Wulff-Nilsen. Dynamic minimum spanning forest with subpolynomial worst-case update time. In *FOCS*, 2017.

31 Hariharan Narayanan and Alexander Rakhlin. Efficient sampling from time-varying log-concave distributions. *J. Mach. Learn. Res.*, 18(1):4017–4045, 2017.

32 Catriona M. Queen and Jim Q. Smith. Multiregression dynamic models. *J. Roy. Statist. Soc. Ser. B*, 55(4):849–870, 1993.

33 Cedric Renggli, Bojan Karlaš, Bolin Ding, Feng Liu, Kevin Schawinski, Wentao Wu, and Ce Zhang. Continuous integration of machine learning models: A rigorous yet practical treatment. In *SysML*, 2019.

34 Padhraic Smyth, Max Welling, and Arthur U Asuncion. Asynchronous distributed learning of topic models. In *NIPS*, 2009.

35 Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *J. ACM*, 56(3):18, 2009.

36 Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets: Part II. Technical Report TR-99-003, International Computer Science Institute, 1999.

37 Martin J. Wainwright and Michael I. Jordan. *Graphical models, exponential families, and variational inference.* Now Publishers Inc, 2008.

38 Christian Wulff-Nilsen. Fully-dynamic minimum spanning forest with improved worst-case update time. In *STOC*, 2017.