中国科学: 信息科学 2022 年 第 52 卷 第 2 期: 287–313

SCIENTIA SINICA Informationis

《中国科学》杂志社
SCIENCE CHINA PRESS

CrossMark
click for updates

论文

分布式梅特罗波利斯算法: 收敛条件与最优并行加速

凤维明*, 尹一通*

南京大学计算机软件新技术国家重点实验室, 南京 210023

* 通信作者. E-mail: fengwm@smail.nju.edu.cn, yinyt@nju.edu.cn

收稿日期: 2021–04–15; 修回日期: 2021–06–02; 接受日期: 2021–07–07; 网络出版日期: 2022–01–20

国家重点研发计划重点专项 (批准号: 2018YFB1003202) 资助

摘要 梅特罗波利斯算法 (Metropolis algorithm) 是一种基本的马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 采样技术, 可用于从概率图模型所表示的高维概率分布 (即吉布斯 (Gibbs) 分布) 中进行随机采样. 传统的梅特罗波利斯算法是一个串行算法. 关于其快速收敛性的一个经典结论是: 当满足梅特罗波利斯算法的 Dobrushin-Shlosman 条件时, 该算法在 $O(n \log n)$ 步内快速收敛, 其中 n 是随机变量的个数. 本文研究了梅特罗波利斯算法的分布式版本 —— 局部梅特罗波利斯算法. 对该算法的正确性与收敛速度进行了分析, 证明了该算法总是收敛于正确的吉布斯分布; 并且对于一类自然的不包含三角形 (triangle-free) 概率图模型, 如果满足相同的 Dobrushin-Shlosman 条件, 则局部梅特罗波利斯算法在 $O(\log n)$ 轮内快速收敛. 相比于传统的串行算法, 实现了 $\Omega(n)$ 倍的渐进最优并行加速比. 具体应用包括图染色、硬核模型和伊辛 (Ising) 模型的分布式采样算法.

关键词 分布式采样, 马尔可夫链蒙特卡洛, 混合时间, 自旋系统, 耦合

1 引言

概率图模型 (probabilistic graphical model) 定义的吉布斯分布 (Gibbs distribution) 是一类重要的概率分布. 给定一个变量的集合, 概率图模型用一系列局部约束定义所有变量整体的联合分布. 吉布斯分布在计算机科学、机器学习、统计物理和概率论等领域有着重要的应用 [1]. 自旋系统 (spin system) 是一类广泛使用的概率图模型, 它通过变量之间的两两相互作用来定义整个系统. 令 $G = (V, E)$ 为一张简单无向图. 图中每个点 $v \in V$ 表示一个在 $[q] \triangleq \{0, 1, 2, \ldots, q-1\}$ 上取值的随机变量, 并且关联了一个 q 维外场 (external field) 向量 $\boldsymbol{b}_v \in \mathbb{R}^q_{\geqslant 0}$. 图中每条边 $e \in E$ 关联了一个 $q \times q$ 的对称矩阵 $\boldsymbol{A}_e \in \mathbb{R}^{q\times q}_{\geqslant 0}$ 用于规定变量之间的相互作用关系, 因此 $\boldsymbol{A}_e$ 也被称为相互作用矩阵 (interaction matrix).

引用格式: 凤维明, 尹一通. 分布式梅特罗波利斯算法: 收敛条件与最优并行加速. 中国科学: 信息科学, 2022, 52: 287–313, doi: 10.1360/SSI-2021-0127
Feng W M, Yin Y T. Distributed Metropolis algorithm: convergence condition and optimal parallel speed-up (in Chinese). Sci Sin Inform, 2022, 52: 287–313, doi: 10.1360/SSI-2021-0127

Algorithm 1 Metropolis algorithm for a spin system

Input: A spin system defined on the graph $G=(V,E)$ with the Gibbs distribution μ, where each variable takes a value from the domain $[q]$, the external fields are $(\boldsymbol{b}_v)_{v\in V}$ and the interaction matrices are $(\boldsymbol{A}_e)_{e\in E}$; and an integer $T>0$.

1: Let $\boldsymbol{X}$ be an arbitrary configuration in $[q]^V$ ($\boldsymbol{X}$ can be infeasible, i.e., $\mu(\boldsymbol{X})=0$);
2: **for** t from 1 to T **do**
3: Sample $v\in V$ uniformly at random ;
4: Sample a random candidate value $c_v\in[q]$ according to the distribution $\boldsymbol{b}_v$;
5: With the probability $\prod_{e=\{u,v\}\in E}A_e(c_v,X_u)$, let $X_v\leftarrow c_v$; otherwise, keep X_v unchanged;
6: **end for**
7: **return** $\boldsymbol{X}$.

自旋系统可以定义吉布斯分布 μ. 对于每一种配置 (configuration) $\boldsymbol{\sigma}\in[q]^V$, 它在吉布斯分布中出现的概率 $\mu(\boldsymbol{\sigma})$ 正比于其所对应的权重 $w(\boldsymbol{\sigma})$:

$$\mu(\boldsymbol{\sigma})\propto w(\boldsymbol{\sigma})\triangleq\prod_{v\in V}b_v(\sigma_v)\prod_{e=\{u,v\}\in E}A_e(\sigma_u,\sigma_v).$$

根据定义, 对向量 $\boldsymbol{b}_v$ 和矩阵 $\boldsymbol{A}_e$ 做归一化操作不改变吉布斯分布. 不失一般性地, 我们假设每个向量 $\boldsymbol{b}_v$ 是一个 $[q]$ 上的概率分布, 每个矩阵 $\boldsymbol{A}_e$ 的最大值为 1, 即

$$\forall v\in V,\quad \sum_{c\in[q]}b_v(c)=1;\qquad \forall e\in E,\quad \max_{c,c'\in[q]}A_e(c,c')=1. \tag{1}$$

自旋系统可以描述很多重要的概率分布, 例如图上合法染色 (coloring) 的均匀分布、硬核模型 (hardcore model)、统计物理学中的伊辛模型 (Ising model) 和玻茨模型 (Potts model) 等 [1].

自旋系统的采样算法是理论计算机科学的重要研究课题. 马尔可夫链蒙特卡洛 (Markov chain Monte Carlo, MCMC) 方法是研究最为深入的采样算法之一 [2,3]. 梅特罗波利斯算法 (Metropolis algorithm) 是一类经典的 MCMC 算法. 算法 1 描述了针对自旋系统的一种梅特罗波利斯算法实现.

算法 1 定义了状态空间 $[q]^V$ 上的一个马尔可夫链. 马尔可夫链从任意初始状态 $\boldsymbol{X}$ 出发, 初始状态未必合法1). 算法一共进行 T 步更新, 每次更新随机选择一个点 v; 从概率分布 $\boldsymbol{b}_v$ 采样一个候选取值; 再以 $\prod_{e=\{u,v\}\in E}A_e(c_v,X_u)$ 的概率接受候选取值, 以 $1-\prod_{e=\{u,v\}\in E}A_e(c_v,X_u)$ 的概率拒绝候选取值. 对于一大类自旋系统, 当 $T\to\infty$ 时, 算法 1 可以收敛到吉布斯分布 μ. 在使用算法 1 时, 我们可以根据自旋系统设定一个足够大的 T, 使得算法输出的分布和目标分布 μ 足够接近.

包括算法 1 在内的许多传统的 MCMC 采样算法都是串行算法. 在大数据时代, 分布式计算模型下的采样算法越来越受到重视. 在分布式机器学习领域, 人们提出一系列分布式采样算法, 有的已经在实践中取得了较好的效果 [4~13]. 近年来, 又有一些工作提出了有严格理论保障的分布式采样算法 [14~19], 其中就包括算法 1 的一种分布式版本 —— 局部梅特罗波利斯算法 (local-Metropolis algorithm) [14~16].

文献 [14] 首次在分布式计算的 LOCAL 模型 [20] 上研究采样算法. LOCAL 模型用简单无向图 $G=(V,E)$ 表示一个通信网络. 图上的每个节点对应一个处理器, 每个处理器有唯一身份 (unique ID, UID), 每条无向边 $e=\{u,v\}$ 对应 u 和 v 之间的双向通信信道. 初始每个节点 $v\in V$ 只能访问自己和邻居的 UID, 本地输入以及独立生成的随机串. LOCAL 模型是一个理想的同步通信模型, 通信按照同

1) 如果一个状态 $\boldsymbol{X}\in[q]^V$ 满足 $\mu(\boldsymbol{X})>0$, 则我们称 $\boldsymbol{X}$ 相对于 μ 合法 (feasible), 否则我们称 $\boldsymbol{X}$ 相对于 μ 非法 (infeasible).

步的轮 (round) 进行. 在每一轮中, 每个点先利用已知信息进行本地计算, 再和邻居通信交换信息. 一个 LOCAL 算法的复杂度是通信的总轮数. 局部梅特罗波利斯算法是一种通用的自旋系统分布式采样算法. 该算法最原始的版本由文献 [14] 提出, 之后文献 [15,16] 在图染色这一具体问题上给出了算法的改进版本. 在本文中, 我们把此类算法统称为局部梅特罗波利斯算法.

文献 [15,16] 中的算法可以由图染色问题推广到一般的自旋系统. 令图 $G=(V,E)$ 为一个 LOCAL 模型的通信网络. 考虑一个定义在 G 上的自旋系统 $\mathcal{I}=(V,E,[q],(\boldsymbol{b}_v)_{v\in V},(\boldsymbol{A}_e)_{e\in E})$. 令 $0\leqslant p\leqslant 1$ 为一个实数参数, $T>0$ 为一个整数参数, p 和 T 的具体取值将在之后确定. 每个点 $v\in V$ 的输入包含取值范围 $[q]$, 向量 $\boldsymbol{b}_v$, 满足 $v\in e$ 的矩阵 $\boldsymbol{A}_e$ 以及参数 p 和 T. 局部梅特罗波利斯算法初始令每个点 $v\in V$ 从 Q 中任意取一个值 $X_v\in Q$; 之后, 算法执行 T 次转移, 假设当前状态为 $\boldsymbol{X}\in[q]^V$, 每次转移算法执行如下更新操作:

- 每个点 $v\in V$ 独立地以概率 p 变为活跃状态, 否则点 v 变成休眠状态.
- 所有活跃的点 $v\in V$ 独立地从概率分布 $\boldsymbol{b}_v$ 中采样一个随机取值 $c_v\in Q$.
- 每条边 $e=\{u,v\}\in E$ 变为活跃状态当且仅当 u 和 v 中至少一个点处于活跃状态, 所有活跃的边 $e\in E$ 独立地抛一枚硬币, 正面向上的概率 p_e 为

$$p_e=\begin{cases}A_e(c_u,c_v)A_e(c_u,X_v)A_e(X_u,c_v), & \text{如果 } u,v \text{ 都活跃},\\ A_e(c_u,X_v), & \text{如果 } u \text{ 活跃}, v \text{ 休眠},\\ A_e(X_u,c_v), & \text{如果 } u \text{ 休眠}, v \text{ 活跃}.\end{cases}\tag{2}$$

- 对于所有点 $v\in V$, 如果点 v 活跃且和 v 相关的所有边的抛硬币结果都是正面向上, 则点 v 更新 $X_v\leftarrow c_v$; 否则点 v 保持 X_v 不变.

算法执行完 T 次转移后结束, 每个点输出当前的 $X_v\in[q]$, 整个网络输出随机配置 $\boldsymbol{X}=(X_v)_{v\in V}$. 通过设置参数 p 和 T, 算法可以控制输出 $\boldsymbol{X}$ 的概率分布, 使之和自旋系统的吉布斯分布 μ 足够接近. 容易看出, 局部梅特罗波利斯算法可以在 LOCAL 模型上实现, 每次转移的复杂度为 $O(1)$ 轮, 因此算法的总复杂度为 $O(T)$. 局部梅特罗波利斯算法的伪代码在算法 2 中给出.

当参数 $p=1$ 时, 算法 2 为文献 [14] 中的原始局部梅特罗波利斯算法. 当参数 $0<p<1$ 时, 把算法 2 应用到图染色问题上, 就可得到文献 [15,16] 中的改进版算法. 文献 [14] 的原始算法的收敛条件较为苛刻, 以均匀采样图的合法 q 染色为例, 如果颜色数 q 和图的最大度数 Δ 满足 $q\geqslant(2+\sqrt{2}+\delta)\Delta$, 其中 $\delta>0$ 是一个常数, 则原始局部梅特罗波利斯算法以 $O(\log n)$ 的时间复杂度近似均匀采样图的合法染色, 其中 $n=|V|$ 是图的点数. 文献 [15,16] 对图染色问题提出了改进版算法并且证明了如果 $q\geqslant(2+\delta)\Delta$, 则通过合理设置参数 $0<p<1$, 算法 2 能以 $O(\log n)$ 的时间复杂度近似均匀采样图的合法染色. $q\geqslant(2+\delta)\Delta$ 是图染色问题 MCMC 算法的经典收敛条件 [21], 在此条件下算法 1 的混合时间 (mixing time) 为 $O(n\log n)$. 因此这个结论说明: 如果图染色问题满足特罗波利斯算法的经典收敛条件, 则局部梅特罗波利斯算法可以做到 $\Omega(n)$ 倍的最优加速.

梅特罗波利斯算法 (算法 1) 是一类基本的串行采样算法, 而局部梅特罗波利斯算法 (算法 2) 是它的并行化版本. 先前的研究主要关注图染色这一具体问题, 一个自然的问题是, 在何种条件下, 局部梅特罗波利斯算法能对一般自旋系统做到 $\Omega(n)$ 倍的最优并行加速. 目前已知的分布式采样算法相对较少, 局部梅特罗波利斯算法是一种重要的分布式马尔可夫链, 因此研究它在一般模型上性能是分布式采样理论的一个重要课题. 本文的主要贡献是给出了局部梅特罗波利斯算法的一种分析, 我们证明了如下结论:

Algorithm 2 Distributed Metropolis algorithm

Input: Each vertex $v \in V$ receives the domain $[q]$, external field $\boldsymbol{b}_v$, all interaction matrices $\boldsymbol{A}_e$ with $v \in e$, a real number $0 \leqslant p \leqslant 1$ and an integer T.

Each $v \in V$ sets X_v to an arbitrary value in $[q]$;
for t from 1 to T **do**
  foreach $v \in V$ **do**
    Become active with probability p; become lazy otherwise;
  foreach active vertex $v \in V$ **do**
    Sample a random candidate value $c_v \in [q]$ according to distribution $\boldsymbol{b}_v$;
  foreach edge $\{u, v\} \in E$ where both u and v are active **do**
    Pass check with probability $A_e(c_u, c_v)A_e(c_u, X_v)A_e(X_u, c_v)$;
  foreach edge $\{u, v\} \in E$ where u is active and v is lazy **do**
    Pass check with probability $A_e(c_u, X_v)$;
  foreach active vertex $v \in V$ **do**
    if all edges incident to v passed their checks **then**
      $X_v \leftarrow c_v$;
    end if
end for
Each $v \in V$ outputs X_v.

- **正确性结论** (定理 1): 对于一般的自旋系统, 如果梅特罗波利斯算法 (算法 1) 可以收敛到目标吉布斯分布, 则局部梅特罗波利斯算法 (算法 1) 也可以收敛到目标吉布斯分布.
- **收敛性结论** (定理 3): 对于定义在不含三角形 (triangle-free) 的图上的自旋系统, 如果自旋系统满足串行梅特罗波利斯算法的 Dobrushin-Shlosman 条件 (条件 1) 以及一个额外条件 (条件 2), 则局部梅特罗波利斯算法的混合时间为 $O(\log n)$, 其中 n 表示自旋系统的点数. 对于很多重要自旋系统, 额外条件 (条件 2) 不仅弱于串行梅特罗波利斯算法的 Dobrushin-Shlosman 条件, 还弱于系统的唯一性条件 (uniqueness condition). 因此我们有如下推论.
- **具体自旋系统上的应用**(推论 1): 对于定义在不含三角形的图上的图染色问题, 硬核模型以及伊辛模型, 如果自旋系统满足串行梅特罗波利斯算法的 Dobrushin-Shlosman 条件 (条件 2), 则局部梅特罗波利斯算法的混合时间为 $O(\log n)$.

Dobrushin-Shlosman 条件是串行马尔可夫链收敛的重要条件 [22~24], 串行梅特罗波利斯算法在此条件下的混合时间为 $O(n \log n)$. 我们的结论说明, 对于满足条件 2 的自旋系统, 如果自旋系统所在的图不含三角形且自旋系统满足梅特罗波利斯算法的 Dobrushin-Shlosman 条件, 则局部梅特罗波利斯算法以 $O(\log n)$ 轮收敛到目标分布. 根据分布式采样下界结论 [14,18], $O(\log n)$ 轮也是分布式采样算法的最优时间复杂度. 这说明对一大类自旋系统, 在串行梅特罗波利斯算法快速收敛的经典条件下, 局部梅特罗波利斯算法也能快速收敛, 并且收敛速度有 $\Omega(n)$ 倍的最优并行加速.

论文结构. 第 2 节严格给出本文的主要结论. 第 3 节给出预备知识. 第 4 节证明正确性结论. 第 5 节证明收敛性结论. 第 6 节把主要结论应用到具体的自旋系统上. 第 7 节给出总结.

2 主要结论

令 $\mathcal{I} = (V, E, [q], \boldsymbol{b}, \boldsymbol{A})$ 为一个定义在图 $G = (V, E)$ 上的自旋系统, 变量取值范围为 $[q]$, 所有外场矩阵为 $\boldsymbol{b} = (\boldsymbol{b}_v)_{v \in V}$, 所有相互作用矩阵为 $\boldsymbol{A} = (\boldsymbol{A}_e)_{e \in E}$. 令 μ 为 $\mathcal{I}$ 定义的吉布斯分布, 令 $\Omega \triangleq [q]^V$ 为 μ 的样本空间. 对于任意配置 $\boldsymbol{\sigma} \in \Omega$, 如果 $\mu(\boldsymbol{\sigma}) > 0$, 则称 $\boldsymbol{\sigma}$ 合法 (feasible); 如果 $\mu(\boldsymbol{\sigma}) = 0$,

则称 $\boldsymbol{\sigma}$ 非法 (infeasible). 本文对比梅特罗波利斯算法 (算法 1) 和局部梅特罗波利斯算法 (算法 2) 的收敛条件. 两种算法都是状态空间 Ω 上的马尔可夫链. 我们总是假设算法 1 可以收敛到吉布斯分布 μ. 严格地, 我们用 $\boldsymbol{M} \in \mathbb{R}_{\geqslant 0}^{\Omega\times\Omega}$ 表示梅特罗波利斯算法的转移矩阵; 对于任意点 $v \in V$, 我们用 $\Gamma_v = \Gamma(v) \triangleq \{u \in V \mid \{u,v\} \in E\}$ 表示 v 在 G 上邻居的集合; 本文只考虑满足如下假设的自旋系统 $\mathcal{I} = (V, E, [q], \boldsymbol{b}, \boldsymbol{A})$.

假设1　对于任意 $v \in V$, 任意 $\boldsymbol{\sigma} \in [q]^{\Gamma_v}$ 都有

$$\sum_{c\in[q]} \left(b_v(c) \prod_{u\in\Gamma_v} A_{uv}(c, \sigma_u) \right) > 0. \tag{3}$$

并且梅特罗波利斯算法 (算法 1) 在 $\mathcal{I}$ 的合法配置空间上不可约 (irreducible), 即

$$\forall \text{ 合法配置 } \boldsymbol{\sigma}, \boldsymbol{\tau} \in \Omega, \quad \exists t \in \mathbb{Z}_{>0}, \quad \text{使得 } M^t(\boldsymbol{\sigma}, \boldsymbol{\tau}) > 0. \tag{4}$$

不等式 (3) 说明给定 v 邻居上任意一种配置 $\boldsymbol{\sigma} \in [q]^{\Gamma_v}$, 总存在取值 c 使得 v 满足所有局部约束, 因此从任何初始状态出发, 算法 1 一定可以收敛到合法状态. 条件 (4) 说明算法 1 在合法状态上连通. 假设 1 保证了算法 1 最终可以收敛到吉布斯分布 μ. 此假设对很多自然的自旋系统成立, 例如满足 $q \geqslant \Delta + 2$ 的图 q – 染色问题 (Δ 表示图的最大度数)、硬核模型, 以及所有由柔性约束 (soft constraint)[2)]的定义的自旋系统.

我们的第 1 个结论说明局部梅特罗波利斯算法的正确性.

定理1　给定任意一个吉布斯分布为 μ 的自旋系统, 以及任意参数 $0 < p < 1$, 局部梅特罗波利斯算法 $(\boldsymbol{X}_t)_{t\geqslant 0}$ 从任意初始状态 $\boldsymbol{X}_0 \in \Omega$ 出发, 当 $t \to \infty$ 时, $\boldsymbol{X}_t$ 的分布收敛到 μ.

文献 [15,16] 对图染色问题证明了算法 2 的正确性. 定理 1 把结论推广到任意自旋系统.

我们的主要结论是算法 2 在一般自旋系统上的快速收敛条件. 我们证明在算法 1 快速收敛的一种充分条件下, 算法 2 也可以快速收敛. 算法 1 和 2 都是状态空间 Ω 上的马尔可夫链, 它们的收敛速度被各自的混合时间刻画. 令 $\boldsymbol{P} \in \mathbb{R}_{\geqslant 0}^{\Omega\times\Omega}$ 表示算法 1 (或算法 2) 的转移矩阵, 则算法 1 (或算法 2) 的混合时间定义为

$$\forall 0 < \epsilon < 1, \quad T_{\mathrm{mix}}(\epsilon) = \max_{\boldsymbol{X}_0\in\Omega} \min\{t \mid d_{\mathrm{TV}}\left(P^t(\boldsymbol{X}_0, \cdot), \mu\right) \leqslant \epsilon\},$$

$P^t(\boldsymbol{X}_0, \cdot)$ 表示马尔可夫链从 $\boldsymbol{X}_0$ 出发, 转移 t 步后的概率分布, $d_{\mathrm{TV}}\left(P^t(\boldsymbol{X}_0, \cdot), \mu\right) \triangleq \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\Omega} |P^t(\boldsymbol{X}_0, \boldsymbol{\sigma}) - \mu(\boldsymbol{\sigma})|$ 表示 $P^t(\boldsymbol{X}_0, \cdot)$ 和 μ 的总变差 (total variation distance).

自旋系统局部节点之间的相互影响和串行马尔可夫链的混合时间有着密切的联系 [22~25]. 对于算法 1 中的串行梅特罗波利斯算法, 我们可以按照如下规则定义节点之间的影响. 考虑梅特罗波利斯算法的一步转移. 算法首先随机等概率选择一个点 $v \in V$. 假设当前配置为 $\boldsymbol{\sigma} \in \Omega$. 令 $\pi_v^{\boldsymbol{\sigma}}$ 为梅特罗波利斯算法一步转移之后, 点 v 上取值的概率分布, 严格地,

$$\forall c \in [q], \quad \pi_v^{\boldsymbol{\sigma}}(c) \triangleq \begin{cases} b_v(c) \prod\limits_{u\in\Gamma_v} A_{uv}(c, \sigma_u), & \text{如果 } c \neq \sigma_v, \\ 1 - \sum\limits_{a\neq\sigma_v} b_v(a) \prod\limits_{u\in\Gamma_v} A_{uv}(a, \sigma_u), & \text{如果 } c = \sigma_v. \end{cases} \tag{5}$$

2) 如果一个自旋系统满足对于任意 $e \in E$, 任意 $c, c' \in [q]$ 都有 $A_e(c, c') > 0$, 则我们称该自旋系统由柔性约束定义, 例如伊辛模型和玻茨模型. 此类系统满足任意 $\boldsymbol{X} \in [q]^V$ 都是合法状态.

定义1 (梅特罗波利斯算法的影响矩阵和总影响) 对于任意点 $u,v\in V$ (u 可以等于 v), 定义 u 对 v 的影响 (influence) 为

$$\rho(u,v)\triangleq\max_{\boldsymbol{\sigma},\boldsymbol{\tau}\in\Omega_u} d_{\mathrm{TV}}\left(\pi_v^{\boldsymbol{\sigma}},\pi_v^{\boldsymbol{\tau}}\right), \tag{6}$$

其中 $\Omega_u\triangleq\{(\boldsymbol{\sigma},\boldsymbol{\tau})\in\Omega\times\Omega\mid\sigma_u\neq\tau_u\wedge\forall w\neq u,\sigma_w=\tau_w\}$ 表示所有只在 u 一点处取值不同的状态对 $(\boldsymbol{\sigma},\boldsymbol{\tau})$. 定义单点产生的总影响 (total influence) 为

$$\alpha\triangleq\max_{u\in V}\sum_{v\in V}\rho(u,v). \tag{7}$$

条件1 (梅特罗波利斯算法的 Dobrushin-Shlosman 条件) 存在常数 $0<\delta<1$ 使得

$$\alpha\leqslant 1-\delta.$$

马尔可夫链的经典结论说明, 如果自旋系统满足梅特罗波利斯算法的 Dobrushin-Shlosman 条件, 则算法 1 可以快速收敛 [22].

定理2 ([22]) 令 $\mathcal{I}$ 为一个定义在图 $G=(V,E)$ 上的自旋系统. 如果存在常数 $\delta>0$ 使得 $\alpha\leqslant 1-\delta$, 则梅特罗波利斯算法 (算法 1) 的混合时间为 $T_{\mathrm{mix}}(\epsilon)=O(n\log\frac{n}{\epsilon})$, 其中 $n=|V|$, $O(\cdot)$ 记号隐藏了一个只和 δ 有关的常数因子.

矩阵 $\boldsymbol{\rho}$ 衡量了在梅特罗波利斯算法 (算法 1) 的转移规则下, 所有点对之间的相互影响. 定理 2 指出, 只要单点发出的总影响 $\alpha<1$, 则算法 1 可以快速收敛. 对于一般的单点更新算法 (single-site dynamics), 我们也可以按照转移规则类似地定义分布 $\pi_\cdot^\cdot(\cdot)$ 和影响矩阵 $\boldsymbol{\rho}$, 只要满足其所对应的 Dobrushin-Shlosman 条件, 相应的马尔可夫链都可以快速收敛 [22~24] . 最经典的 Dobrushin-Shlosman 条件 [26, 27] 由吉布斯采样算法 (Gibbs sampling, 又称 Glauber dynamics) 的转移规则定义, 在这种定义下, 分布 $\pi_v^{\boldsymbol{\sigma}}$ 为 $\mu_v(\cdot\mid\sigma_{V\setminus\{v\}})$, 其中 $\mu_v(\cdot\mid\sigma_{V\setminus\{v\}})$ 表示在给定 $\sigma_{V\setminus\{v\}}$ 的条件下, 吉布斯分布 μ 在点 v 上的边缘分布.

局部梅特罗波利斯算法 (算法 2) 是梅特罗波利斯算法的分布式版本. 一个自然的问题是: 在经典的 Dobrushin-Shlosman 条件下, 局部梅特罗波利斯算法是否也能快速收敛? 本文对这个问题给出了肯定的回答, 但是我们额外要求图 G 不含三角形 (triangle-free) 并且自旋系统满足如下条件.

条件2 存在一个常数 $C>0$ 使得对于任意点 $v\in V$, 任意和 v 关联的边 $e=\{u,v\}\in E$,

$$\forall c_u\in[q],\quad\sum_{c_v\in[q]}b_v(c_v)A_e(c_v,c_u)\geqslant 1-\frac{C}{\Delta},$$

其中 Δ 是图 $G=(V,E)$ 的最大度数.

条件 2 说明, 对于任意边 $\{u,v\}\in E$, 任意 u 的当前值 c_u, 只要 v 按照外场采样一个随机值, 则边 $\{u,v\}$ 上的约束能以 $1-O(\frac{1}{\Delta})$ 的概率被满足. 在一些典型的自旋系统上, 条件 2 是一个非常弱的条件, 不光弱于定理 2 中的收敛条件 $\alpha<1$, 甚至弱于对应系统的唯一性条件 (uniqueness condition).

• 图染色问题. 给定一个图 $G=(V,E)$ 以及颜色集合 $[q]=\{0,1,\ldots,q-1\}$, 吉布斯分布 μ 为 G 上所有合法图染色的均匀分布. 条件 2 等价于

$$\exists\text{常数 } C'>0\ \text{ 使得 } q\geqslant C'\Delta,$$

而图染色问题的唯一性条件为 $q \geqslant \Delta + 1$ [28, 29].

• 硬核模型. 给定一个图 $G=(V,E)$ 以及一个逸度参数 $\lambda \in \mathbb{R}_{\geqslant 0}$, 吉布斯分布 μ 为 G 上所有独立集的加权分布, 每种独立集 $I \subseteq V$ 出现的概率为 $\mu(I) \propto \lambda^{|I|}$. 条件 2 等价于

$$\exists \text{ 常数 } C' > 0 \text{ 使得 } \lambda \leqslant \frac{C'}{\Delta},$$

而硬核模型的唯一性条件为 $\lambda < \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^{\Delta}} \approx \frac{\mathrm{e}}{\Delta}$ [30, 31].

• 伊辛模型. 给定一个图 $G=(V,E)$ 以及一个温度参数 $\beta \in \mathbb{R}$, 吉布斯分布 μ 定义在 $\{-1,+1\}^V$ 上, 对于任意 $\boldsymbol{\sigma} \in \{-1,+1\}^V$ 都有 $\mu(\boldsymbol{\sigma}) \propto \exp(\beta \sum_{\{u,v\}\in E} \sigma_u \sigma_v)$. 条件 2 等价于

$$\exists \text{ 常数 } C' > 0 \text{ 使得 } 1 - \exp(-2|\beta|) \leqslant \frac{C'}{\Delta},$$

而伊辛模型的唯一性条件为 $1 - \exp(-2|\beta|) < \frac{2}{\Delta}$ [32, 33].

对于局部梅特罗波利斯算法 (算法 2), 我们证明了如下定理.

定理3　令 $\mathcal{I} = (V, E, [q], \boldsymbol{b}, \boldsymbol{A})$ 为一个定义在图 $G=(V,E)$ 上的自旋系统. 如果 $\mathcal{I}$ 同时满足

• 图 G 不含三角形, 即不存在 3 个点 $u, v, w \in V$ 使得 u, v, w 之间都有边;

• $\mathcal{I}$ 以常数 $C > 0$ 满足条件 2;

• 梅特罗波利斯算法的 Dobrushin-Shlosman 条件: 存在常数 $0 < \delta < 1$ 使得 $\alpha \leqslant 1 - \delta$, 则存在 $p = p(C, \delta)$ 使得以 p 为参数的局部梅特罗波利斯算法 (算法 2) 的混合时间为

$$T_{\mathrm{mix}}(\epsilon) = O\left(\log \frac{n}{\epsilon}\right),$$

其中 $n = |V|$ 为总点数, $O(\cdot)$ 记号隐藏了一个只和 δ, C 有关的常数因子.

所有满足定理 3 中条件的自旋系统都存在复杂度为 $O(\log n)$ 的分布式采样算法. 根据文献 [14,18] 中的下界结论3), $O(\log n)$ 是 LOCAL 模型上分布式采样算法的最优运行时间, 相对于串行梅特罗波利斯算法 (算法 1), 分布式算法做到了 $\Omega(n)$ 倍的最优加速比. 应用到具体模型上, 我们有如下推论.

推论1　令 $\delta > 0$ 为一个常数, $G=(V,E)$ 为一个不含三角形的图. 令 $n = |V|$ 为 G 的点数, Δ 为 G 的最大度数. 对于如下定义在 G 上的自旋系统:

• 满足 $q \geqslant (2+\delta)\Delta$ 的图染色问题;

• 满足 $\lambda \leqslant \frac{1-\delta}{\Delta}$ 的硬核模型;

• 满足 $1 - \exp(-2|\beta|) \leqslant \frac{(1-\delta)\eta}{\Delta+1}$ 的伊辛模型, 其中 $\eta \approx 0.703$ 满足方程 $\eta = \exp(-\eta/2)$;

都存在 $p = p(\delta)$ 使得以 p 为参数的局部梅特罗波利斯算法 (算法 2) 的混合时间为

$$T_{\mathrm{mix}}(\epsilon) = O\left(\log \frac{n}{\epsilon}\right),$$

其中 $O(\cdot)$ 记号隐藏了一个只和 δ 有关的常数因子.

推论 1 的证明在第 6 节给出. 在推论 1 的 3 类自旋系统中, 条件 2 都严格弱于梅特罗波利斯算法的 Dobrushin-Shlosman 条件. 因此推论 1 给出的 3 个参数条件实质上是把 Dobrushin-Shlosman 条件应用到具体模型上的结果. 因此对于这 3 类重要的自旋系统, 如果图 G 不含三角形, 那么 Dobrushin-Shlosman 条件不仅能推出串行梅特罗波利斯算法 $O(n \log n)$ 的混合时间, 也能推出局部梅特罗波利斯算法 $O(\log n)$ 的混合时间. 事实上, 对于图染色问题, 文献 [14, 16] 中的分析可以证明对于任意图

3) 文献 [14, 18] 证明下界使用的图类都不含三角形, 因此下界结论对不含三角形的图依然成立.

$G=(V,E)$, 如果 $q \geqslant (2+\delta)\Delta$, 则局部梅特罗波利斯算法的混合时间为 $O(\log n)$. 但是文献 [14,16] 中的分析技术只适用于图染色问题, 而本文为一般的自旋系统给出了一种统一的分析技术.

相对于梅特罗波利斯算法 (算法 1), 局部梅特罗波利斯算法 (算法 2) 的分析更加复杂, 这是因为算法 1 每次只更新一个点的取值, 而算法 2 一步可能更新 $\Theta(n)$ 个点的取值. 在证明正确性结论的过程中 (第 4 节), 我们需要构造一个双射来验证算法 2 的细致平衡方程 (detailed balance equation). 在证明收敛性结论的过程中 (第 5 节), 我们对算法 2 使用路径耦合分析技术. 文献 [15,16] 的分析只适用于图染色问题, 本文对一般的自旋系统设计了一种全新的路径耦合方案. 在分析耦合的过程中, 根据算法 2 的转移规则 (式 (2)), 每条边通过测试的概率相对于算法 1 有所降低, 因此需要条件 2 来保证耦合成功的概率, 同时我们还需要图 G 不含三角形的性质来克服节点之间的相关性. 分布式马尔可夫链的耦合需要更加细致而复杂的分析, 相关分析技术也是一个值得深入研究的方向.

3 预备知识

3.1 总变差与耦合

令 Ω 为一个样本空间, μ 和 ν 是定义在 Ω 上的两种概率分布. 定义 μ 和 ν 的总变差为

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

因为 μ 和 ν 是两个概率分布, 它们满足 $\sum_{x\in\Omega}\mu(x)=1$ 且 $\sum_{x\in\Omega}\nu(x)=1$, 所以总变差可以写成

$$d_{\mathrm{TV}}(\mu,\nu)=\sum_{x\in\Omega:\mu(x)>\nu(x)}(\mu(x)-\nu(x))=\sum_{y\in\Omega:\nu(y)>\mu(y)}(\nu(y)-\mu(y)). \tag{8}$$

令 $(X,Y)\in\Omega\times\Omega$ 为一对服从某种联合分布的随机变量. 如果 X 的边缘分布为 μ, Y 的边缘分布为 ν, 则 (X,Y) 称为 μ 和 ν 的一种耦合 (coupling). 随机变量的耦合和总变差之间满足如下耦合不等式 (coupling inequality).

命题1 ([3, 命题4.2]) 令 μ 和 ν 为两种概率分布. 任意一种 μ 和 ν 的耦合 (X,Y) 满足

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

且存在一种最优耦合 (optimal coupling) (X,Y) 使得 $\Pr[X\neq Y]=d_{\mathrm{TV}}(\mu,\nu)$.

在本文中, 我们会使用如下关于耦合的引理.

引理1 令 Ω 和 Ω' 为两个样本空间, $X,Y\in\Omega$ 为两个随机变量, X 的概率分布为 μ, Y 的概率分布为 ν, 令 f,g 为两个从 Ω 到 Ω' 的映射, 定义随机变量 $X'=f(X)\in\Omega'$ 和 $Y'=g(Y)\in\Omega'$, X' 的概率分布为 μ', Y' 的概率分布为 ν'. μ 和 ν 的任意一种耦合 (X,Y) 满足

$$\Pr[f(X)\neq g(Y)]\geqslant d_{\mathrm{TV}}(\mu',\nu'),$$

且存在 μ 和 ν 的一种耦合 (X,Y) 使得

$$\Pr[f(X)\neq g(Y)]=d_{\mathrm{TV}}(\mu',\nu').$$

证明　对于 μ 和 ν 的任意一种耦合 (X,Y), $(X',Y')=(f(X),g(Y))$ 构成了 μ' 和 ν' 之间的一种耦合. 根据耦合不等式, 我们有 $\Pr[f(X)\neq g(Y)]\leqslant d_{\mathrm{TV}}(\mu',\nu')$.

考虑随机变量 X',Y'. 根据耦合不等式, 存在最优耦合 (X',Y') 使得 $\Pr[X'\neq Y']=d_{\mathrm{TV}}(\mu',\nu')$. 我们先根据最优耦合采样 (X',Y'). 然后在 $f(X)=X'$ 的条件下采样 X, 在 $g(Y)=Y'$ 的条件下采样 Y. 由此可以得到耦合 (X,Y). 此耦合满足 $\Pr[f(X)\neq g(Y)]=d_{\mathrm{TV}}(\mu',\nu')$.

3.2　马尔可夫链与路径耦合

令 Ω 为一个样本空间. 令 $(X_t)_{t\geqslant 0}$ 为一个定义在 Ω 上的马尔可夫链, 设其转移矩阵为 $\boldsymbol{P}\in\mathbb{R}_{\geqslant 0}^{\Omega\times\Omega}$. 本文经常用 $\boldsymbol{P}$ 指代其所对应的马尔可夫链. 如果一个 Ω 上的概率分布 μ (μ 视为行向量) 满足 $\mu=\mu\boldsymbol{P}$, 则我们称 μ 为 $\boldsymbol{P}$ 的一种平稳分布 (stationary distribution). 如果对于任意状态 $\boldsymbol{\sigma},\boldsymbol{\tau}\in\Omega$, 都存在一个整数 $t>0$ 使得 $P^t(\boldsymbol{\sigma},\boldsymbol{\tau})>0$, 则我们称 $\boldsymbol{P}$ 不可约. 对于不可约的链 $\boldsymbol{P}$, 如果对于任意 $\boldsymbol{\sigma}\in\Omega$ 都有 $\gcd\{t>0\mid P^t(\boldsymbol{\sigma},\boldsymbol{\sigma})>0\}=1$, 则我们称 $\boldsymbol{P}$ 非周期 (aperiodic). 如果 $\boldsymbol{P}$ 既是不可约的又是非周期的, 则 $\boldsymbol{P}$ 有唯一平稳分布. 我们称 $\boldsymbol{P}$ 相对于 Ω 上的分布 μ 可逆 (reversible) 当且仅当如下细致平衡方程成立:

$$\forall\boldsymbol{\sigma},\boldsymbol{\tau}\in\Omega:\quad \mu(\boldsymbol{\sigma})P(\boldsymbol{\sigma},\boldsymbol{\tau})=\mu(\boldsymbol{\tau})P(\boldsymbol{\tau},\boldsymbol{\sigma}),$$

在此条件下, μ 是 $\boldsymbol{P}$ 的一种平稳分布.

容易验证, 梅特罗波利斯算法 (算法 1) 相对于吉布斯分布 μ 可逆, 且对于任意合法配置 $\boldsymbol{\sigma}$, 算法 1 从 $\boldsymbol{\sigma}$ 转移到 $\boldsymbol{\sigma}$ 的概率非零. 因此, 如果输入的自旋系统满足假设 1, 则算法 1 有唯一平稳分布 μ.

假设马尔可夫链 $\boldsymbol{P}$ 有唯一平稳分布 μ, 则 $\boldsymbol{P}$ 的混合时间定义为

$$\forall 0<\epsilon<1,\quad T_{\mathrm{mix}}(\epsilon)=\max_{X_0\in\Omega}\min\{t\mid d_{\mathrm{TV}}\left(P^t(\boldsymbol{X}_0,\cdot),\mu\right)\leqslant\epsilon\}.$$

马尔可夫链的耦合是混合时间的主要分析工具之一. 一个联合随机过程 $(X_t,Y_t)_{t\geqslant 0}$ 称为马尔可夫链 $\boldsymbol{P}$ 的耦合当且仅当边缘随机过程 $(X_t)_{t\geqslant 0}$ 和 $(Y_t)_{t\geqslant 0}$ 各自都服从 $\boldsymbol{P}$ 所定义的转移规则, 且如果 $X_k=Y_k$ 则对所有的 $\ell\geqslant k$ 都有 $X_\ell=Y_\ell$. 根据耦合不等式, 我们有

$$d_{\mathrm{TV}}\left(P^t(X_0,\cdot),\mu\right)\leqslant\max_{X_0,Y_0\in\Omega}d_{\mathrm{TV}}(X_t,Y_t)\leqslant\max_{X_0,Y_0\in\Omega}\Pr[X_t\neq Y_t].$$

因此, 给定一个耦合过程, 我们只要分析 $X_t\neq Y_t$ 的概率就能分析马尔可夫链的混合时间.

路径耦合 (path coupling) [22] 是一种构造马尔可夫链耦合的工具. 令 V 为变量集合, $[q]$ 为一个有限的取值范围, 在本文中, 考虑 $[q]^V$ 上的马尔可夫链. 对于任意 $\boldsymbol{X},\boldsymbol{Y}\in[q]^V$, 定义汉明距离 (Hamming distance) 为

$$d_{\mathrm{ham}}(\boldsymbol{X},\boldsymbol{Y})\triangleq|\{v\in V\mid X_v\neq Y_v\}|.$$

在本文中, 我们会使用如下路径耦合引理.

引理2 ([22])　令 $0<\delta<1$ 为一个参数. 令 V 为一个大小为 n 变量集合, $[q]$ 为一个有限的取值范围. 令 $\boldsymbol{P}$ 为定义在样本空间 $[q]^V$ 上不可约、非周期的马尔可夫链. 如果对于任意满足 $d_{\mathrm{ham}}(\boldsymbol{X},\boldsymbol{Y})=1$ 的状态对 $\boldsymbol{X},\boldsymbol{Y}\in[q]^V$ 都存在 $\boldsymbol{P}$ 一步转移的耦合 $(\boldsymbol{X},\boldsymbol{Y})\to(\boldsymbol{X}',\boldsymbol{Y}')$ 使得

$$\mathbb{E}[d_{\mathrm{ham}}(\boldsymbol{X}',\boldsymbol{Y}')\mid\boldsymbol{X},\boldsymbol{Y}]\leqslant 1-\delta,$$

则马尔可夫链 $\boldsymbol{P}$ 的混合时间满足

$$T_{\mathrm{mix}}(\epsilon) \leqslant \left\lceil \frac{1}{\delta} \log \frac{n}{\epsilon} \right\rceil.$$

利用路径耦合技术, 在构造马尔可夫链耦合的时候, 我们不需要考虑任意状态对 $\boldsymbol{X},\boldsymbol{Y}$, 而是只需要考虑仅在一点取值不同的状态对. 这可以极大地简化耦合的构造.

4 正确性分析

本节证明定理 1. 令 $\mathcal{I}=(V,E,[q],\boldsymbol{b},\boldsymbol{A})$ 为一个定义在图 $G=(V,E)$ 上的自旋系统, 变量取值范围为 $[q]$, 所有外场矩阵为 $\boldsymbol{b}=(\boldsymbol{b}_v)_{v\in V}$, 所有相互作用矩阵为 $\boldsymbol{A}=(\boldsymbol{A}_e)_{e\in E}$. 我们先证明从任意初始状态出发, 局部梅特罗波利斯算法一定可以收敛到合法状态. 因为 $\mathcal{I}$ 满足假设 1, 对于每个点 $v\in V$, 如果点 v 活跃且点 v 的所有邻居都休眠, 那么以非零的概率, 点 v 在一次马尔可夫链更新之后, 可以满足点 v, 以及和 v 关联的边上的约束, 即, $b_v(X_v)\prod_{u\in\Gamma_v}A_{uv}(X_v,X_u)>0$. 注意到局部梅特罗波利斯算法不可能把一个合法状态更新成非法状态. 所以局部梅特罗波利斯算法一定可以收敛到合法状态.

我们接着证明, 局部梅特罗波利斯算法在合法状态空间上是连通的. 令 n 为总点数. 令 $\boldsymbol{P}$ 为局部梅特罗波利斯算法的转移矩阵. 我们证明对于任何合法状态 X,Y, 存在 t 使得 $P^t(X,Y)>0$. 根据假设, 梅特罗波利斯算法在合法解空间上是连通的. 在局部梅特罗波利斯算法上, 以概率 $np(1-p)^{(n-1)}$, 只有一个点活跃且其他点都休眠. 在此条件下, 局部梅特罗波利斯算法的转移规则和梅特罗波利斯算法一致. 所以每一步以非零概率, 局部梅特罗波利斯算法可以完全模拟梅特罗波利斯算法的转移. 因此局部梅特罗波利斯算法在合法解空间上也是连通的. 注意到局部梅特罗波利斯算法每一步以概率 $(1-p)^n$ 所有点都休眠. 因此 $P(X,X)>0$, 局部梅特罗波利斯算法是非周期马尔可夫链. 所以局部梅特罗波利斯算法存在唯一平稳分布, 且平稳分布只在合法状态上有非零概率.

最后我们验证局部梅特罗波利斯算法满足细致平衡方程

$$\forall \boldsymbol{X},\boldsymbol{Y}\in[q]^V,\quad \mu(\boldsymbol{X})P(\boldsymbol{X},\boldsymbol{Y})=\mu(\boldsymbol{Y})P(\boldsymbol{Y},\boldsymbol{X}). \tag{9}$$

从而说明吉布斯分布 μ 是局部梅特罗波利斯算法的唯一平稳分布. 注意到如果 $\boldsymbol{X},\boldsymbol{Y}$ 都是非法状态, 显然式 (9) 成立. 如果一个合法一个非法, 假设 $\boldsymbol{X}$ 合法 $\boldsymbol{Y}$ 非法, 因为 $P(\boldsymbol{X},\boldsymbol{Y})=0$, 所以式 (9) 成立. 现在我们考虑 $\boldsymbol{X},\boldsymbol{Y}$ 都合法的情况. 局部梅特罗波利斯算法所定义的马尔可夫链的每一步转移有以下 3 个步骤:

- 生成一个活跃点的集合 $S\subseteq V$;
- 每一个活跃的点 $v\in V$ 按照概率分布 $\boldsymbol{b}_v$ 随机生成 $c_v\in[q]$, 记向量 $\boldsymbol{c}=(c_v)_{v\in S}$;
- 记集合 $E(S)$ 为两个端点都活跃的所有边的集合, 记集合 $\delta(S)$ 为恰好只有一个端点活跃的边的集合, 我们称 $E(S)\cup\delta(S)$ 里的边为活跃的边; 所有活跃的边 $e\in E(S)\cup\delta(S)$ 抛一枚硬币, 记 $r_e\in\{0,1\}$ 为抛硬币结果, $r_e=1$ 表示正面向上, $r_e=0$ 表示背面向上, 记向量 $\boldsymbol{r}=(r_e)_{e\in E(S)\cup\delta(S)}$.

给定当前状态, 局部梅特罗波利斯算法生成随机三元组 $(\mathbb{S},\mathbb{c},\mathbb{r})$, 再利用三元组完成转移. 记 $\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$ 是所有能把 X 转移成 Y 的三元组的集合, 则

$$P(\boldsymbol{X},\boldsymbol{Y})=\sum_{(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}}\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})].$$

同理可以对 $P(\boldsymbol{Y},\boldsymbol{X})$ 定义 $\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}$ 并写出类似的等式. 定义 $D\triangleq\{v\in V\mid X_v\neq Y_v\}$ 为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 取值不同的点的集合. 记 $E(D)$ 为两个端点都在集合 D 中的边的集合, 记 $\delta(D)$ 为恰好只有一个端点在 D 中的边的集合. 要验证细致平衡方程 (9), 我们只需要验证如下两种关系:

- $P(\boldsymbol{X},\boldsymbol{Y})=P(\boldsymbol{Y},\boldsymbol{X})=0$;
- 或 $P(\boldsymbol{X},\boldsymbol{Y}),P(\boldsymbol{Y},\boldsymbol{X})>0$ 满足以下关系:

$$\begin{aligned}\frac{P(\boldsymbol{X},\boldsymbol{Y})}{P(\boldsymbol{Y},\boldsymbol{X})}&=\frac{\sum_{(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}}\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]}{\sum_{(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}}\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]},\\ \text{验证等号 }(\star)\quad&\overset{(\star)}{=}\frac{\prod_{v\in D}b_v(Y_v)\prod_{e=E(D)\cup\delta(D)}A_e(Y_e)}{\prod_{v\in D}b_v(X_v)\prod_{e=E(D)\cup\delta(D)}A_e(X_e)}=\frac{\mu(\boldsymbol{Y})}{\mu(\boldsymbol{X})}.\end{aligned}$$

我们证明方法是先构造一个双射 $h_{\boldsymbol{X}\to\boldsymbol{Y}}:\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}\to\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}$. 然后, 对于任意的 $(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$, 记 $(S',\boldsymbol{c}',\boldsymbol{r}')=h_{\boldsymbol{X}\to\boldsymbol{Y}}(S,\boldsymbol{c},\boldsymbol{r})$, 我们只需要证明如下关系就能验证细致平衡方程

$$\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]=\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S',\boldsymbol{c}',\boldsymbol{r}')]=0, \tag{10}$$

或 $\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]$, $\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S',\boldsymbol{c}',\boldsymbol{r}')]>0$ 满足

$$\frac{\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]}{\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S',\boldsymbol{c}',\boldsymbol{r}')]}=\frac{\prod_{v\in D}b_v(Y_v)\prod_{e=E(D)\cup\delta(D)}A_e(Y_e)}{\prod_{v\in D}b_v(X_v)\prod_{e=E(D)\cup\delta(D)}A_e(X_e)}. \tag{11}$$

对于任意点 $v\in V$, 我们用 $\Gamma(v)$ 表示 v 在图 $G=(V,E)$ 上的邻居. 注意到一个三元组 $(S,\boldsymbol{c},\boldsymbol{r})$ 在集合 $\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$ 中当且仅当满足以下条件:

- $D\subseteq S$;
- 对于任意的 $v\in D, c_v=Y_v$, 且对于任意邻居 $u\in\Gamma(v)$, $r_{uv}=1$;
- 对于任意的 $v\in S\setminus D$, $c_v=X_v=Y_v$ 或存在邻居 $u\in\Gamma(v)$, $r_{uv}=0$.

双射的构造规则如下, 给定一个 $(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$, 构造 $(S',\boldsymbol{c}',\boldsymbol{r}')=h_{\boldsymbol{X}\to\boldsymbol{Y}}(S,\boldsymbol{c},\boldsymbol{r})$ 如下:

- $S'=S$;
- 对于任意的 $v\in D$, $c'_v=X_v$; 对于任意的 $v\in S\setminus D$, $c'_v=c_v$;
- 对于任意的 $e\in E(S)\cup\delta(S)$, $r'_e=r_e$.

现在证明 $h_{\boldsymbol{X}\to\boldsymbol{Y}}$ 是一个双射, 首先容易验证 $(S',\boldsymbol{c}',\boldsymbol{r}')$ 一定可以把 $\boldsymbol{Y}$ 变成 $\boldsymbol{X}$, 所以 $h_{\boldsymbol{X}\to\boldsymbol{Y}}$ 的值域是 $\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}$. 在上述定义中, 我们交换 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的位置, 可以得到一个从 $\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}$ 到 $\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$ 的映射 $h_{\boldsymbol{Y}\to\boldsymbol{X}}$. 容易验证 $h_{\boldsymbol{X}\to\boldsymbol{Y}}$ 和 $h_{\boldsymbol{Y}\to\boldsymbol{X}}$ 为互逆映射, 即对于任意 $(S,\boldsymbol{c},\boldsymbol{r})\in\Omega_{\boldsymbol{X}\to\boldsymbol{Y}}$, $h_{\boldsymbol{Y}\to\boldsymbol{X}}(h_{\boldsymbol{X}\to\boldsymbol{Y}}(S,\boldsymbol{c},\boldsymbol{r}))=(S,\boldsymbol{c},\boldsymbol{r})$; 对于任意 $(S',\boldsymbol{c}',\boldsymbol{r}')\in\Omega_{\boldsymbol{Y}\to\boldsymbol{X}}, h_{\boldsymbol{X}\to\boldsymbol{Y}}(h_{\boldsymbol{Y}\to\boldsymbol{X}}(S',\boldsymbol{c}',\boldsymbol{r}'))=(S',\boldsymbol{c}',\boldsymbol{r}')$. 所以 $h_{\boldsymbol{X}\to\boldsymbol{Y}}$ 是双射.

现在证明等式 (10) 和 (11). 由链式法则

$$\begin{aligned}&\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S,\boldsymbol{c},\boldsymbol{r})]=\Pr[\mathbb{S}=S]\cdot\Pr[\mathbb{c}=\boldsymbol{c}\mid\mathbb{S}=S]\cdot\Pr[\mathbb{r}=\boldsymbol{r}\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}];\\&\Pr[(\mathbb{S},\mathbb{c},\mathbb{r})=(S',\boldsymbol{c}',\boldsymbol{r}')]=\Pr[\mathbb{S}=S']\cdot\Pr[\mathbb{c}=\boldsymbol{c}'\mid\mathbb{S}=S']\cdot\Pr[\mathbb{r}=\boldsymbol{r}'\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}'].\end{aligned}$$

根据双射的定义, 有 $S=S'$. 因为 $0<p<1$, 则一定有 $\Pr[\mathbb{S}=S]=\Pr[\mathbb{S}=S']>0$. 又因为对于每个活跃的点 v, 随机值 c_v 是从概率分布 $\boldsymbol{b}_v$ 中独立产生的. 注意到 $\boldsymbol{X},\boldsymbol{Y}$ 都是合法状态, $c(D)=Y(D)$, $c'(D)=X(D)$, 且 $c(S\setminus D)=c'(S\setminus D)$. 于是要么 $\Pr[\mathbb{c}=\boldsymbol{c}\mid\mathbb{S}=S]=\Pr[\mathbb{c}=\boldsymbol{c}'\mid\mathbb{S}=S']=0$, 要么两者同时非零. 如果同时取 0, 则等式 (10) 成立. 我们假设

$$\Pr[\mathbb{S}=S]\cdot\Pr[\mathbb{c}=\boldsymbol{c}\mid\mathbb{S}=S]>0\text{ 且 }\Pr[\mathbb{S}=S']\cdot\Pr[\mathbb{c}=\boldsymbol{c}'\mid\mathbb{S}=S']>0.$$

于是有

$$\frac{\Pr[\mathbb{S}=S]\Pr[\mathbb{c}=\boldsymbol{c}\mid\mathbb{S}=S]}{\Pr[\mathbb{S}=S']\Pr[\mathbb{c}=\boldsymbol{c}'\mid\mathbb{S}=S']}=\frac{\Pr[\mathbb{c}=\boldsymbol{c}\mid\mathbb{S}=S]}{\Pr[\mathbb{c}=\boldsymbol{c}'\mid\mathbb{S}=S']}=\frac{\prod_{v\in D}b_v(Y_v)}{\prod_{v\in D}b_v(X_v)}.$$

最后考虑 $\Pr[\mathbb{r}=\boldsymbol{r}\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]$ 和 $\Pr[\mathbb{r}=\boldsymbol{r}'\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']$ 的比值. 因为所有的边独立地抛硬币, 我们研究每一条边上的概率的比值关系. 我们把活跃的边 $E(S)\cup\delta(S)$ 分为 3 类. 依次考虑每一类的比值:

• $E(D)$ 中的边. 两个端点都在集合 D 中的边, 这种边一定有 $r_e=r'_e=1$. 对任意 $e=\{u,v\}\in E(D)$, u,v 都是活跃的点而且 $r_e=r'_e=1$, $c_u=Y_u$, $c_v=Y_v$, $c'_u=X_u$, $c'_v=X_v$, 于是有

$$\begin{aligned}\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]&=A_e(c_u,c_v)A_e(c_u,X_v)A_e(X_u,c_v)\\&=A_e(Y_u,Y_v)A_e(Y_u,X_v)A_e(X_u,Y_v),\\\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']&=A_e(c'_u,c'_v)A_e(c'_u,Y_v)A_e(Y_u,c'_v)\\&=A_e(X_u,X_v)A_e(X_u,Y_v)A_e(Y_u,X_v).\end{aligned}$$

如果 $A_e(X_u,Y_v)A_e(Y_u,X_v)=0$, 则两个概率值 $\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]$ 和 $\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']$ 都为 0; 否则 $\frac{\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]}{\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']}=\frac{A_e(Y_e)}{A_e(X_e)}$.

• $\delta(D)$ 中的边. 一个端点在 D 中, 一个端点不在 D 中的边, 这种边一定有 $r_e=r'_e=1$. 考虑一条边 $e=\{u,v\}\in\delta(D)$, 我们假设 $u\in D$, $v\notin D$. 根据双射的定义, 一定有 u 是活跃的点, $r_e=1$, $c_u=Y_u$, $c'_u=X_u$. 根据点 v 是否为活跃的点考虑如下两种子情况. (1) 点 v 为活跃的点. 根据双射的定义, 有 $c_v=c'_v$. 因为 $v\notin D$, 所以 $X_v=Y_v$. 于是

$$\begin{aligned}\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]&=A_e(c_u,c_v)A_e(c_u,X_v)A_e(X_u,c_v)\\&=A_e(Y_u,c_v)A_e(Y_u,Y_v)A_e(X_u,c_v),\\\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']&=A_e(c'_u,c'_v)A_e(c'_u,Y_v)A_e(Y_u,c'_v)\\&=A_e(X_u,c_v)A_e(X_u,X_v)A_e(Y_u,c_v).\end{aligned}$$

如果 $A_e(Y_u,c_v)A_e(X_u,c_v)=0$, 则上述两个概率值 $\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]$ 和 $\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']$ 都为 0; 否则 $\frac{\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]}{\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']}=\frac{A_e(Y_e)}{A_e(X_e)}$. (2) 点 v 不为活跃的点. 因为 $v\in D$, 所以 $X_v=Y_v$. 于是

$$\begin{aligned}\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]&=A_e(c_u,X_v)=A_e(Y_u,X_v)=A_e(Y_u,Y_v),\\\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']&=A_e(c'_u,Y_v)=A_e(X_u,Y_v)=A_e(X_u,X_v).\end{aligned}$$

此时一定有 $\frac{\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]}{\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}']}=\frac{A_e(Y_e)}{A_e(X_e)}$.

• 既不在 $E(D)$ 也不在 $\delta(D)$ 中的边. 对于这种边 $e=\{u,v\}$, 一定有 $X_u=Y_u$, $X_v=Y_v$. 根据双射的定义, 如果 u 是活跃的点, 一定有 $c_u=c'_u$; 如果 v 是活跃的点, 一定有 $c_v=c'_v$; 如果 e 的端点存在活跃的点, 则一定有 $r_e=r'_e$. 所以

$$\Pr[\mathbb{r}_e=1\mid\mathbb{S}=S\wedge\mathbb{c}=\boldsymbol{c}]=\Pr[\mathbb{r}'_e=1\mid\mathbb{S}=S'\wedge\mathbb{c}=\boldsymbol{c}'].$$

因为所有的边独立地抛硬币, 综合以上 3 种情况, 可以验证等式 (10) 或者 (11) 成立. 所以局部梅特罗波利斯算法满足细致平衡方程. 正确性得证.

5　收敛性分析

本节用路径耦合的分析技术来证明定理 3. 为了说明耦合的构造, 我们先定义一些记号. 令 $G=(V,E)$ 为一张图. 对任意 $v\in V$, 定义

$$\Gamma(v)\triangleq\{u\in V\mid\{u,v\}\in E\},\quad \Gamma^+(v)\triangleq\Gamma(v)\cup\{v\}.$$

对于任意 $u,v\in V$, 定义 $\mathrm{dist}_G(u,v)$ 为 u 和 v 在图 G 上的最短路距离. 对于任意点 $v\in V$, 任意整数 $\ell\geqslant 0$, 定义到 v 距离恰好为 ℓ 的点集为

$$S_\ell(v)\triangleq\{u\in V\mid \mathrm{dist}_G(u,v)=\ell\},$$

注意到 $S_1(v)=\Gamma(v)$. 在一些英文文献中, 集合 $S_\ell(v)$ 被称为以 v 为中心, ℓ 为半径的球面 (sphere). 对于任意集合 $\Lambda\subseteq V$, 定义集合 Λ 的内部边集为

$$E(\Lambda)\triangleq\{\{u,v\}\in E\mid u\in\Lambda\wedge v\in\Lambda\}.$$

定义集合 Λ 的边界边集为

$$\delta(\Lambda)\triangleq\{\{u,v\}\in E\mid u,v\text{ 中有一个点属于 }\Lambda,\text{ 有一个点不属于 }\Lambda\}.$$

定义集合 Λ 所有相关边的集合为

$$E^+(\Lambda)\triangleq E(\Lambda)\cup\delta(\Lambda).$$

5.1　耦合的构造

令 $\mathcal{I}=(V,E,[q],\boldsymbol{b},\boldsymbol{A})$ 为一个定义在图 $G=(V,E)$ 上的自旋系统, 变量取值范围为 $[q]$, 所有外场矩阵为 $\boldsymbol{b}=(\boldsymbol{b}_v)_{v\in V}$, 所有相互作用矩阵为 $\boldsymbol{A}=(\boldsymbol{A}_e)_{e\in E}$. 假设算法 1 的参数为 $0<p<1$. 假设算法 1 的当前状态为 $\boldsymbol{X}\in[q]^V$. 执行转移时, 算法按照如下规则生成三元组 $(S,\boldsymbol{c},\boldsymbol{r})$:

- 每个点独立以概率 p 变成活跃状态, 记活跃点的集合为 $S\subseteq V$;
- 每个 $v\in S$ 独立地从分布 $\boldsymbol{b}_v$ 中采样 $c_v\in[q]$, 令 $\boldsymbol{c}=(c_v)_{v\in S}$;
- 对任意活跃的边 $e=\{u,v\}\in E^+(S)=E(S)\cup\delta(S)$, 独立采样 $r_e\in\{0,1\}$ 满足 $\Pr[r_e=1]=p_e$, 令 $\boldsymbol{r}=(r_e)_{e\in E^+(S)}$, 其中概率 p_e 定义为

$$p_e=\begin{cases}A_e(c_u,c_v)A_e(c_u,X_v)A_e(X_u,c_v), & \text{如果 } u,v \text{ 都活跃},\\ A_e(c_u,X_v), & \text{如果 } u \text{ 活跃}, v \text{ 休眠},\\ A_e(X_u,c_v), & \text{如果 } u \text{ 休眠}, v \text{ 活跃}.\end{cases}\tag{12}$$

显然, 局部梅特罗波利斯算法的一步转移 $\boldsymbol{X}\to\boldsymbol{X}'$ 完全由当前状态 $\boldsymbol{X}\in[q]^V$ 和三元组 $(S,\boldsymbol{c},\boldsymbol{r})$ 确定. 在下文中, 首先给出一种生成三元组 $(S,\boldsymbol{c},\boldsymbol{r})$ 的特定过程, 然后基于这个特定的生成过程构造局部梅特罗波利斯算法一步转移的耦合.

三元组生成过程 $\mathcal{P}(v_0)$. 固定一个特殊点 $v_0\in V$. 假设当前状态为 $\boldsymbol{X}\in[q]^V$. 我们把如下生成三元组的过程记为 $\mathcal{P}(v_0)$. 首先每个点独立以概率 p 变成活跃状态, 记活跃点的集合为 $S\subseteq V$. 然后, 把所有活跃的点分成 3 类 $S_{\text{中心}}\uplus S_{\text{邻居}}\uplus S_{\text{外围}}=S$:

$$S_{\text{外围}}\triangleq S\cap(V\setminus\Gamma^+(v_0)),\quad S_{\text{邻居}}\triangleq S\cap\Gamma(v_0),\quad S_{\text{中心}}\triangleq S\cap\{v_0\}.$$

3 类集合 $S_{中心}, S_{邻居}, S_{外围}$ 都有可能是空集. 接着, 我们再把活跃的边集 $E^{+}(S)$ 分成 3 类:

$$\begin{aligned}
&E_{中心\text{-}邻居} \triangleq \{\{v_0, v\} \in E \mid v_0 \in S_{中心} \vee v \in S_{邻居}\}, \\
&E_{邻居\text{-}外围} \triangleq \{\{u, v\} \in E \mid u \in \Gamma(v_0) \wedge v \in S_2(v_0) \wedge (u \in S_{邻居} \vee v \in S_{外围})\}, \\
&E_{外围} \triangleq \{\{u, v\} \in E \mid \{u, v\} \cap \Gamma(v_0) = \emptyset \wedge \{u, v\} \cap S_{外围} \neq \emptyset\}.
\end{aligned}$$

$E_{中心\text{-}邻居}$ 表示 v_0 和 v_0 的邻居之间至少有一个活跃端点的边. 如果 v_0 活跃, 则 $E_{中心\text{-}邻居} = \{\{v_0, v\} \mid v \in \Gamma(v_0)\}$; 如果 v_0 休眠, 则 $E_{中心\text{-}邻居} = \{\{v_0, v\} \mid v \in \Gamma(v_0) \wedge v \in S_{邻居}\}$. $E_{邻居\text{-}外围}$ 表示 $\Gamma(v_0)$ 和 $S_2(v_0)$ 之间至少有一个活跃端点的边, 即对于任意 $\{u, v\} \in E_{邻居\text{-}外围}$, 一定有 $u \in \Gamma(v_0)$, $v \in S_2(v_0)$ 且 u, v 中至少有一个活跃点. $E_{外围}$ 表示一个端点在 $S_{外围}$ 中, 另一端点不在 $\Gamma(v_0)$ 中的活跃的边.

引理3 如果图 $G = (V, E)$ 不含三角形, 则 $E_{中心\text{-}邻居} \uplus E_{邻居\text{-}外围} \uplus E_{外围} = E^{+}(S)$, 所以 3 种边集 $E_{中心\text{-}邻居}$, $E_{邻居\text{-}外围}$, $E_{外围}$ 构成所有活跃边的一种划分.

证明 因为图 G 不含三角形, 所以对于 v_0 任意两个不同的邻居 $u, v \in \Gamma(v_0)$, u, v 之间不存在边. 我们可以把所有边的集合 E 划分成 3 类 $E_1 = \{\{v_0, v\} \mid v \in \Gamma(v_0)\}$, $E_2 = \{\{u, v\} \mid u \in \Gamma(v_0) \wedge v \in S_2(v_0)\}$, $E_3 = E \setminus (E_1 \cup E_2)$. 容易验证 $E_{中心\text{-}邻居} = E_1 \cap E^{+}(S)$, $E_{邻居\text{-}外围} = E_2 \cap E^{+}(S)$, $E_{外围} = E^{+}(S) \cap E_3$. 因此 $E_{中心\text{-}邻居}$, $E_{邻居\text{-}外围}$, $E_{外围}$ 构成所有活跃边的一种划分.

三元组中的 $\boldsymbol{c}, \boldsymbol{r}$ 按照如下规则生成:

(1) 所有 $v \in S_{外围}$ 独立地从分布 $\boldsymbol{b}_v$ 中采样 $c_v \in [q]$; 所有边 $e \in E_{外围}$ 独立地采样 $r_e \in \{0, 1\}$, 概率满足 $\Pr[r_e = 1] = p_e$.

(2) 所有边 $e \in E_{邻居\text{-}外围}$ 采样 $r_e \in \{0, 1\}$. 考虑 $E_{邻居\text{-}外围}$ 中的边 $e = \{u, v\}$, 不失一般性地, 假设 $u \in \Gamma(v_0)$ 且 $v \in S_2(v_0)$, 根据点 u 是否活跃, 可以把 $E_{邻居\text{-}外围}$ 中的边分成两类

$$\begin{aligned}
E^{(1)}_{邻居\text{-}外围} &\triangleq \{\{u, v\} \in E_{邻居\text{-}外围} \mid u \in \Gamma(v_0) \wedge u \notin S_{邻居}\} \\
E^{(2)}_{邻居\text{-}外围} &\triangleq \biguplus_{u \in S_{邻居}} E_u \triangleq \biguplus_{u \in S_{邻居}} \{e \in E_{邻居\text{-}外围} \mid u \in e\} \\
&= \biguplus_{u \in S_{邻居}} \{\{u, v\} \mid v \in \Gamma(u) \setminus \{v_0\}\}.
\end{aligned} \tag{13}$$

所有 $e = \{u, v\} \in E^{(1)}_{邻居\text{-}外围}$ 独立采样 $r_e \in \{0, 1\}$, 概率满足

$$\Pr[r_e = 1] = p_e = A_e(X_u, c_v). \tag{14}$$

对于任意 $u \in S_{邻居}$, 因为此时 c_u 尚未生成, 所以目前无法对 $e = \{u, v\} \in E_u$ 计算 p_e, 因此对任意 $e = \{u, v\} \in E_u$ 定义函数 $q_e : [q] \to [0, 1]$:

$$\forall a \in [q], \quad q_e(a) = \begin{cases} A_e(a, c_v) A_e(a, X_v) A_e(X_u, c_v), & \text{如果 } v \text{ 活跃}, \\ A_e(a, X_v), & \text{如果 } v \text{ 休眠}. \end{cases} \tag{15}$$

根据定义, $q_e(a)$ 表示 $c_u = a$ 时的 p_e. 注意到对于 $u \in S_{邻居}$, 所有 E_u 互不相交. 任意 $u \in S_{邻居}$ 独立地执行以下操作: 所有边 $e \in E_u$ 联合采样 $(r_e)_{e \in E_u}$, 概率满足对于任意 $\boldsymbol{\sigma} \in \{0, 1\}^{E_u}$,

$$\Pr\left[\forall e \in E_u, r_e = \sigma_e\right] = \sum_{z \in [q]} b_u(z) \prod_{e \in E_u} \left(q_e(z) \cdot \mathbf{1}_{\sigma_e = 1} + (1 - q_e(z)) \cdot \mathbf{1}_{\sigma_e = 0}\right), \tag{16}$$

上式中 $\mathbf{1}_A \in \{0,1\}$ 是一个指示事件 A 是否发生的指示变量. 容易验证式 (16) 定义了 $\{0,1\}^{E_u}$ 上一个合法的概率分布.

(3) 所有点 $u \in S_{邻居}$ 独立采样 $c_u \in [q]$, 概率满足

$$\forall a \in [q], \quad \Pr[c_u = a] = \frac{b_u(a)\prod_{e\in E_u}(q_e(a)\cdot \mathbf{1}_{r_e=1} + (1-q_e(a))\cdot \mathbf{1}_{r_e=0})}{\sum_{z\in[q]} b_u(z)\prod_{e\in E_u}(q_e(z)\cdot \mathbf{1}_{r_e=1} + (1-q_e(z))\cdot \mathbf{1}_{r_e=0})}. \tag{17}$$

(4) 根据 v_0 是否为活跃的点 ($S_{中心} = \{v_0\}$ 或 $S_{中心} = \emptyset$) 分两种情况:

• 如果 $S_{中心} = \emptyset$, 则对于任意 $u \in S_{邻居}$, 令 $e = \{v_0, u\}$, 独立采样 $r_e \in \{0,1\}$, 概率满足 $\Pr[r_e = 1] = p_e = A_e(X_{v_0}, c_u)$;

• 如果 $S_{中心} = \{v_0\}$, 则从分布 $\boldsymbol{b}_{v_0}$ 中独立采样 c_{v_0}; 对于任意邻居 $u \in \Gamma(v_0)$, 令 $e = \{v_0, u\}$, 独立采样 $r_{e,1} \in \{0,1\}$, 概率满足 $\Pr[r_{e,1} = 1] = A_e(c_{v_0}, X_u)$; 对于任意 $w \in S_{邻居}$, 令 $e' = \{v_0, w\}$, 独立采样 $r_{e',2} \in \{0,1\}$, 概率满足 $\Pr[r_{e',2} = 1] = A_{e'}(c_{v_0}, c_w)A_{e'}(X_{v_0}, c_w)$. 对于任意满足 $u \in \Gamma(v_0) \setminus S_{邻居}$ 的边 $e = \{v_0, u\}$, 令 $r_e = r_{e,1}$; 对于任意满足 $w \in S_{邻居}$ 的边 $e' = \{v_0, u\}$, 令 $r'_e = r_{e',1} \wedge r_{e',2}$.

容易验证, 上述过程结束后, 所有的 $e \in E_{中心\text{-}邻居}$ 都采样了 r_e.

引理4 如果图 $G = (V, E)$ 不含三角形, 则对于任意当前配置 $\boldsymbol{X} \in [q]^V$, 任意点 $v_0 \in V$, 随机过程 $\mathcal{P}(v_0)$ 生成的三元组 $(S, \boldsymbol{c}, \boldsymbol{r})$ 和算法 1 生成的三元组同分布.

证明 假设当前状态是 $\boldsymbol{X} \in [q]^V$. 固定一个点 $v_0 \in V$. $\mathcal{P}(v_0)$ 过程会产生一个随机三元组 $(\mathbb{S}, \mathbb{c}, \mathbb{r})$. 我们证明对于任意集合 $S \subseteq V$, 任意 $\boldsymbol{c} \in [q]^S$, 任意 $\boldsymbol{r} \in \{0,1\}^{E^+(S)}$, 随机过程 $\mathcal{P}(v_0)$ 满足

$$\Pr[(\mathbb{S}, \mathbb{c}, \mathbb{r}) = (S, \boldsymbol{c}, \boldsymbol{r})] = p^{|S|}(1-p)^{|V|-|S|}\prod_{v\in S} b_v(c_v)\prod_{e\in E^+(S)} f_e(r_e), \tag{18}$$

其中 $f_e(r_e)$ 定义为

$$f_e(r_e) = \begin{cases} p_e, & \text{如果 } r_e = 1, \\ 1 - p_e, & \text{如果 } r_e = 0, \end{cases}$$

其中 p_e 定义在式 (12) 中. 这样就可以证明引理.

首先, 过程 $\mathcal{P}(v_0)$ 每个点独立地以概率 p 变成活跃状态, 有

$$\Pr[\mathbb{S} = S] = p^{|S|}(1-p)^{|V|-|S|}. \tag{19}$$

之后所有的 $v \in S_{外围}$ 从 $\boldsymbol{b}_v$ 中采样 c_v, 所有的 $e \in E_{外围}$ 采样 $r_e \in \{0,1\}$. 有

$$\Pr[\mathbb{c}(S_{外围}) = c(S_{外围}) \wedge \mathbb{r}(E_{外围}) = r(E_{外围}) \mid \mathbb{S} = S] = \prod_{v\in S_{外围}} b_v(c_v)\prod_{e\in E_{外围}} f_e(r_e).$$

接下来, 在 $\mathbb{S} = S$, $\mathbb{c}(S_{外围}) = c(S_{外围})$, $\mathbb{r}(E_{外围}) = r(E_{外围})$ 的条件下, 分析 $\mathbb{c}(S_{邻居}) = c(S_{邻居}) \wedge \mathbb{r}(E_{邻居\text{-}外围}) = r(E_{邻居\text{-}外围})$ 的概率. 根据式 (14) 可知, $\mathbb{r}(E^{(1)}_{邻居\text{-}外围}) = r(E^{(1)}_{邻居\text{-}外围})$ 会对此概率贡献一个因子 $\prod_{e\in E^{(1)}_{邻居\text{-}外围}} f_e(r_e)$. 结合式 (16) 和 (17) 可知, 对于每个 $u \in S_{邻居}$, $\mathbb{c}(u) = c(u)$ 且 $\mathbb{r}(E_u) = r(E_u)$ 会对此概率贡献一个因子,

$$b_u(c_u)\prod_{e\in E_u}(q_e(c_u)\cdot \mathbf{1}_{r_e=1} + (1-q_e(c_u))\cdot \mathbf{1}_{r_e=0}) = b_u(c_u)\prod_{e\in E_u} f_e(r_e).$$

令 $\mathcal{E}_1$ 表示事件 $\mathbb{c}(S_{外围}) = c(S_{外围}) \wedge \mathbb{r}(E_{外围}) = r(E_{外围}) \wedge \mathbb{S} = S$, 有

$$\Pr[\mathbb{c}(S_{邻居}) = c(S_{邻居}) \wedge \mathbb{r}(E_{邻居\text{-}外围}) = r(E_{邻居\text{-}外围}) \mid \mathcal{E}_1] = \prod_{u \in S_{邻居}} b_u(c_u) \prod_{e \in E_{邻居\text{-}外围}} f_e(r_e). \tag{20}$$

最后考虑 $S_{中心}$ 里面的点以及 $E_{中心\text{-}邻居}$ 里面的边. 定义事件 $\mathcal{E}_2$ 为 $\mathcal{E}_1 \wedge \mathbb{c}(S_{邻居}) = c(S_{邻居}) \wedge \mathbb{r}(E_{邻居\text{-}外围}) = r(E_{邻居\text{-}外围})$. 我们对 v_0 是否活跃分两种情况.

• v_0 不活跃. 此时有 $S_{中心} = \emptyset$. 根据 $\mathcal{P}(v_0)$ 生成 $(\boldsymbol{c}, \boldsymbol{r})$ 的第 5.1 步可知

$$\begin{aligned} &\Pr[\mathbb{c}(S_{邻居}) = c(S_{邻居}) \wedge \mathbb{r}(E_{中心\text{-}邻居}) = r(E_{中心\text{-}邻居}) \mid \mathcal{E}_2] \\ &= \prod_{e \in E_{中心\text{-}邻居}} f_e(r_e) = \prod_{v \in S_{中心}} b_v(c_v) \prod_{e \in E_{中心\text{-}邻居}} f_e(r_e), \end{aligned} \tag{21}$$

最后一个等号成立是因为 $S_{中心} = \emptyset$.

• v_0 活跃. 根据 $\mathcal{P}(v_0)$ 生成 $(\boldsymbol{c}, \boldsymbol{r})$ 的第 5.1 步, v_0 从分布 $\boldsymbol{b}_{v_0}$ 中生成 c_{v_0}. 此时 $E_{中心\text{-}邻居}$ 为与 v_0 关联的所有边. 容易验证对于任意 $e \in E_{中心\text{-}邻居}$, $r_e = 1$ 的概率都为 p_e, 所以有

$$\Pr[\mathbb{c}(S_{邻居}) = c(S_{邻居}) \wedge \mathbb{r}(E_{中心\text{-}邻居}) = r(E_{中心\text{-}邻居}) \mid \mathcal{E}_2] = \prod_{v \in S_{中心}} b_v(c_v) \prod_{e \in E_{中心\text{-}邻居}} f_e(r_e). \tag{22}$$

注意到 $S_{中心} \uplus S_{邻居} \uplus S_{外围} = S$. 如果图 G 不含三角形, 由引理 3, $E_{中心\text{-}邻居} \uplus E_{邻居\text{-}外围} \uplus E_{外围} = E$. 结合式 (19)~(22) 并使用链式法则 (chain rule) 可以证明式 (18).

耦合过程 $\mathcal{C}(v_0)$. 固定一个点 $v_0 \in V$. 令 $\boldsymbol{X}, \boldsymbol{Y} \in [q]^V$ 为只在 v_0 处取值不同的两个配置. 我们构造局部梅特罗波利斯算法的一步耦合 $(\boldsymbol{X}, \boldsymbol{Y}) \to (\boldsymbol{X}', \boldsymbol{Y}')$. 令 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$ 为实现 $(\boldsymbol{X} \to \boldsymbol{X}')$ 的随机三元组, 令 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$ 为实现 $(\boldsymbol{Y} \to \boldsymbol{Y}')$ 的随机三元组. 注意到给定 $\boldsymbol{X}$ 和 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$ 后, $\boldsymbol{X}'$ 被完全确定; 给定 $\boldsymbol{Y}$ 和 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$ 后, $\boldsymbol{Y}'$ 被完全确定. 要构造耦合 $(\boldsymbol{X}, \boldsymbol{Y}) \to (\boldsymbol{X}', \boldsymbol{Y}')$, 只需要构造随机三元组 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$ 和 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$ 之间的耦合. 我们用 $\mathcal{P}(v_0)$ 来构造耦合 $\mathcal{C}(v_0)$:

(1) 两者生成相同的随机集合 $S^X = S^Y$, 记 $S = S^X = S^Y$.

(2) 对于任意 $v \in S_{外围}$, 两者生成相同的 $c_v^X = c_v^Y$, 对于任意边 $e \in E_{外围}$, 生成相同的 $r_e^X = r_e^Y$.

(3) 对于任意 $e \in E_{邻居\text{-}外围}$, 两者生成相同的 $r_e^X = r_e^Y$.

(4) 根据 v_0 是否为活跃的点 ($S_{中心} = \{v_0\}$ 或 $S_{中心} = \emptyset$) 分两种情况:

• 情况一: $S_{中心} = \emptyset$. 独立地考虑每个点 $u \in S_{邻居}$, 令 $e = \{v_0, u\}$. 在当前条件下, X'_u 完全由 c_u^X, r_e^X 确定, Y'_u 完全由 c_u^Y, r_e^Y 确定, 我们用一种特定的耦合来生成 c_u^X, c_u^Y, r_e^X, r_e^Y, 这种耦合最大化 $X'_u = Y'_u$ 的概率.

• 情况二: $S_{中心} = \{v_0\}$. 首先两者生成相同的 $c_{v_0}^X = c_{v_0}^Y$. 对任意 $u \in \Gamma(v_0)$, 令 $e = \{v_0, u\}$, 两者生成相同的 $r_{e,1}^X = r_{e,1}^Y$. 独立地考虑每个点 $w \in S_{邻居}$, 令 $e' = \{v_0, w\}$, 在当前条件下, X'_w 完全由 c_w^X, $r_{e',2}^X$ 确定, Y'_w 完全由 c_w^Y, $r_{e',2}^Y$ 确定, 我们用一种特定的耦合来生成 c_w^X, c_w^Y, $r_{e',2}^X$, $r_{e',2}^Y$, 这种耦合最大化 $X'_w = Y'_w$ 的概率.

(5) 给定三元组 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$ 和 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$ 之后, 按照算法 1 的规则把 $(\boldsymbol{X}, \boldsymbol{Y})$ 更新成 $(\boldsymbol{X}', \boldsymbol{Y}')$.

引理5 如果图 $G = (V, E)$ 不含三角形, 则对于任意只在 v_0 处取值不同的配置 $\boldsymbol{X}, \boldsymbol{Y} \in [q]^V$, $\mathcal{C}(v_0)$ 是随机三元组 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$ 和 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$ 的一种合法耦合.

证明 根据引理 4, 随机过程 $\mathcal{P}(v_0)$ 会生成服从正确概率分布的三元组. 我们只需要说明, 给定 $\boldsymbol{X}$ 后, 按照 $\mathcal{P}(v_0)$ 生成的三元组 $(S^X, \boldsymbol{c}^X, \boldsymbol{r}^X)$, 与给定 $\boldsymbol{Y}$ 后, 按照 $\mathcal{P}(v_0)$ 生成的三元组 $(S^Y, \boldsymbol{c}^Y, \boldsymbol{r}^Y)$

可以按照 $\mathcal{C}(v_0)$ 的规则耦合即可. 因为两条链产生活跃点集时, 每个点活跃的概率都是 p, 所以两者可以生成相同的 $S^X = S^Y$. 因为所有活跃的点 $v \in S_{外围}$ 都按照 $\boldsymbol{b}_v$ 采样候选取值, 所以两者可以生成相同的 $c_v^X = c_v^Y$. 注意到 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 只在 v_0 一点处取值不同, 在其他所有点取值都相同. 所以对于所有的 $e \in E_{外围} \cup E_{邻居\text{-}外围}$, r_e^X 和 r_e^Y 同分布, 因此两者可以生成相同的 $r_e^X = r_e^Y$. 如果 v_0 活跃, 两条链都从 $\boldsymbol{b}_{v_0}$ 生成候选取值, 因此两者可以生成相同的 $c_{v_0}^X = c_{v_0}^Y$. 对于 $S_{邻居}$ 里面的点 v 和 $E_{中心\text{-}邻居}$ 里面的边 e, $\mathcal{C}(v_0)$ 都会用一种合法耦合来耦合 c_v^X, c_v^Y 以及 r_e^X, r_e^Y. 所以 $\mathcal{C}(v_0)$ 是一种合法的耦合.

5.2 耦合的分析

令 $\mathcal{I} = (V, E, [q], \boldsymbol{b}, \boldsymbol{A})$ 为一个定义在图 $G = (V, E)$ 上的自旋系统. 假设 $\mathcal{I}$ 满足定理 3 中的 3 个条件: (1) G 不含三角形; (2) 存在常数 $0 < \delta < 1$ 使得 $\mathcal{I}$ 单点产生的总影响 $\alpha \leqslant 1 - \delta$; (3) 存在常数 $C > 0$ 使得 $\mathcal{I}$ 以参数 C 满足条件 2. 不失一般性地, 我们假设

$$0 < C \leqslant \Delta. \tag{23}$$

若 $C > \Delta$, 则根据条件 2, 因为 C 是常数, 所以 Δ 也是常数, 容易验证 $C = \Delta$ 时条件 2 也成立. 固定一个点 $v_0 \in V$. 令 $\boldsymbol{X}, \boldsymbol{Y} \in [q]^V$ 为只在 v_0 处取值不同的两个配置. 考虑 $(\boldsymbol{X}, \boldsymbol{Y}) \to (\boldsymbol{X}', \boldsymbol{Y}')$ 的耦合过程 $\mathcal{C}(v_0)$. 重申 $\boldsymbol{\rho}$ 是式 (6) 中定义的影响矩阵, Δ 是图 G 的最大度数, p 是算法 2 的参数. 我们有如下 3 个引理.

引理6 如果 $p \leqslant \frac{1}{4}$, 则耦合过程 $\mathcal{C}(v_0)$ 满足

$$\Pr[X'_{v_0} = Y'_{v_0} \mid \boldsymbol{X}, \boldsymbol{Y}] \geqslant p(1 - \rho(v_0, v_0)) \left(1 - \frac{4pC}{\Delta}\right)^{\Delta}.$$

引理7 耦合过程 $\mathcal{C}(v_0)$ 满足对任意 $u \in \Gamma(v_0)$, $\Pr[X'_u \neq Y'_u \mid \boldsymbol{X}, \boldsymbol{Y}] \leqslant p \cdot \rho(v_0, u)$.

引理8 耦合过程 $\mathcal{C}(v_0)$ 满足对任意 $w \in V \setminus \Gamma^+(v_0)$, $\Pr[X'_w \neq Y'_w \mid \boldsymbol{X}, \boldsymbol{Y}] = 0$.

我们首先用上述 3 个引理证明定理 3, 然后分别证明这 3 个引理.

5.2.1 定理 3 的证明

我们把参数 p 设置为

$$p = \min\left\{\frac{1}{8}, \frac{\delta}{8C}\right\}.$$

对于任意配置对 $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Omega$, 令 $d_{\mathrm{ham}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \triangleq |\{v \in V \mid \sigma_v \neq \tau_v\}|$ 表示 $\boldsymbol{\sigma}$ 和 $\boldsymbol{\tau}$ 之间的汉明距离. 根据期望的线性性质, 有

$$\mathbb{E}[d_{\mathrm{ham}}(\boldsymbol{X}', \boldsymbol{Y}') \mid \boldsymbol{X}, \boldsymbol{Y}] = \sum_{v \in V} \Pr[X'_v \neq Y'_v \mid \boldsymbol{X}, \boldsymbol{Y}].$$

注意到 $p \leqslant \frac{1}{4}$. 根据引理 6~8, 有

$$\mathbb{E}[d_{\mathrm{ham}}(\boldsymbol{X}', \boldsymbol{Y}') \mid \boldsymbol{X}, \boldsymbol{Y}] \leqslant 1 - p(1 - \rho(v_0, v_0)) \left(1 - \frac{4pC}{\Delta}\right)^{\Delta} + p \sum_{u \in \Gamma(v_0)} \rho(v_0, u).$$

由伯努利不等式, 对于任意整数 $N \geqslant 0$, 任意实数 $x \geqslant -1$, 都有 $(1+x)^N \geqslant 1+Nx$. 于是有 $(1-\frac{4pC}{\Delta})^\Delta \geqslant 1-4pC$. 上式可以化简为

$$\begin{aligned}\mathbb{E}[d_{\mathrm{ham}}(\boldsymbol{X}', \boldsymbol{Y}') \mid \boldsymbol{X}, \boldsymbol{Y}] &\leqslant 1 - p(1-\rho(v_0, v_0))(1-4pC) + p\sum_{u\in\Gamma(v_0)} \rho(v_0, u) \\ &= 1 - p\left(1 - 4pC + 4pC\rho(v_0, v_0) - \sum_{u\in\Gamma^+(v_0)} \rho(v_0, u)\right).\end{aligned}$$

根据定义可知 $0 \leqslant \rho(v_0, v_0) \leqslant 1$. 根据总影响 α 的定义有 $\sum_{u\in\Gamma^+(v_0)} \rho(v_0, u) \leqslant \alpha \leqslant 1-\delta$. 所以

$$\mathbb{E}[d_{\mathrm{ham}}(\boldsymbol{X}', \boldsymbol{Y}') \mid \boldsymbol{X}, \boldsymbol{Y}] \leqslant 1 - p(1 - 4pC - (1-\delta)) \leqslant 1 - p(\delta - 4pC) \leqslant 1 - \frac{\delta}{2}p.$$

根据路径耦合引理, 局部梅特罗波利斯算法 (算法 2) 的混合时间为

$$T_{\mathrm{mix}}(\epsilon) \leqslant \left\lceil \frac{2}{\delta p} \log\frac{n}{\epsilon} \right\rceil = \left\lceil \frac{2}{\delta \min\left\{\frac{1}{8}, \frac{\delta}{8C}\right\}} \log\frac{n}{\epsilon} \right\rceil \leqslant \frac{20(1+C)}{\delta^2} \log\frac{n}{\epsilon}.$$

定理 3 得证.

5.2.2 引理 6 的证明

令 $\boldsymbol{X}, \boldsymbol{Y}$ 为只在 v_0 处不同的两个配置. 在证明中, 我们省去 $\Pr[\cdot]$ 记号中的条件 $\boldsymbol{X}, \boldsymbol{Y}$. 根据耦合 $\mathcal{C}(v_0)$, 如果点 v_0 成功在 $\boldsymbol{X} \to \boldsymbol{X}'$ 和 $\boldsymbol{Y} \to \boldsymbol{Y}'$ 的转移过程中更新取值, 则一定有 $X_{v_0} = Y_{v_0}$. 所以我们考虑如下 3 个事件同时发生的概率:

- $\mathcal{E}_1$: 点 v_0 在两条链中都是活跃的点. 有 $\Pr[\mathcal{E}_1] = p$;
- $\mathcal{E}_2$: 对于任意的 $u \in \Gamma(v_0)$, 令 $e = \{u, v_0\}$, $r^X_{e,1} = r^Y_{e,1} = 1$. 根据耦合的定义, 有

$$\Pr[\mathcal{E}_2 \mid \mathcal{E}_1] = \sum_{c\in[q]} \boldsymbol{b}_{v_0}(c) \prod_{u\in\Gamma(v_0)} A_{v_0 u}(c, X_u).$$

注意到上式右侧表示在串行梅特罗波利斯算法 (算法 1) 中, 给定 $X_{\Gamma(v_0)}$ 的条件下, 点 v_0 随机产生一个候选取值并接受的概率. 考虑影响矩阵中的 $\rho(v_0, v_0)$. 根据式 (5) 和 (6) 可知

$$\Pr[\mathcal{E}_2 \mid \mathcal{E}_1] \geqslant \min_{\boldsymbol{\sigma}\in[q]^{\Gamma(v_0)}} \left(\sum_{c\in[q]} \boldsymbol{b}_{v_0}(c) \prod_{u\in\Gamma(v_0)} A_{v_0 u}(c, \sigma_u)\right) = 1 - \rho(v_0, v_0).$$

- $\mathcal{E}_3$: 对于任意 $w \in \Gamma(v_0)$, 令 $e = \{v_0, w\}$, 如下两个事件中任意一个发生 (1) $w \notin S_{\text{邻居}}$; (2) $w \in S_{\text{邻居}}$ 且 $r^X_{e,2} = r^Y_{e,2} = 1$. 我们断言

$$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2] \geqslant \left(1 - \frac{4pC}{\Delta}\right)^\Delta. \tag{24}$$

在给出式 (24) 的证明之前, 先证明引理. 易见如果 $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ 同时发生, 则 v_0 在两条链都会接受候选取值. 根据耦合, v_0 会在两条链中采样相同的候选取值, 所以有

$$\Pr[X'_{v_0} = Y'_{v_0}] \geqslant \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geqslant p(1-\rho(v_0, v_0))\left(1 - \frac{4pC}{\Delta}\right)^\Delta.$$

现在证明式 (24). 定义事件

• $\mathcal{F}(c)$: 事件 $\mathcal{E}_1$ 和 $\mathcal{E}_2$ 同时发生, 且 v_0 在两条链中生成的候选取值都是 $c \in [q]$.

只需要证明

$$\forall c \in [q], \quad \Pr[\mathcal{E}_3 \mid \mathcal{F}(c)] \geqslant \left(1 - \frac{4pC}{\Delta}\right)^{\Delta}. \tag{25}$$

固定一个取值 $c \in [q]$, 固定一个点 $w \in \Gamma(v_0)$, 令 $e = \{v_0, w\}$. 定义事件 $\mathcal{E}_3(e)$ 为如下两个事件中任意一个发生 (1) $w \notin S_{\text{邻居}}$; (2) $w \in S_{\text{邻居}}$ 且 $r^X_{e,2} = r^Y_{e,2} = 1$. 显然 $\Pr[w \notin S_{\text{邻居}} \mid \mathcal{F}(c)] = 1 - p$ 且 $\Pr[w \in S_{\text{邻居}} \mid \mathcal{F}(c)] = p$. 我们分析概率

$$\begin{aligned} &\Pr[r^X_{e,2} = r^Y_{e,2} = 1 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}] \\ &\geqslant 1 - \Pr[r^X_{e,2} = 0 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}] - \Pr[r^Y_{e,2} = 0 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}]. \end{aligned}$$

上述不等式可由联合界 (union bound) 得出. 我们现在分析 $r^X_{e,2} = 0$ 的概率. 考虑从 $\boldsymbol{X} \to \boldsymbol{X}'$ 的转移. 给定事件 $\mathcal{F}(c) \wedge w \in S_{\text{邻居}}$ 之后, w 生成的候选取值服从分布 $\boldsymbol{b}_w$. 于是有

$$\begin{aligned} \Pr[r^X_{e,2} = 0 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}] &= \sum_{a \in [q]} b_w(a)(1 - A_e(c,a)A_e(X_{v_0}, a)) \\ (\star) \qquad &\leqslant \sum_{a \in [q]} b_w(a)(2 - A_e(c,a) - A_e(X_{v_0}, a)) \\ &= 2 - \sum_{a \in [q]} b_w(a)A_e(c,a) - \sum_{a \in [q]} b_w(a)A_e(X_{v_0}, a) \\ (\text{由条件 } 2) \qquad &\leqslant \frac{2C}{\Delta}. \end{aligned}$$

上述不等式中 $(\star)$ 成立是因为对任意 $0 \leqslant x, y \leqslant 1$ 都有 $1 - xy \leqslant 2 - x - y$. 同理可得

$$\Pr[r^Y_{e,2} = 0 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}] \leqslant \frac{2C}{\Delta}.$$

综合可得

$$\begin{aligned} \Pr[\mathcal{E}_3(e) \mid \mathcal{F}(c)] &\geqslant \Pr[w \notin S_{\text{邻居}} \mid \mathcal{F}(c)] + \Pr[w \in S_{\text{邻居}} \mid \mathcal{F}(c)] \Pr[r^X_{e,2} = r^Y_{e,2} = 1 \mid \mathcal{F}(c) \wedge w \in S_{\text{邻居}}] \\ &\geqslant (1 - p) + p\left(1 - \frac{4C}{\Delta}\right) = 1 - \frac{4pC}{\Delta}. \end{aligned}$$

根据式 (23), 有 $C \leqslant \Delta$. 因为 $p \leqslant \frac{1}{4}$, 所以有 $1 - \frac{4pC}{\Delta} \geqslant 0$. 注意到 $\mathcal{E}_3 = \wedge_{e \in E: v_0 \in e} \mathcal{E}_3(e)$. 在给定 $\mathcal{F}(c)$ 的条件下, 对于每个 $u \in \Gamma(v_0)$, 耦合 $\mathcal{C}(v_0)$ 独立地确定 u 是否活跃, 对于每个活跃的 $u \in S_{\text{邻居}}$, 耦合 $\mathcal{C}(v_0)$ 独立地生成四元组 $(c^X_u, c^Y_u, r^X_{e',2}, r^Y_{e',2})$. 所以

$$\Pr[\mathcal{E}_3 \mid \mathcal{F}(c)] \geqslant \left(1 - \frac{4pC}{\Delta}\right)^{\Delta}.$$

式 (25) 得证.

5.2.3 引理 7 的证明

令 $\boldsymbol{X}, \boldsymbol{Y}$ 为只在 v_0 处不同的两个配置. 在证明中, 我们省去 $\Pr[\cdot]$ 记号中的条件 $\boldsymbol{X}, \boldsymbol{Y}$. 固定一个邻居节点 $u \in \Gamma(v_0)$. 如果 $X'_u \neq Y'_u$, 则如下事件一定同时发生:

- $\mathcal{E}_1$: u 是一个活跃的点, 即 $u \in S_{\text{邻居}}$.
- $\mathcal{E}_2$: 对于任意边 $e \in E_u$, 都有 $r_e^X = r_e^Y = 1$. 重申边集 $E_u \subseteq E$ 定义在等式 (13) 中, E_u 表示所有满足 $v \in S_2(v_0)$ 的边 $\{u, v\}$.
- $\mathcal{E}_3$: 在给定 $\mathcal{E}_1$ 和 $\mathcal{E}_2$ 的条件下, 经过耦合之后有 $X'_u \neq Y'_u$.

如果 $X'_u \neq Y'_u$, u 一定活跃, 所以 $\mathcal{E}_1$ 一定发生. 注意到在耦合过程 $\mathcal{C}(v_0)$ 中, 对于任意 $e \in E_u$, 一定有 $r_e^X = r_e^Y$. 如果 $X'_u \neq Y'_u$, 一定有一条链接收了候选取值, 一条链拒绝了候选取值. 如果 $\mathcal{E}_2$ 不发生, 则两条链同时拒绝候选取值, 所以不可能有 $X'_u \neq Y'_u$. 在 $\mathcal{E}_1$ 和 $\mathcal{E}_2$ 都发生的条件下, 如果 $X'_u \neq Y'_u$, 显然 $\mathcal{E}_3$ 发生.

为了分析 $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ 的概率, 对于 $V \setminus \Gamma^+(v_0)$ 中的所有点, 令 $S_{\text{外围}}$ 为这些点中活跃的点, 首先固定一个集合 $S_{\text{外围}}$; 对所有的 $v \in S_{\text{外围}}$, 根据耦合规则, 它们会生成一样的候选取值, 记 v 的候选取值为 c_v, 我们固定所有的候选取值 $(c_v)_{v \in S_{\text{外围}}}$. 在本小节之后的分析中, 只考虑固定 $S_{\text{外围}}$, 以及 $(c_v)_{v \in S_{\text{外围}}}$ 之后, 耦合 $C(v_0)$ 生成的条件概率空间. 在此条件下, 证明 $\Pr[X'_u \neq Y'_u] \leqslant p \cdot \rho(v_0, u)$, 再对 $S_{\text{外围}}$ 和 $(c_v)_{v \in S_{\text{外围}}}$ 利用全概率公式即可证明定理.

首先分析事件 $\mathcal{E}_1$ 的概率. 显然有

$$\Pr[\mathcal{E}_1] = p. \tag{26}$$

接着分析事件 $\mathcal{E}_2$ 的概率. 给定集合 $S_{\text{外围}}$ 和候选取值 $(c_v)_{v \in S_{\text{外围}}}$ 之后, 由式 (16), 有

$$\Pr[\mathcal{E}_2 \mid \mathcal{E}_1] = \sum_{a \in [q]} b_u(a) \prod_{e \in E_u} q_e(a),$$

其中 $q_e(a)$ 定义在式 (15) 中, 在上式中, $q_e(a)$ 可以直接按照定义计算是因为 $S_{\text{外围}}$, $(c_v)_{v \in S_{\text{外围}}}$ 以及 $\boldsymbol{X} \in [q]^V$ 都已给定. 不失一般性地, 假设

$$\Pr[\mathcal{E}_2 \mid \mathcal{E}_1] > 0, \tag{27}$$

否则 $X'_u \neq Y'_u$ 的概率为 0, 引理显然成立.

最后分析事件 $\mathcal{E}_3$ 的概率. 注意到在 $\mathcal{E}_1$ 和 $\mathcal{E}_2$ 都发生的条件下, 根据式 (17), 点 u 需要从如下分布 $\boldsymbol{b}'_u$ 中产生候选取值 $c_u \in [q]$:

$$\begin{aligned} \forall a \in [q], \quad b'_u(a) &\triangleq \frac{b_u(a) \prod_{e \in E_u} q_e(a)}{\sum_{z \in [q]} b_u(z) \prod_{e \in E_u} q_e(z)} = \frac{b_u(a) \prod_{e \in E_u} q_e(a)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} \\ &\leqslant \frac{b_u(a) \prod_{e = \{u,v\} \in E_u} A_e(a, X_v)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}. \end{aligned} \tag{28}$$

上式中最后一个不等式成立是因为对任意边 $e = \{u, v\} \in E_u$, 任意 $a \in [q]$ 都有 $q_e(a) \leqslant A_e(a, X_v)$. 我们分两种情况来考虑 $\mathcal{E}_3$ 的概率.

情况一: $S_{\text{中心}} = \emptyset$. 在给定 $\mathcal{E}_1, \mathcal{E}_2$ 的条件下, 我们可以计算出 X'_u 和 Y'_u 的概率分布满足:

$$\begin{aligned} &\forall a \notin [q] \setminus \{X_u\} = [q] \setminus \{Y_u\}, \quad \Pr[X'_u = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{\text{中心}} = \emptyset] = b'_u(a) A_{v_0 u}(X_{v_0}, a), \\ &\forall a \notin [q] \setminus \{X_u\} = [q] \setminus \{Y_u\}, \quad \Pr[Y'_u = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{\text{中心}} = \emptyset] = b'_u(a) A_{v_0 u}(Y_{v_0}, a). \end{aligned} \tag{29}$$

在当前条件下, X_u' 完全由 c_u^X, r_e^X 确定, Y_u' 完全由 c_u^Y, r_e^Y 确定, $\mathcal{C}(v_0)$ 在耦合 c_u^X, c_u^Y, r_e^X, r_e^Y 时最大化 $X_u' = Y_u'$ 的概率. 根据引理 1, $X_u' \neq Y_u'$ 的概率等于两者对应分布的总变差, 于是有

$$\begin{aligned}&\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] \\ &= \frac{1}{2}\sum_{a\in[q]} \left|\Pr[X_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] - \Pr[Y_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset]\right|. \end{aligned}\tag{30}$$

注意到 $X_u = Y_u$. 根据式 (29) 可知对于任意的 $a \in [q] \setminus \{X_u\}$, $\Pr[X_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] > \Pr[Y_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset]$ 当且仅当 $A_{v_0u}(X_{v_0}, a) > A_{v_0u}(Y_{v_0}, a)$. 我们定义如下两个集合:

$$\begin{aligned}H_X &\triangleq \{a \in [q] \setminus \{X_u\} \mid A_{v_0u}(X_{v_0}, a) > A_{v_0u}(Y_{v_0}, a)\},\\ H_Y &\triangleq \{a \in [q] \setminus \{X_u\} \mid A_{v_0u}(Y_{v_0}, a) > A_{v_0u}(X_{v_0}, a)\}.\end{aligned}\tag{31}$$

由 H_X 和 H_Y, 我们可以定义出如下两个量:

$$\begin{aligned}D_X &\triangleq \sum_{a\in H_X} (\Pr[X_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] - \Pr[Y_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset]),\\ D_Y &\triangleq \sum_{a\in H_Y} (\Pr[Y_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] - \Pr[X_u' = a \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset]).\end{aligned}$$

因为 $D_X \cup D_Y = [q] \setminus \{X_u\}$. 根据式 (8), 等式 (30) 中的总变差可以写成如下形式:

$$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] = \max\{D_X, D_Y\}.$$

根据式 (29) 和 (28), 我们有

$$\begin{aligned}D_X &= \sum_{a\in H_X} b_u'(a)(A_{v_0u}(X_{v_0}, a) - A_{v_0u}(Y_{v_0}, a))\\ &\leqslant \frac{\sum_{a\in H_X} b_u(a)(A_{v_0u}(X_{v_0}, a) - A_{v_0u}(Y_{v_0}, a)) \prod_{e=\{u,v\}\in E_u} A_e(a, X_v)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} \triangleq \frac{B_X}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}.\end{aligned}$$

$$\begin{aligned}D_Y &= \sum_{a\in H_Y} b_u'(a)(A_{v_0u}(Y_{v_0}, a) - A_{v_0u}(X_{v_0}, a))\\ &\leqslant \frac{\sum_{a\in H_Y} b_u(a)(A_{v_0u}(Y_{v_0}, a) - A_{v_0u}(X_{v_0}, a)) \prod_{e=\{u,v\}\in E_u} A_e(a, X_v)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} \triangleq \frac{B_Y}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}.\end{aligned}$$

所以有

$$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{中心} = \emptyset] \leqslant \max\left\{\frac{B_X}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}, \frac{B_Y}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}\right\}. \tag{32}$$

情况二: $S_{中心} = \{v_0\}$. 令 $e = \{v_0, u\}$. 此时 v_0 要从概率分布 $\boldsymbol{b}_{v_0}$ 中生成一个随机候选取值 $c_{v_0} \in [q]$. 根据耦合, 两条链会生成一样的候选取值. 令事件 $\mathcal{F}(c_{v_0})$ 表示 v_0 活跃且 v_0 生成的候选取值为 c_{v_0}. 耦合过程在采样了 $r_{e,1}^X, r_{e,1}^Y, r_{e,2}^X, r_{e,2}^Y$ 且一定有 $r_{e,1}^X = r_{e,1}^Y$. 如果 $r_{e,1}^X = r_{e,1}^Y = 0$, 那么一定有 $X_u' = X_u = Y_u = Y_u'$. 我们考虑 $r_{e,1}^X = r_{e,1}^Y = 1$ 的情况. 记事件 $\mathcal{G}$ 为 $\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{F}(c_{v_0}) \wedge r_{e,1}^X = r_{e,1}^Y = 1$, 不妨假设 $\mathcal{G}$ 发生的概率大于 0. 我们有

$$\begin{aligned}\forall a \notin [q] \setminus \{X_u\} = [q] \setminus \{Y_u\}, \qquad & \Pr[X_u' = a \mid \mathcal{G}] = b_u'(a) A_{v_0u}(X_{v_0}, a) A_{v_0u}(c_{v_0}, a)\\ (\gamma(a) \triangleq A_{v_0u}(c_{v_0}, a)) \quad & \phantom{\Pr[X_u' = a \mid \mathcal{G}]} = \gamma(a) b_u'(a) A_{v_0u}(X_{v_0}, a),\\ \forall a \notin [q] \setminus \{X_u\} = [q] \setminus \{Y_u\}, \qquad & \Pr[Y_u' = a \mid \mathcal{G}] = b_u'(a) A_{v_0u}(Y_{v_0}, a) A_{v_0u}(c_{v_0}, a)\\ (\gamma(a) \triangleq A_{v_0u}(c_{v_0}, a)) \quad & \phantom{\Pr[Y_u' = a \mid \mathcal{G}]} = \gamma(a) b_u'(a) A_{v_0u}(Y_{v_0}, a).\end{aligned}\tag{33}$$

在当前条件下, X'_u 完全由 c_u^X, $r_{e,2}^X$ 确定, Y'_u 完全由 c_u^Y, $r_{e,2}^Y$ 确定, $\mathcal{C}(v_0)$ 在耦合 c_u^X, c_u^Y, $r_{e,2}^X$, $r_{e,2}^Y$ 时最大化 $X'_u = Y'_u$ 的概率. 根据引理 1, $X'_u \neq Y'_u$ 的概率等于两者对应分布的总变差, 于是有

$$\Pr[X'_u \neq Y'_u \mid \mathcal{G}] = \frac{1}{2} \sum_{a \in [q]} \left| \Pr[X'_u = a \mid \mathcal{G}] - \Pr[Y'_u = a \mid \mathcal{G}] \right|. \tag{34}$$

根据式 (33) 可知, 对于任意的 $a \in [q] \setminus \{X_u\}$, $\Pr[X'_u = a \mid \mathcal{G}] > \Pr[Y'_u = a \mid \mathcal{G}]$ 当且仅当 $A_{v_0u}(X_{v_0,a}, a) > A_{v_0u}(Y_{v_0}, a)$. 重申集合 H_X 和集合 H_Y 定义在式 (31) 中. 类似地, 定义

$$\begin{aligned} D'_X &\triangleq \sum_{a \in H_X} (\Pr[X'_u = a \mid \mathcal{G}] - \Pr[Y'_u = a \mid \mathcal{G}]), \\ D'_Y &\triangleq \sum_{a \in H_Y} (\Pr[Y'_u = a \mid \mathcal{G}] - \Pr[X'_u = a \mid \mathcal{G}]). \end{aligned}$$

根据式 (8), 等式 (34) 中的总变差可以写成如下形式:

$$\Pr[X'_u \neq Y'_u \mid \mathcal{G}] = \max\{D'_X, D'_Y\}.$$

结合式 (28), (33) 以及 $0 \leqslant \gamma(a) \leqslant 1$, 有

$$\begin{aligned} D'_X &= \sum_{a \in H_X} \gamma(a) b'_u(a) (A_{v_0u}(X_{v_0}, a) - A_{v_0u}(Y_{v_0}, a)) \leqslant \sum_{a \in H_X} b'_u(a) (A_{v_0u}(X_{v_0}, a) - A_{v_0u}(Y_{v_0}, a)) \\ &\leqslant \frac{\sum_{a \in H_X} b_u(a) (A_{v_0u}(X_{v_0}, a) - A_{v_0u}(Y_{v_0}, a)) \prod_{e=\{u,v\} \in E_u} A_e(a, X_v)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} = \frac{B_X}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}. \end{aligned}$$

$$\begin{aligned} D'_Y &= \sum_{a \in H_Y} \gamma(a) b'_u(a) (A_{v_0u}(Y_{v_0}, a) - A_{v_0u}(X_{v_0}, a)) \leqslant \sum_{a \in H_Y} b'_u(a) (A_{v_0u}(Y_{v_0}, a) - A_{v_0u}(X_{v_0}, a)) \\ &\leqslant \frac{\sum_{a \in H_Y} b_u(a) (A_{v_0u}(Y_{v_0}, a) - A_{v_0u}(X_{v_0}, a)) \prod_{e=\{u,v\} \in E_u} A_e(a, X_v)}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} = \frac{B_Y}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}. \end{aligned}$$

注意到事件 $\mathcal{G}$ 的定义为 $\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{F}(c_{v_0}) \wedge r_{e,1}^X = r_{e,1}^Y = 1$, 而且 $X'_u \neq Y'_u$ 仅当 $r_{e,1}^X = r_{e,1}^Y = 1$. 于是有如下不等式:

$$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{F}(c_{v_0})] = \Pr[X'_u \neq Y'_u \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{F}(c_{v_0})] \leqslant \max\{B_X, B_Y\}.$$

注意到 B_X, B_Y 都和 c_{v_0} 无关, 对条件 $\mathcal{F}(c_{v_0})$ 使用全概率公式可得

$$\Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge S_{\text{中心}} = \{v_0\}] \leqslant \max\left\{ \frac{B_X}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]}, \frac{B_Y}{\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]} \right\}. \tag{35}$$

综合两种情况 (32) 和 (35), 再结合式 (26) 和 (27) 可知

$$\Pr[X'_u \neq Y'_u] \leqslant \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] = \Pr[\mathcal{E}_1] \Pr[\mathcal{E}_2 \mid \mathcal{E}_1] \Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2] \leqslant p \max\{B_X, B_Y\}. \tag{36}$$

因为图 $G = (V, E)$ 不含三角形, 所以 $E_u \cup \{v_0u\}$ 恰好是 u 关联的所有边. 注意到两个配置 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 只在点 v_0 处取值不同. 根据 H_X, H_Y 的定义 (等式 (31)) 以及式 (5), (6) 可知

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{X}}, \pi_u^{\boldsymbol{Y}}\right) = \frac{1}{2} \sum_{a \in [q]} \left| \pi_u^{\boldsymbol{X}}(a) - \pi_u^{\boldsymbol{Y}}(a) \right|$$

$$\begin{aligned}(\star)\quad &= \max\left\{\sum_{a\in H_X}(\pi_u^{\boldsymbol{X}}(a)-\pi_u^{\boldsymbol{Y}}(a)),\ \sum_{a\in H_Y}(\pi_u^{\boldsymbol{Y}}(a)-\pi_u^{\boldsymbol{X}}(a))\right\}\\ (\text{由 } B_X,B_Y \text{ 的定义})\quad &= \max\{B_X,B_Y\}.\end{aligned}$$

其中等式 $(\star)$ 可以由以下两点推出:

- 由定义 (5) 可知, 对于任意 $a\neq X_u$, $\pi_u^{\boldsymbol{X}}(a)>\pi_u^{\boldsymbol{Y}}(a)$ 当且仅当 $A_{v_0u}(a,X_{v_0})>A_{v_0u}(a,Y_{v_0})$;
- 全变差可以写成式 (12) 的形式, 且 $H_X\cup H_Y=[q]\setminus\{X_u\}$.

根据 $\boldsymbol{\rho}$ 的定义, 有

$$\max\{B_X,B_Y\}=d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{X}},\pi_u^{\boldsymbol{Y}}\right)\leqslant\rho(v_0,u).\tag{37}$$

结合式 (36) 和 (37) 可知 $\Pr[X_u'\neq Y_u']\leqslant p\cdot\rho(v_0,u)$.

5.2.4 引理 8 的证明

令集合 $\Lambda=V\setminus\Gamma^+(v_0)$. 首先注意到 $X_\Lambda=Y_\Lambda$. 考虑耦合过程 $\mathcal{C}(v_0)$. 两条链生成相同的活跃点集. 所有 Λ 中活跃的点生成相同的候选取值. 所有和 Λ 中活跃的点关联的边 e 采样相同的 $r_e^X=r_e^Y$. 所以一定有 $X_\Lambda'=Y_\Lambda'$.

6 具体自旋系统上的应用

本节证明推论 1. 容易验证, 推论中所有的自旋系统都满足假设 1. 推论已经假设自旋系统所在的图 $G=(V,E)$ 不含三角形. 对于推论 1 里所有的自旋系统, 只需要给 α 一个上界, 并证明系统满足条件 2. 这样, 推论 1 就可以由定理 3 推出.

根据式 (5) 和 (6) 可知, 对于任意自旋系统, 影响矩阵 $\rho(u,v)>0$ 仅当 $v\in\Gamma^+(u)$. 因此, 只需要考虑相邻点对之间, 以及一个点对自身的影响.

首先考虑满足 $q\geqslant(2+\delta)\Delta$ 的图染色问题. 固定图上一个点 $u\in V$, 假设 u 取到式 (7) 中的最大值. 容易验证 $\rho(u,u)\leqslant\frac{\Delta}{q}$. 对于任意点 $v\in\Gamma(u)$, 容易验证 $\rho(u,v)\leqslant\frac{1}{q}$. 于是有

$$\alpha=\sum_{v\in V}\rho(u,v)=\sum_{v\in\Gamma^+(u)}\rho(u,v)\leqslant\frac{2\Delta}{q}\leqslant\frac{2}{2+\delta}=1-\frac{\delta}{2+\delta}.$$

令 $\delta'=\frac{\delta}{2+\delta}$, 则有 $\alpha\leqslant1-\delta'$. 容易验证, 当 $q\geqslant(2+\delta)\Delta$ 时, 图染色问题以常数 $C=\frac{1}{2}$ 满足条件 2.

接着考虑满足 $\lambda\leqslant\frac{1-\delta}{\Delta}$ 的硬核模型. 不妨假设 $0<\delta\leqslant1$. 固定图上一个点 $u\in V$, 假设 u 取到式 (7) 中的最大值. 容易验证 $\rho(u,u)\leqslant\frac{\lambda}{1+\lambda}$. 对于任意点 $v\in\Gamma(u)$, 容易验证 $\rho(u,v)\leqslant\frac{\lambda}{1+\lambda}$. 于是有

$$\begin{aligned}\alpha=\sum_{v\in V}\rho(u,v)=\sum_{v\in\Gamma^+(u)}\rho(u,v)&\leqslant\frac{(\Delta+1)\lambda}{1+\lambda}\leqslant\frac{\Delta+1}{1+\frac{\Delta}{1-\delta}}\\ (\text{由 } \Delta\geqslant1)\quad&\leqslant1-\frac{\delta}{2-\delta}.\end{aligned}$$

令 $\delta'=\frac{\delta}{2-\delta}$, 则有 $\alpha\leqslant1-\delta'$. 容易验证, 当 $\lambda\leqslant\frac{1-\delta}{\Delta}$ 时, 硬核模型以常数 $C=1$ 满足条件 2.

最后考虑满足 $\exp(-2|\beta|)\geqslant1-\frac{(1-\delta)\eta}{\Delta+1}$ 的伊辛模型. 不妨假设 $0<\delta\leqslant1$. 固定图上一个点 $u\in V$, 假设 u 取到式 (7) 中的最大值. 令

$$\gamma=\exp(-2|\beta|).\tag{38}$$

设 u 的度数为 d, 其中 $1 \leqslant d \leqslant \Delta$. 考虑一对配置 $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \{-1,+1\}^V$, 其中 $\boldsymbol{\sigma}$ 和 $\boldsymbol{\tau}$ 只在 u 处取值不同, 不妨假设 $\sigma_u = +1$ 且 $\tau_u = -1$. 因为 $\sigma_{\Gamma(u)} = \tau_{\Gamma(u)}$, 假设 d 个邻居中有 s 个邻居取值为 $+1$, 有 $d-s$ 个邻居取值为 -1, 其中 $0 \leqslant s \leqslant d$. 由式 (8), 有

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = \left|\pi_u^{\boldsymbol{\sigma}}(+1) - \pi_u^{\boldsymbol{\tau}}(+1)\right| = \left|1 - \pi_u^{\boldsymbol{\sigma}}(-1) - \pi_u^{\boldsymbol{\tau}}(+1)\right|.$$

如果 $\beta \geqslant 0$, 则

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = 1 - \frac{1}{2}\gamma^s - \frac{1}{2}\gamma^{d-s} \leqslant 1 - \frac{1}{2}\sqrt{\gamma^d} \leqslant 1 - \frac{\gamma^{\frac{\Delta}{2}}}{2}.$$

如果 $\beta < 0$, 则

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = 1 - \frac{1}{2}\gamma^s - \frac{1}{2}\gamma^{d-s} \leqslant 1 - \frac{1}{2}\sqrt{\gamma^d} \leqslant 1 - \frac{\gamma^{\frac{\Delta}{2}}}{2}.$$

所以

$$\rho(u,u) \leqslant 1 - \frac{\gamma^{\frac{\Delta}{2}}}{2}.$$

考虑点 u 的一个邻居节点 v. 现在假设 v 的度数为 d. 因为 $\boldsymbol{\sigma}$ 和 $\boldsymbol{\tau}$ 只在点 u 处取值不同, 我们假设在 $\boldsymbol{\sigma}$ 中, v 的邻居里有 $s+1$ 个邻居取值为 $+1$, 在 $\boldsymbol{\tau}$ 中, v 的邻居里有 s 个邻居取值为 $+1$, 其中 $0 \leqslant s \leqslant d-1$. 我们先假设 $\sigma_v = \tau_v = +1$. 此时根据式 (8), 有

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = \left|\pi_v^{\boldsymbol{\sigma}}(-1) - \pi_v^{\boldsymbol{\tau}}(-1)\right|.$$

如果 $\beta \geqslant 0$, 则

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = \frac{1}{2}\gamma^s - \frac{1}{2}\gamma^{s+1} = \frac{1}{2}\gamma^s(1-\gamma) \leqslant \frac{1-\gamma}{2}.$$

如果 $\beta < 0$, 则

$$d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = \frac{1}{2}\gamma^{d-s-1} - \frac{1}{2}\gamma^{d-s} = \frac{1}{2}\gamma^{d-s-1}(1-\gamma) \leqslant \frac{1-\gamma}{2}.$$

如果 $\sigma_v = \tau_v = -1$, 可以计算 $d_{\mathrm{TV}}\left(\pi_u^{\boldsymbol{\sigma}}, \pi_u^{\boldsymbol{\tau}}\right) = \left|\pi_v^{\boldsymbol{\sigma}}(+1) - \pi_v^{\boldsymbol{\tau}}(+1)\right|$, 同理得出一样的上界. 所以

$$\forall v \in \Gamma_u, \quad \rho(u,v) \leqslant \frac{1-\gamma}{2}.$$

所以有

$$\alpha = \sum_{v \in V} \rho(u,v) = \sum_{v \in \Gamma^+(u)} \rho(u,v) \leqslant 1 - \frac{\gamma^{\frac{\Delta}{2}}}{2} + \frac{(1-\gamma)\Delta}{2}.$$

注意到 $\eta \approx 0.703$, 有 $\gamma \geqslant 1 - \frac{(1-\delta)\eta}{\Delta+1} \geqslant 1 - \frac{(1-\delta)\eta}{\Delta+(1-\delta)\eta}$, 所以

$$\begin{aligned}
&\gamma^{\frac{\Delta}{2}} \geqslant \left(1 - \frac{(1-\delta)\eta}{\Delta+(1-\delta)\eta}\right)^{\frac{\Delta}{2}} = \left(\frac{1}{1+\frac{(1-\delta)\eta}{\Delta}}\right)^{\frac{\Delta}{2}} \geqslant \exp\left(-\frac{(1-\delta)\eta}{2}\right),\\
&(1-\gamma)\Delta \leqslant \frac{\Delta(1-\delta)\eta}{(\Delta+1)} \leqslant (1-\delta)\eta.
\end{aligned}$$

根据定义 $\eta=\exp(-\frac{\eta}{2})$, 注意到 $f(x)=\exp(-x/2)-x$ 单调递减. 令

$$\delta'=\frac{1}{2}\exp\left(-\frac{(1-\delta)\eta}{2}\right)-\frac{(1-\delta)\eta}{2}>0.$$

则有 $\alpha\leqslant 1-\delta'$. 容易验证, 当 $\gamma\geqslant 1-\frac{(1-\delta)\eta}{\Delta+1}$ 时, 伊辛模型以常数 $C=1$ 满足条件 2.

在上述 3 种模型中, 参数 δ' 都只和 δ 有关, 参数 C 都是一个绝对常数. 由定理 3 可得推论 1.

7　总结

本文研究了分布式采样的局部梅特罗波利斯算法, 证明了算法在一般自旋系统上的正确性和收敛性结论. 对于较自然 (具体由条件 2 所刻画) 的自旋系统, 如果所在图不含三角形且自旋系统满足梅特罗波利斯算法的 Dobrushin-Shlosman 条件, 则局部梅特罗波利斯算法的混合时间为 $O(\log n)$, 相对串行算法有 $\Omega(n)$ 倍的最优并行加速. 相比之前的结果 [15, 16], 本文结论对一般自旋系统成立. 在技术上, 我们给了一种全新的路径耦合方案, 它能给一般自旋系统一个统一的分析.

除去前文中提到的, 也有其他工作研究分布式采样算法. 文献 [34] 研究在 LOCAL 模型上直接模拟串行梅特罗波利斯算法, 我们强调相关结论不适用于局部梅特罗波利斯算法. 文献 [14] 提出了另一种分布式马尔可夫链: 卢比 – 格劳伯算法. 文献 [18,19] 提出了非马尔可夫链分布式采样算法.

分布式采样的研究理论研究处于起步阶段, 目前有很多重要的公开问题. 例如设计新的分布式采样算法和探索新的分布式采样算法分析工具. 对于本文研究的局部梅特罗波利斯算法, 一个公开问题是能否在一般图上得到更好的收敛条件.

参考文献

1　Mezard M, Montanari A. Information, Physics, and Computation. Oxford: Oxford University Press, 2009

2　Sinclair A, Jerrum M. Approximate counting, uniform generation and rapidly mixing Markov chains. Inf Comput, 1989, 82: 93–133

3　Levin D A, Peres Y. Markov Chains and Mixing Times. Providence: American Mathematical Society, 2017

4　Newman D, Asuncion A, Smyth P, et al. Distributed inference for latent Dirichlet allocation. In: Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS), 2007. 1081–1088

5　Doshi-Velez F, Mohamed S, Ghahramani Z, et al. Large scale nonparametric Bayesian inference: data parallelisation in the Indian buffet process. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS), 2009. 1294–1302

6　Smyth P, Welling M, Asuncion A U. Asynchronous distributed learning of topic models. In: Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS), 2009. 81–88

7　Yan F, Xu N Y, Qi Y. Parallel inference for latent Dirichlet allocation on graphics processing units. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS), 2009. 2134–2142

8　Gonzalez J E, Low Y, Gretton A, et al. Parallel Gibbs sampling: from colored fields to thin junction trees. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), 2011. 324–332

9　Ahmed A, Aly M, Gonzalez J E, et al. Scalable inference in latent variable models. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM), 2012. 123–132

10　de Sa C, Olukotun K, Ré C. Ensuring rapid mixing and low bias for asynchronous Gibbs sampling. In: Proceedings of the 33rd International Conference on Machine Learning (ICML), 2016. 1567–1576

11　de Sa C, Zhang C, Olukotun K, et al. Rapidly mixing Gibbs sampling for a class of factor graphs using hierarchy width. In: Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS), 2015. 3097–3105

12　Daskalakis C, Dikkala N, Jayanti S. Hogwild!-Gibbs can be panaccurate. In: Proceedings of the 31st Advances in Neural Information Processing Systems (NIPS), 2018. 32–41

13 Kandasamy K, Krishnamurthy A, Schneider J, et al. Parallelised Bayesian optimisation via Thompson sampling. In: Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS), 2018. 133–142
14 Feng W M, Sun Y X, Yin Y T. What can be sampled locally? Distrib Comput, 2020, 33: 227–253
15 Feng W M, Hayes T P, Yin Y T. Distributed symmetry breaking in sampling (optimal distributed randomly coloring with fewer colors). 2018. ArXiv:1802.06953
16 Fischer M, Ghaffari M. A simple parallel and distributed sampling technique: local Glauber dynamics. In: Proceedings of the 32nd International Symposium on Distributed Computing (DISC), 2018
17 Feng W M, Yin Y T. On local distributed sampling and counting. In: Proceedings of the 37th ACM Symposium on Principles of Distributed Computing (PODC), 2018. 189–198
18 Guo H, Jerrum M, Liu J C. Uniform sampling through the Lovász local lemma. J ACM, 2019, 66: 1–31
19 Feng W M, Vishnoi N K, Yin Y T. Dynamic sampling from graphical models. In: Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC), 2019. 1070–1081
20 Linial N. Distributive graph algorithms global solutions from local data. In: Proceedings of the 28th IEEE Annual Symposium on Foundations of Computer Science (FOCS), 1987. 331–335
21 Jerrum M. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. Random Struct Alg, 1995, 7: 157–165
22 Bubley R, Dyer M. Path coupling: a technique for proving rapid mixing in Markov chains. In: Proceedings of the 38th IEEE Annual Symposium on Foundations of Computer Science (FOCS), 1997. 223–231
23 Hayes T P. A simple condition implying rapid mixing of single-site dynamics on spin systems. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2006. 39–46
24 Dyer M, Goldberg L A, Jerrum M. Dobrushin conditions and systematic scan. In: Proceedings of the 10th International Workshop on Randomization and Computation (RANDOM), 2006. 327–338
25 Dobrushin R L. Prescribing a system of random variables by conditional distributions. Theory Probab Appl, 1970, 15: 458–486
26 Dobrushin R L, Shlosman S B. Completely analytical Gibbs fields. In: Proceedings of Statistical Physics and Dynamical Systems, 1985
27 Dobrushin R L, Shlosman S B. Constructive criterion for the uniqueness of Gibbs field. In: Proceedings of Statistical Physics and Dynamical Systems, 1985
28 Jonasson J. Uniqueness of uniform random colorings of regular trees. Stat Probab Lett, 2002, 57: 243–248
29 Galanis A, Štefankovič D, Vigoda E. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. J ACM, 2015, 62: 50
30 Weitz D. Counting independent sets up to the tree threshold. In: Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), 2006. 140–149
31 Sly A. Computational transition at the uniqueness threshold. In: Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010. 287–296
32 Sly A, Sun N. Counting in two-spin models on d-regular graphs. Ann Probab, 2014, 42: 2383–2416
33 Sinclair A, Srivastava P, Thurley M. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. J Stat Phys, 2014, 155: 666–686
34 Feng W M, Hayes T P, Yin Y T. Distributed metropolis sampler with optimal parallelism. In: Proceedings of the 32nd ACM-SIAM Symposium on Discrete Algorithms (SODA), 2021. 2121–2140

Distributed Metropolis algorithm: convergence condition and optimal parallel speed-up

Weiming FENG* & Yitong YIN*

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023*, China*
* Corresponding author. E-mail: fengwm@smail.nju.edu.cn, yinyt@nju.edu.cn

Abstract　The Metropolis algorithm is a fundamental Markov chain Monte Carlo (MCMC) sampling technique, which can be used for drawing random samples from high-dimensional probability distributions (i.e., Gibbs distributions) represented by probabilistic graphical models. The traditional Metropolis algorithm is a sequential algorithm. A classic result regarding the fast convergence of the Metropolis algorithm is: when the Dobrushin-Shlosman condition for the Metropolis algorithm is satisfied, the algorithm converges rapidly within $O(n \log n)$ step, where n is the number of variables. This paper studies a distributed variant of the Metropolis algorithm, called the local-Metropolis algorithm. We provide an analysis of the correctness and convergence of this new algorithm, and show: the algorithm always converges to the correct Gibbs distribution; moreover, for a natural class of triangle-free probabilistic graphical models, as long as the same Dobrushin-Shlosman condition is satisfied, the local-Metropolis algorithm converges within $O(\log n)$ rounds of distributed computing. Compared to the traditional sequential Metropolis algorithm, this achieves an asymptotically optimal $\Omega(n)$ factor of speed-ups. Concrete applications include the distributed sampling algorithms for graph coloring, hardcore model, and Ising model.

Keywords　distributed sampling, Markov chain Monte Carlo, mixing time, spin system, coupling