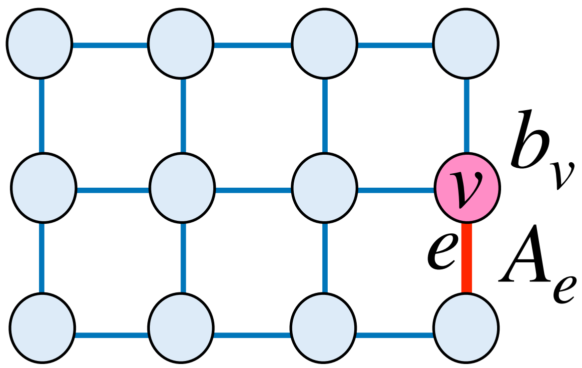
Dynamic and Distributed Algorithms for Sampling from Gibbs Distributions

Yitong Yin
Nanjing University

Gibbs Distribution

$G = (V, E)$



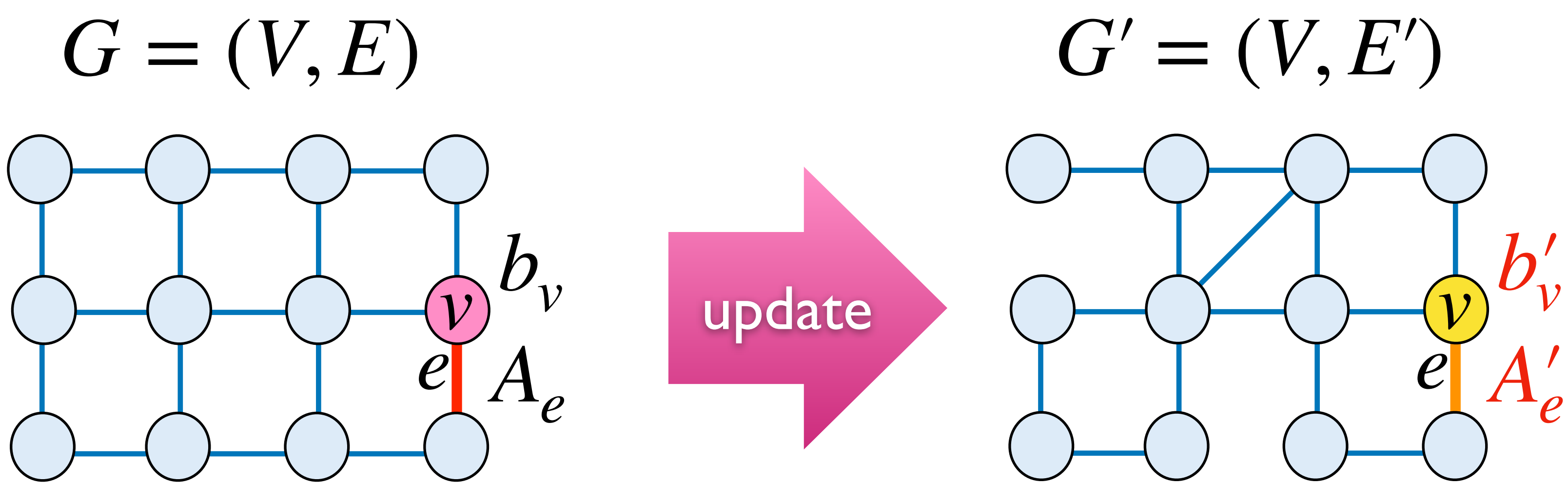
- $q \geq 2$ spin states
- each $v \in V$, distribution $b_v : [q] \to [0,1]$
- each $e \in E$, symmetric $A_e : [q]^2 \to [0,1]$

$\forall$configuration $\sigma \in [q]^V$:

$$w(\sigma) = \prod_{e=\{u,v\} \in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v \in V} b_v\left(\sigma_v\right)$$

Gibbs distribution: $\mu(\sigma) = \dfrac{w(\sigma)}{Z}$ where $Z = \sum_{\sigma \in [q]^V} w(\sigma)$

Dynamic Sampling

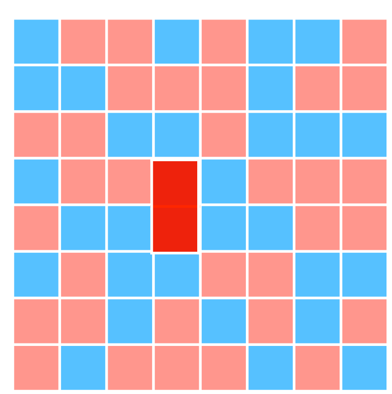


dynamic sampling algorithm:

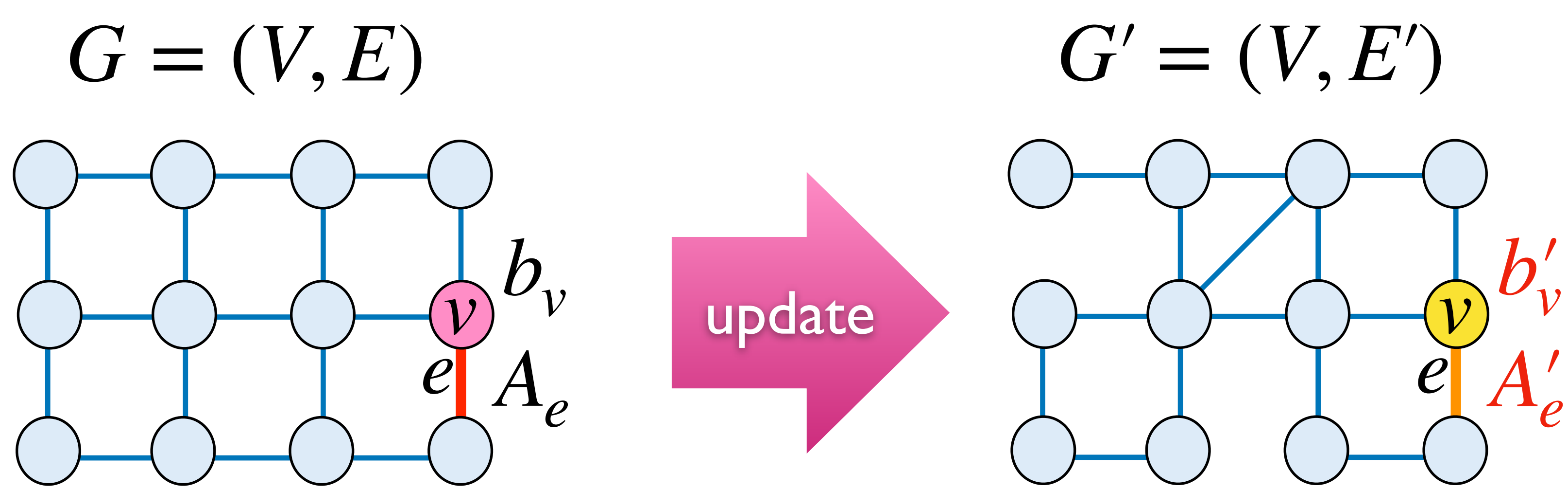
$$X \sim \mu \longrightarrow X' \sim \mu'$$

with cost that depends on

$|\texttt{update}| \triangleq$ # changed vertices and edges



Dynamic Sampling

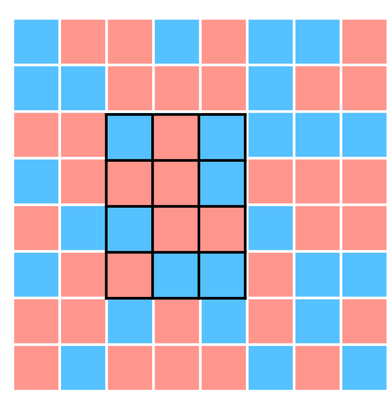


dynamic sampling algorithm:

$$X \sim \mu \quad \longrightarrow \quad X' \sim \mu'$$

with cost $\tilde{O}(|\texttt{update}|)$

$|\texttt{update}| \triangleq$ # changed vertices and edges



Dynamic Sampling

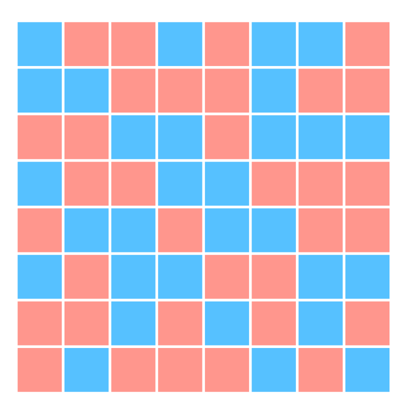
empty graph $(V, \varnothing)$ $G = (V, E)$

dynamic sampling algorithm:

$$X^{(0)} \sim \bigoplus_v b_v \longrightarrow X \sim \mu$$

$\tilde{O}(|\texttt{update}|)$ dynamic sampling

$\Longrightarrow \tilde{O}(|E|)$ static sampling



A Moser-Tardos style algorithm

[Feng, Vishnoi, Y. '19]

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v\in V} b_v\left(\sigma_v\right)$

current sample: $\boldsymbol{X} \sim \mu$

$R \leftarrow \{v \in V \mid v$ is updated or incident to updated $e\}$;

while $R \neq \emptyset$ **do**

for every $v \in R$, resample $X_v \sim b_v$ independently;

every *internal* $e = \{u, v\} \subseteq R$ accepts ind. w.p. $A_e(X_u, X_v)$;

every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts ind. w.p.

$\frac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)} \min A_e\left(X_u^{\text{old}}, \cdot\right)$; // X_u^{old}: X_u before resampling

$R \leftarrow \bigcup_{e \text{ rejects}} e$;

A Moser-Tardos style algorithm

[Feng, Vishnoi, Y. '19]

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v\in V} b_v\left(\sigma_v\right)$

current sample: $X \sim \mu$

$R \leftarrow \{v \in V \mid v$ is updated or incident to updated $e\}$;

while $R \neq \emptyset$ **do**

for every $v \in R$, resample $X_v \sim b_v$ independently;

every *internal* $e = \{u, v\} \subseteq R$ accepts ind. w.p. $A_e(X_u, X_v)$;

every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts ind. w.p.

$\propto \dfrac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)}$; // X_u^{old}: X_u before resampling

$R \leftarrow \bigcup_{e \text{ rejects}} e$;

Rejection Sampling

Gibbs distribution: $$\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v\in V} b_v\left(\sigma_v\right)$$

Rejection sampling: $$(\boldsymbol{X} \mid R = \emptyset) \sim \mu$$

for every $v \in R$, sample $X_v \sim b_v$ independently;

every edge $e = \{u, v\} \in E$ accepts independently w.p. $A_e(X_u, X_v)$;

$$R \leftarrow \bigcup_{e \text{ rejects}} e$$

A Moser-Tardos style algorithm

[Feng, Vishnoi, Y. '19]

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e(\sigma_u, \sigma_v) \prod_{v\in V} b_v(\sigma_v)$

current sample: $X \sim \mu$

$R \leftarrow \{v \in V \mid v$ is updated or incident to updated $e\}$;

while $R \neq \emptyset$ **do**

for every $v \in R$, resample $X_v \sim b_v$ independently;

every *internal* $e = \{u, v\} \subseteq R$ accepts ind. w.p. $A_e(X_u, X_v)$;

every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts ind. w.p.

$\propto \frac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)}$; // X_u^{old}: X_u before resampling

$R \leftarrow \bigcup_{e \text{ rejects}} e$;

Partial Rejection Sampling (PRS): [Guo, Jerrum, Liu '17]

A *heat-bath* based algorithm

[Feng, Guo, Y. '19]

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e(\sigma_u, \sigma_v) \prod_{v\in V} b_v(\sigma_v)$

current sample: $X \sim \mu$

$R \leftarrow \{v \in V \mid v$ is updated or incident to updated $e\}$;

while $R \neq \emptyset$ **do**

pick a random $u \in R$;

with probability $\propto \dfrac{1}{\mu_u(X_u \mid X_{N(u)})}$ **do**

resample $X_u \sim \mu_u(\cdot \mid X_{N(u)})$;

delete u from R;

else

add all neighbors of u to R;

constant factor depends only on $X_{R\cap N(u)}$

heat-bath

a.k.a. Glauber dynamics

Gibbs sampling

$N(u) \triangleq$ neighborhood of u

M-T dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \emptyset$ **do**
  for every $v \in R$, resample $X_v \sim b_v$ independently;
  every *internal* $e = \{u, v\} \subseteq R$ accepts w.p. $A_e(X_u, X_v)$;
  every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts w.p. $\propto \dfrac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)}$;
  $R \leftarrow \bigcup_{e \text{ rejects}} e$;

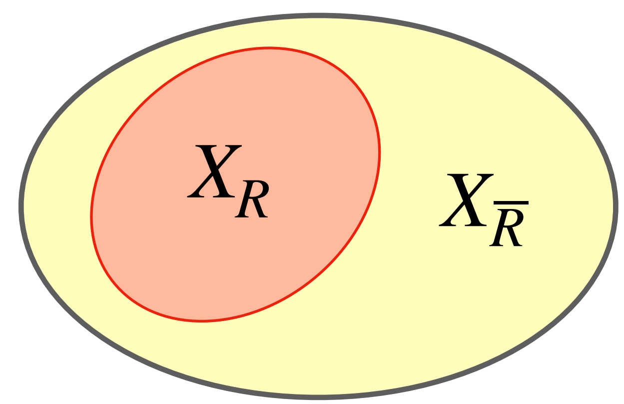
heat-bath dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \emptyset$ **do**
  pick a random $u \in R$;
  with probability $\propto \dfrac{1}{\mu_u(X_u \mid X_{N(u)})}$ **do**
    resample $X_u \sim \mu_u(\cdot \mid X_{N(u)})$;
    delete u from R;
  else
    add all neighbors of u to R;

chain: $(\boldsymbol{X}, R) \longrightarrow (\boldsymbol{X}', R')$

configuration $\boldsymbol{X} \in [q]^V$

set $R \subseteq V$ of "incorrect" vertices



Conditional Gibbs property:

Given any R and X_R, the $X_{\overline{R}}$ always follows $\mu_{\overline{R}}^{X_R}$.

$\Rightarrow$ $\boldsymbol{X} \sim \mu$ when $R = \emptyset$

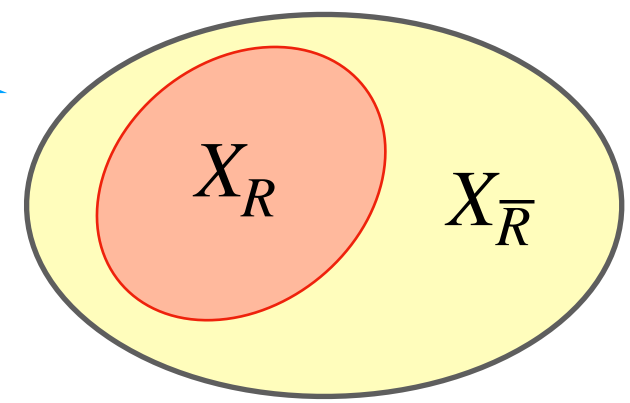
(marginal distribution on $\overline{R}$ conditioned on X_R)

Equilibrium Condition

chain: $(x, R) \xrightarrow{P} (y, R')$

Conditional Gibbs property:

Given any R and X_R, the $X_{\overline{R}}$ always follows $\mu_{\overline{R}}^{X_R}$.



Fix any $\sigma \in [q]^R, \tau \in [q]^{R'}$.

$\forall y \in [q]^V$ that $y_{R'} = \tau$:

$$\mu_{\overline{R'}}^{\tau}(y_{\overline{R'}}) \propto \sum_{\substack{x \in [q]^V \\ x_R = \sigma}} \mu_{\overline{R}}^{\sigma}(x_{\overline{R}}) \cdot P((x, R), (y, R'))$$

M-T dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \varnothing$ **do**
for every $v \in R$, resample $X_v \sim b_v$ independently;
every *internal* $e = \{u, v\} \subseteq R$ accepts w.p. $A_e(X_u, X_v)$;
every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts w.p. $\propto \frac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)}$;
$R \leftarrow \bigcup_{e \text{ rejects}} e$;

heat-bath dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \varnothing$ **do**
pick a random $u \in R$;
with probability $\propto \frac{1}{\mu_u(X_u \mid X_{N(u)})}$ **do**
resample $X_u \sim \mu_u(\cdot \mid X_{N(u)})$;
delete u from R;
else
add all neighbors of u to R;

chain: $(X, R) \longrightarrow (X', R')$

Conditional Gibbs property:

Given any R and X_R, the $X_{\overline{R}}$ always follows $\mu_{\overline{R}}^{X_R}$.

- defined in [Feng, Vishnoi, Y. '19], also implicitly in [Guo, Jerrum '18]
- satisfied invariantly by the M-T and heat-bath dynamic samplers
$\Longrightarrow$ Las Vegas perfect samplers (*interruptible*)
- retrospectively, holds for *Partial Rejection Sampling* [Guo, Jerrum, Liu '17] and *Randomness Recycler* [Fill, Huber '00]

heat-bath dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \emptyset$ **do**
  pick a random $u \in R$;
  with probability $\propto \dfrac{1}{\mu_u(X_u \mid X_{N(u)})}$ **do**
    resample $X_u \sim \mu_u(\cdot \mid X_{N(u)})$;
    delete u from R;
  else
    add all neighbors of u to R;

chain:

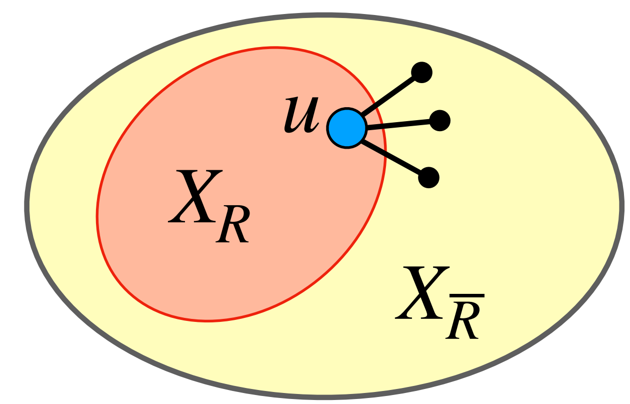
$$(\boldsymbol{X}, R) \longrightarrow (\boldsymbol{X}', R')$$

Conditional Gibbs property:

Given any R and X_R,
$X_{\overline{R}}$ always follows $\mu_{\overline{R}}^{X_R}$.

success case:

$$R' = R \setminus \{u\}$$



invariant CGP: $X_{\overline{R}} \sim \mu_{\overline{R}}^{X_R}$

filter

Bayes law   depends only on X_R

$$\Pr[\text{ filter succeeds }] \propto \frac{\mu_{\overline{R}}^{X_{R'}}(X_{\overline{R}})}{\mu_{\overline{R}}^{X_R}(X_{\overline{R}})} = \frac{\mu_u^{X_{R'}}(X_u)}{\mu_u^{X_{N(u)}}(X_u)} \propto \frac{1}{\mu_u^{X_{N(u)}}(X_u)}$$

$$X_{\overline{R}} \sim \mu_{\overline{R}}^{X_{R'}} \;+\; X_u \sim \mu_u(\cdot \mid X_{N(u)}) \implies X_{\overline{R'}} \sim \mu_{\overline{R'}}^{X_{R'}}$$

invariant CGP

heat-bath dynamic sampler

$R \leftarrow$ {vertices affected by update};

while $R \neq \emptyset$ **do**

pick a random $u \in R$;

with probability $\propto \dfrac{1}{\mu_u(X_u \mid X_{N(u)})}$ **do**

resample $X_u \sim \mu_u(\cdot \mid X_{N(u)})$;

delete u from R;

else

add all neighbors of u to R;

invariant CGP: $X_{\overline{R}} \sim \mu_{\overline{R}}^{X_R}$

chain:

$$(\boldsymbol{X}, R) \longrightarrow (\boldsymbol{X}', R')$$

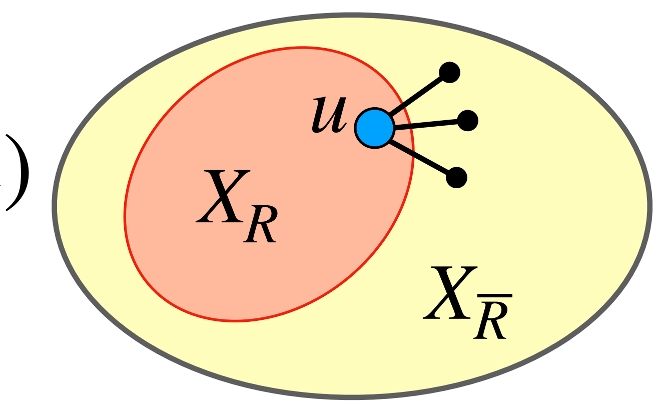
Conditional Gibbs property:

Given any R and X_R,

$X_{\overline{R}}$ always follows $\mu_{\overline{R}}^{X_R}$.

failure case:

$$R' = R \cup N(u)$$



all vertices whose spins are revealed are includes in R'

➡ invariant CGP: $X_{\overline{R'}} \sim \mu_{\overline{R'}}^{X_{R'}}$

M-T dynamic sampler

$R \leftarrow$ {vertices affected by update};
while $R \neq \emptyset$ do
  for every $v \in R$, resample $X_v \sim b_v$ independently;
  every *internal* $e = \{u, v\} \subseteq R$ accepts w.p. $A_e(X_u, X_v)$;
  every *boundary* $e = \{u, v\}$ with $u \in R, v \notin R$ accepts w.p. $\propto \frac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)}$;
  $R \leftarrow \bigcup_{e \text{ rejects}} e$;

Efficiency Analysis:

$$\mathbf{E}[H(R') \mid R] < H(R)$$

set R (or some potential of it) decays in expectation in every step in the worst case

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\} \in E} A_e(\sigma_u, \sigma_v) \prod_{v \in V} b_v(\sigma_v)$

- $\min A_e > 1 - \frac{1}{4\Delta}$, where Δ is the max-degree
- Ising model with inverse temp. β: $\mathrm{e}^{-2|\beta|} > 1 - \frac{1}{2.221\Delta + 1}$
- hardcore model with fugacity $\lambda < \frac{1}{\sqrt{2}\Delta - 1}$

- $X' \sim \mu'$ is returned within $O(\Delta \lvert \texttt{update} \rvert)$ resamples
- $O(\Delta \lvert E \rvert)$ time Las-Vegas perfect sampler

heat-bath dynamic sampler (*block* version)

$R \leftarrow$ {vertices affected by update};
while $R \neq \emptyset$ **do**
  pick a random $u \in R$ and r-ball $B = B_r(u)$;
  with probability $\propto \dfrac{1}{\mu_u(X_u \mid X_{\partial B})}$ **do**
    resample $X_B \sim \mu_B(\cdot \mid X_{\partial B})$;
    delete u from R;
  else
    add all boundary vertices in ∂B to R;

strong spatial mixing (SSM):

$$d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^\tau) \leq \exp(-\Omega(\mathrm{dist}(v, \sigma \oplus \tau)))$$

sub-exp neighborhood growth:

$$\forall v, \quad |\partial B_r(v)| \leq \exp(o(r))$$

E.g. $\mathbb{Z}^d$

On graphs with *sub-exp* neighborhood growth:

SSM → [Feng, Guo, Y. '19] → $O(|\texttt{update}|)$ dynamic sampling → $O(n)$-time perfect sampler

SSM ↔ [Dyer, Sinclair, Vigoda, Weitz '04] [Goldberg, Martin, Paterson'05] ↔ $O(n \log n)$ mixing of block MC

A data structure approach

[Feng, He, Sun, Y. '20]

trajectory for single-site dynamics:

$$X_0,\ X_1,\ X_2,\ \cdots\cdots\cdots\cdots,\ X_T$$
$$\|$$
$$X'_0,\ X'_1,\ X'_2,\ \cdots\cdots\cdots\cdots,\ X'_T$$

very rare

Gibbs distribution: $\mu(\sigma) \propto \exp\left(\sum_{v\in V}\phi_v(\sigma_v) + \sum_{e=\{u,v\}\in E}\phi_e(\sigma_u,\sigma_v)\right)$

Update of graphical model: $\mathbf{\Phi} \to \mathbf{\Phi}'$ with $\texttt{diff} \triangleq \|\Phi - \Phi'\|_1$

<u>Dobrushin-Shlosman condition</u>
(path coupling cond.)

➡

$O(\texttt{diff} \cdot \Delta \log n)$ steps
differ in single-site transition

efficient data structure (with a space overhead)
for resolving such dynamic update

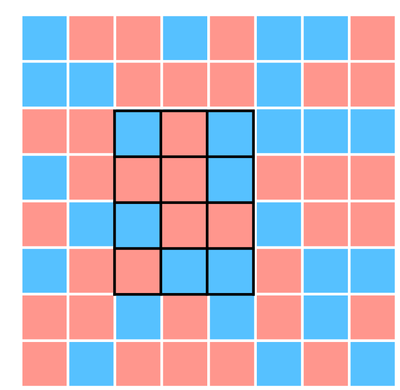
Caveats

equilibrium:
conditional Gibbs property

→ Correctness:

- dynamic sampling (succinct in space)
- perfect sampling (interruptible)

- Does the *conditional Gibbs property* require stronger condition to maintain on general graphs?
  - e.g. expanders
- In dynamic sampling: the updated sample and original sample are correlated.
  - *far-apart* spins: decay of correlation
  - *nearby* spins: possibly resampled



Distributed Gibbs Sampling

Moser-Tardos sampler

$R \leftarrow V$; // used for *static* sampling

while $R \neq \emptyset$ **do**

in parallel: for every $v \in R$, resample $X_v \sim b_v$ independently;

in parallel: every internal $e = \{u, v\} \subseteq R$ accepts w.p. $A_e(X_u, X_v)$;

in parallel: every boundary $e = \{u, v\}$ with $u \in R, v \notin R$ accepts w.p.

$$\propto \frac{A_e(X_u, X_v)}{A_e(X_u^{\text{old}}, X_v)};$$

$$R \leftarrow \bigcup_{e \text{ rejects}} e;$$

- $\min A_e > 1 - \frac{1}{4\Delta}$
- Ising model: $e^{-2|\beta|} > 1 - \frac{1}{2.221\Delta + 1}$
- hardcore model: $\lambda < \frac{1}{\sqrt{2}\Delta - 1}$

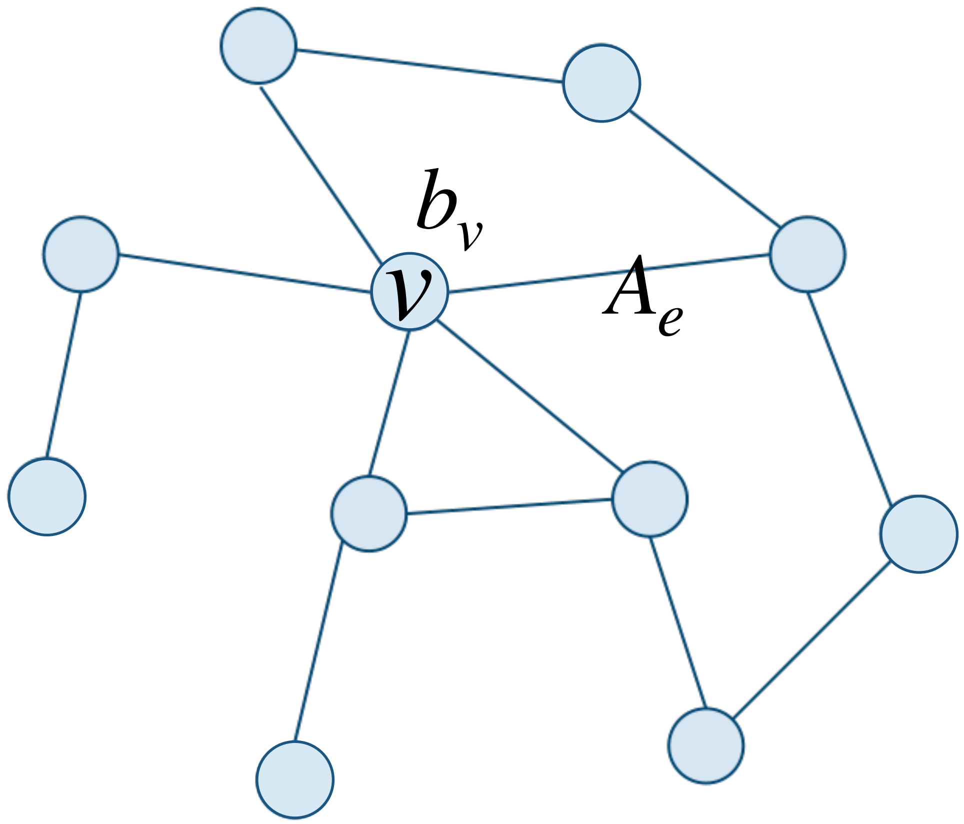
$\Rightarrow$ $X \sim \mu$ is returned in $O(\log n)$ rounds in expectation

Distributed Gibbs Sampling

Gibbs distribution:

$$\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v\in V} b_v\left(\sigma_v\right)$$

network $G = (V, E)$



Distributed algorithm:

upon termination return $\boldsymbol{X} \in [q]^V$

- perfect sampling: $\boldsymbol{X} \sim \mu$
- approx. sampling: $d_{\mathrm{TV}}(\boldsymbol{X}, \mu) \leq \epsilon$

[Guo, Jerrum, Liu '17] [Feng, Sun, Y. '17]:

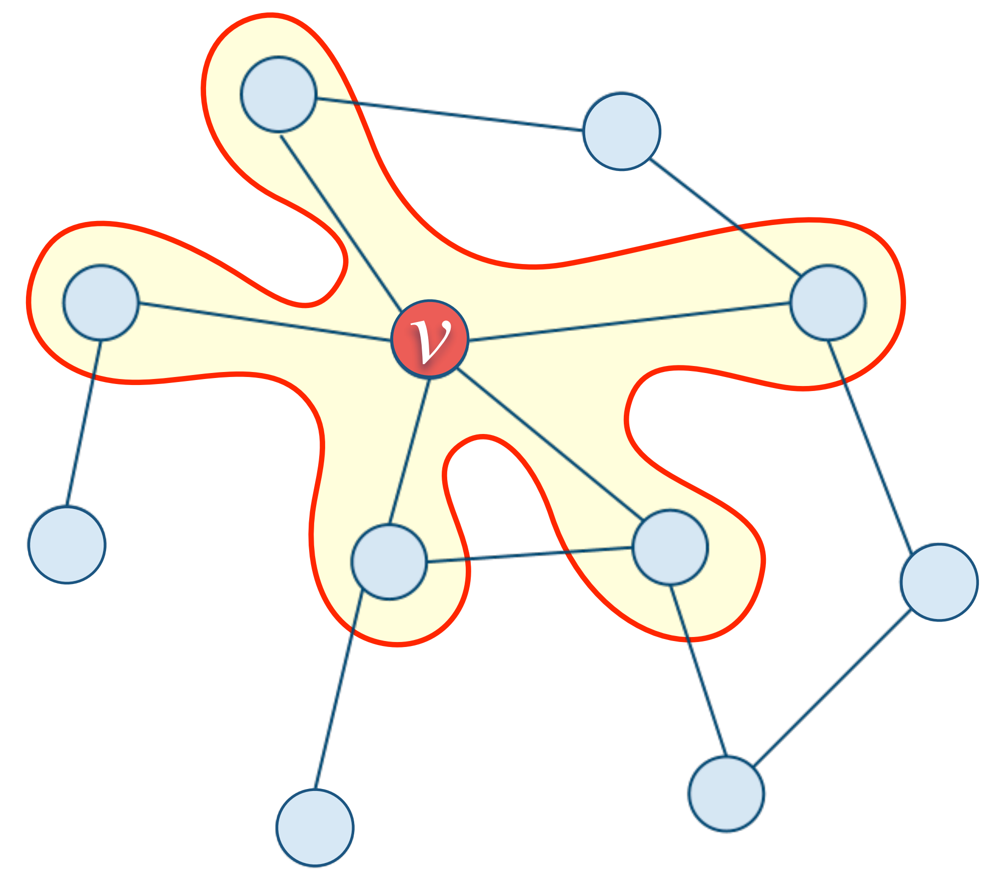
approx. sampling requires $\Omega(\log n)$ rounds for $\epsilon < 1/3$

Distributed Gibbs Sampling

Single-site dynamics $\boldsymbol{X} \to \boldsymbol{X}'$:

pick a random $v \in V$;

update X_v according to $X_{N^+(v)}$;



- typical rapid mixing time: $O(n \log n)$
- requires $\Omega(n \log n)$ steps to mix
  [Hayes, Sinclair '07]

Parallelize single-site dynamics: $O(n \log n)$ steps $\to$ $O(\log n)$ rounds

- chromatic scheduler: no adjacent concurrent update $\Longrightarrow$ $\Omega(\Delta \log n)$ rounds
- *Hogwild*! (independently random scheduler): biased,
  but may be good enough for local or Lipschitz estimators

  [Niu, Recht, Ré, Wright '11], [De Sa, Olukotun, Ré '16], [Daskalakis, Dikkala, Jayanti '18]

Parallel Metropolis Filters

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v \in V} b_v\left(\sigma_v\right)$

A Metropolis chain:

$\boldsymbol{X} \to \boldsymbol{X}'$

pick a random $v \in V$;

propose a random $c_v \sim b_v$;

accept and $X_v \leftarrow c_v$ w.p. $\prod_{u \in N(v)} A_{\{u,v\}}(X_u, c_v)$;

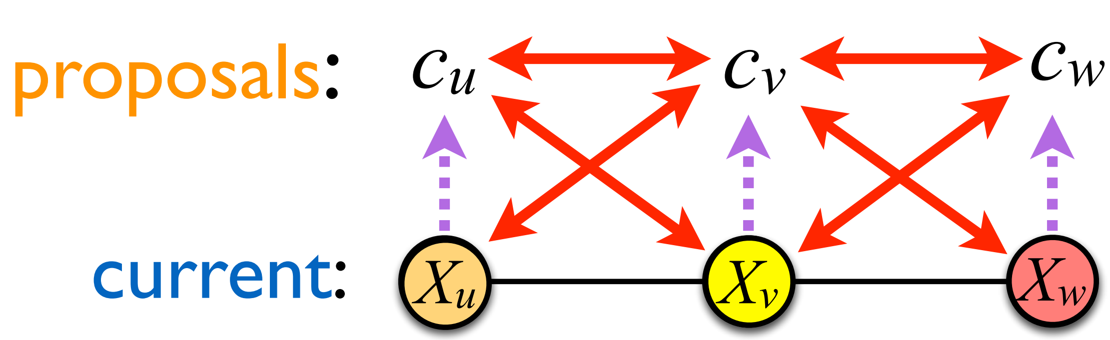
Local-Metropolis chain: [Feng, Sun, Y. '17]

every $v \in V$ independently proposes $c_v \sim b_v$;

every $e = \{u, v\} \in E$ accepts independently w.p.

$$A_e(X_u, c_v) \cdot A_e(c_u, X_v) \cdot A_e(c_u, c_v);$$

every $v \in V$ accepts and $X_v \leftarrow c_v$ if all its incident edges accepted;



Parallel Metropolis Filters

Gibbs distribution: $\mu(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e\left(\sigma_u, \sigma_v\right) \prod_{v\in V} b_v\left(\sigma_v\right)$

Local-Metropolis chain: [Feng, Sun, Y. '17]

every $v \in V$ independently proposes $c_v \sim b_v$;

every $e = \{u, v\} \in E$ accepts independently w.p.

$$A_e(X_u, c_v) \cdot A_e(c_u, X_v) \cdot A_e(c_u, c_v);$$

every $v \in V$ accepts and $X_v \leftarrow c_v$ if all its incident edges accepted;

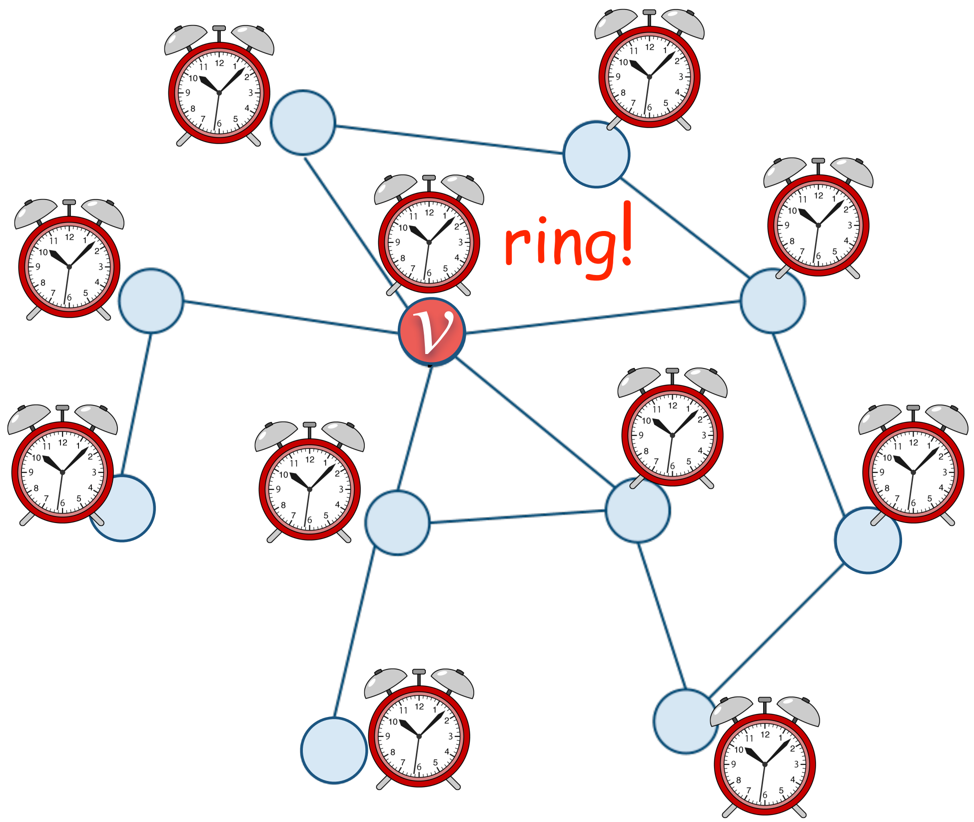
- sample from μ when stationary
- improved in [Fischer, Ghaffari '18] [Feng, Hayes, Y. '18]:

  path coupling for single-site Metropolis $\Longrightarrow O(\log n)$ rounds mixing

- applied in LCA model [Biswas, Rubinfeld, Yodpinyanee '19]

Distributed simulation of Continuous chain

rate-1 Poisson clocks



when the clock at $v \in V$ rings:

update X_v according to $X_{N^+(v)}$;

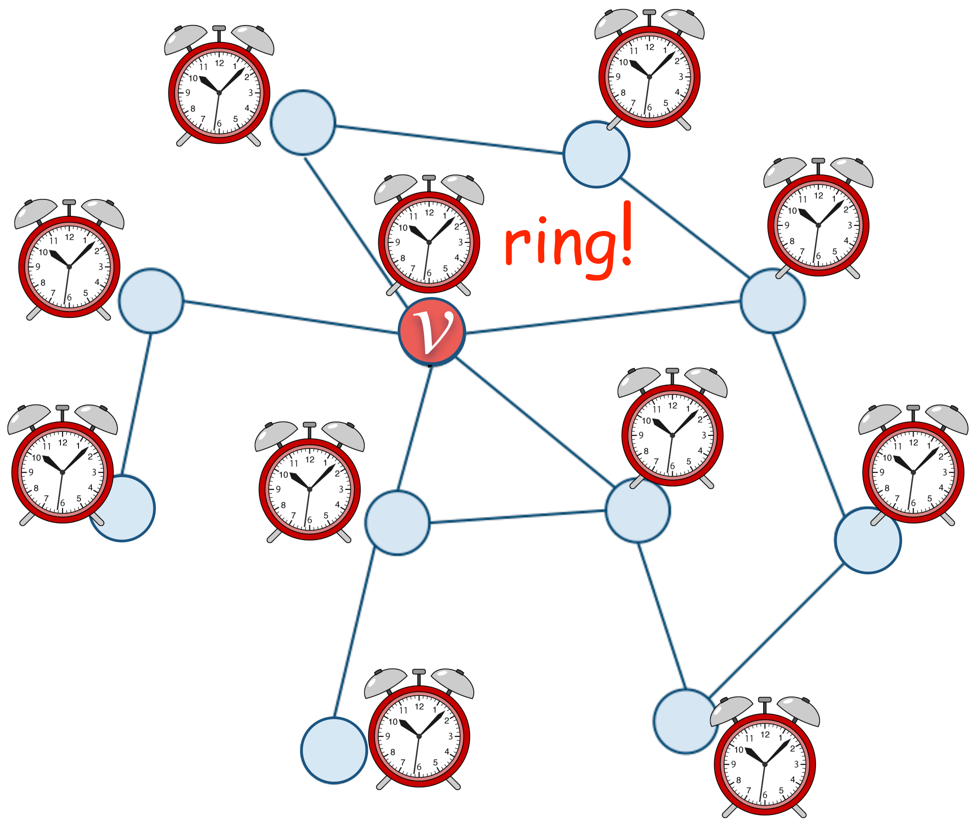
We want: faithfully simulate continuous time T in $O(T)$ rounds

To resolve an update at $v \in V$ at time t:

- naive: wait until $X_{N^+(v)}$ at time t is known to v $\implies \Omega(\Delta T)$ rounds
- *resolve update in advance*: [Feng, Hayes, Y. '19]

Distributed simulation of Continuous chain

rate-1 Poisson clocks



Metropolis Chain

when the clock at $v \in V$ rings:

propose a random c_v;
accept and $X_v \leftarrow c_v$ w.p. $\mathtt{Bias}(c_v, X_{N^+(v)})$;

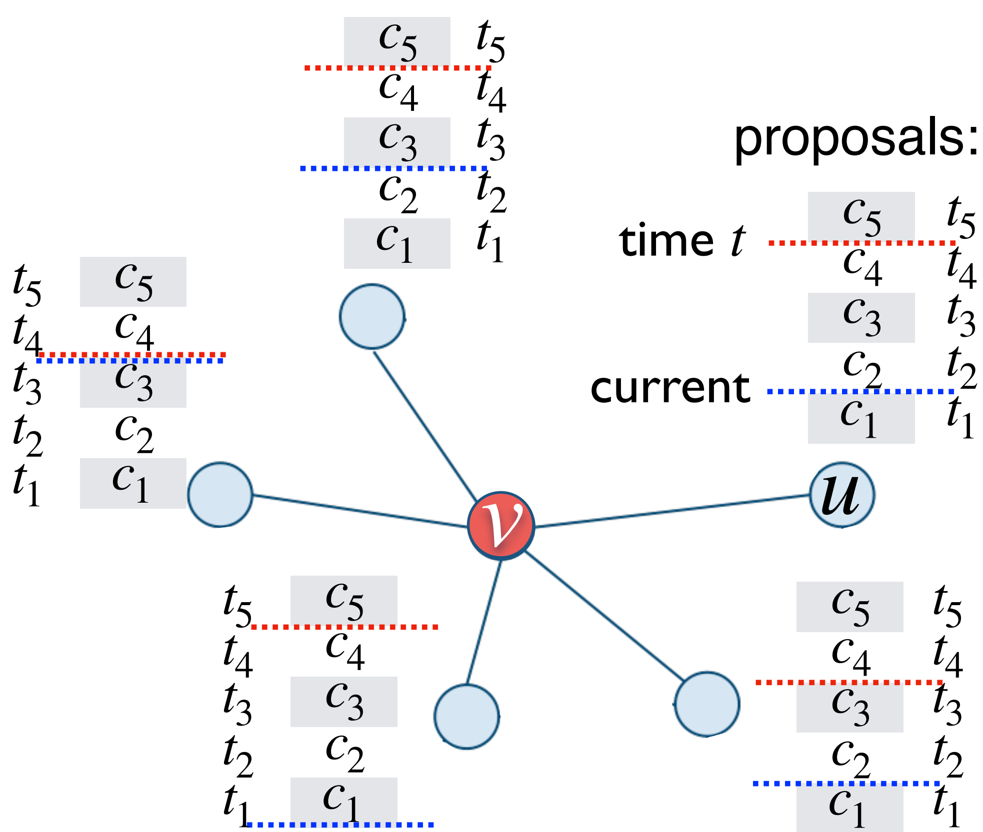
We want: faithfully simulate continuous time T in $O(T)$ rounds

To resolve a proposal c_v of $v \in V$ at time t:

- naive: wait until $X_{N^+(v)}$ at time t is known to v $\implies \Omega(\Delta T)$ rounds
- *resolve update in advance*: [Feng, Hayes, Y. '19]

Distributed simulation of Continuous chain

rate-1 Poisson clocks



Metropolis Chain

when the clock at $v \in V$ rings:

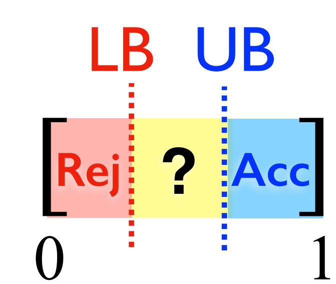
> propose a random c_v;
>
> accept and $X_v \leftarrow c_v$ w.p. $\texttt{Bias}(c_v, X_{N^+(v)})$;

We want: faithfully simulate continuous time T in $O(T)$ rounds

To resolve a proposal c_v of $v \in V$ at time t:

- naive: wait until $X_{N^+(v)}$ at time t is known to v $\Longrightarrow \Omega(\Delta T)$ rounds
- *resolve update in advance*: [Feng, Hayes, Y. '19]

  flip a coin with $\texttt{Bias}(c_v, X_{N^+(v)})$ before $X_{N^+(v)}$ is fully known



Faithfully simulate time-T continuous Metropolis chain in $O(T + \log n)$ rounds.

[Feng, Hayes, Y. '19]

model	Efficient simulation	Necessary condition for mixing
q-coloring	∃ constant $C>0$ $q>C\Delta$	$q \geq \Delta+2$
Ising model with temperature β	∃ constant $C>0$ $1 - e^{-2\|\beta\|} < \frac{C}{\Delta}$	$1 - e^{-2\|\beta\|} < \frac{2}{\Delta}$
hardcore model with fugacity λ	∃ constant $C>0$ $\lambda < \frac{C}{\Delta}$	$\lambda < \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-2}$

Summary

- Many new ideas for dynamic/distributed sampling.
- **Open problems:**
  - conditional Gibbs property *vs*. phase transition
    - e.g. q-coloring on general graphs for $q = O(\Delta)$
  - impact of correlations in dynamic sampling applications
    - e.g. inference, approximate counting
  - parallelization of general single-site dynamics
    - e.g. Glauber dynamics
  - use these new ideas to improve sampling in classic setting
    - e.g. Moser-Tardos style tight analysis of sampling

Thank you!