

What can be sampled *locally*?

Yitong Yin
Nanjing University

Joint work with: Weiming Feng, Yuxin Sun

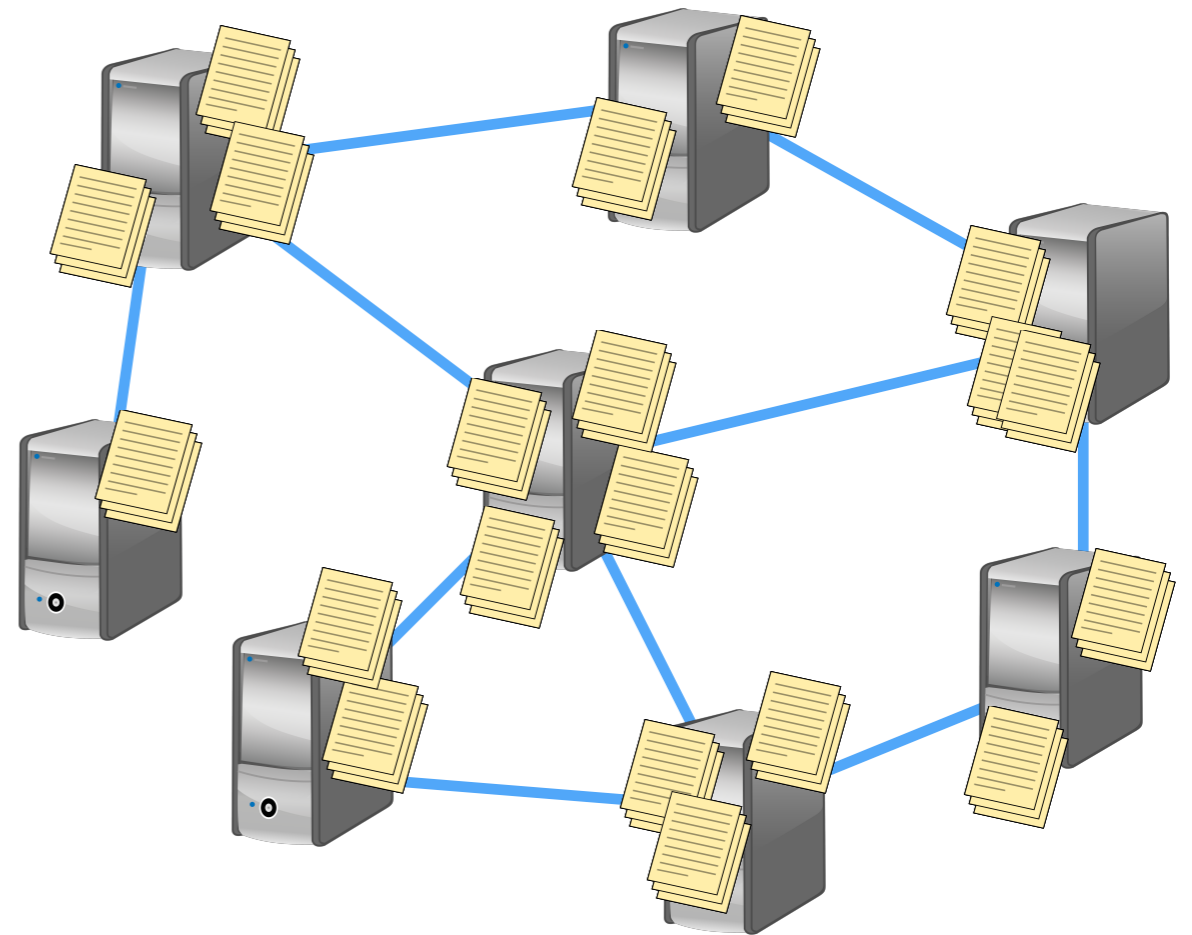
Local Computation

“What can be computed locally?”

[Noar, Stockmeyer, STOC’93, SICOMP’95]

the *LOCAL* model [Linial ’87]:

- Communications are **synchronized**.
- In each **round**: each node can send messages of **unbounded sizes** to all its neighbors.
- Local computations are **free**.
- **Complexity**: # of rounds to terminate in the worst case.
- In t rounds: each node can collect information up to distance t .



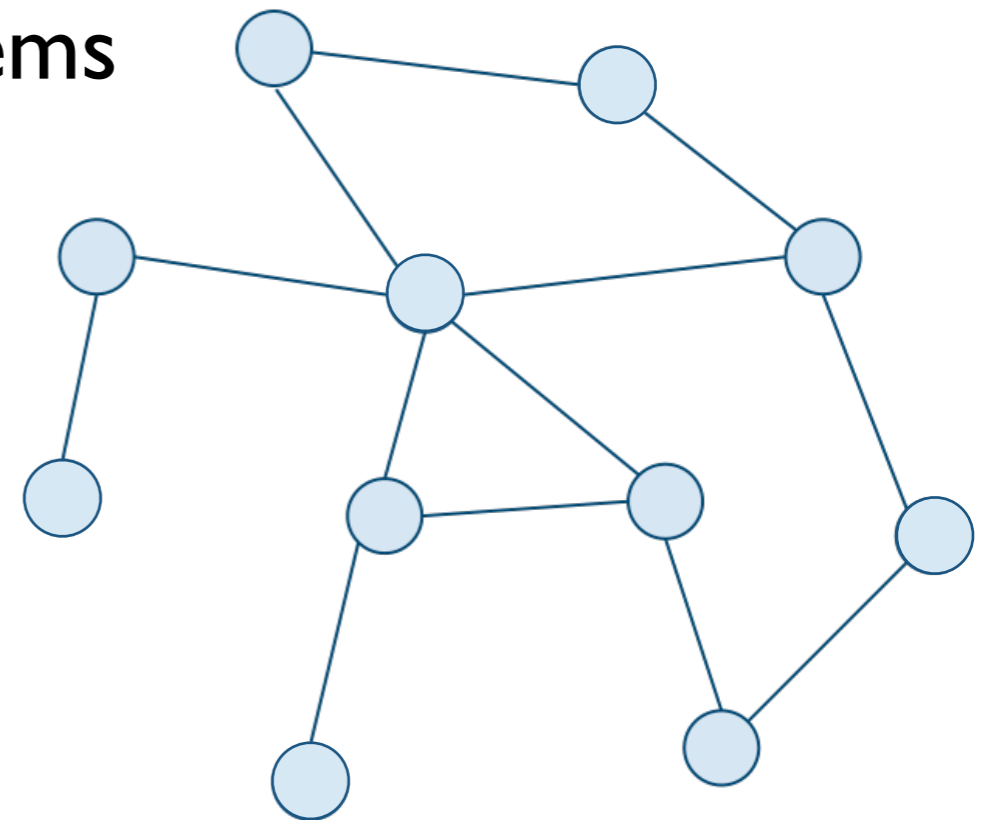
Local Computation

the *LOCAL* model [Linial '87]:

- In t rounds: each node can collect information up to distance t .

Locally Checkable Labeling (LCL) problems [Noar, Stockmeyer '93]:

- CSPs with **local constraints**.
- **Construct a feasible solution:**
vertex/edge coloring, Lovász local lemma
 - **Find local optimum:** MIS, MM
 - **Approximate global optimum:**
maximum matching, minimum vertex cover, minimum dominating set



network $G(V, E)$

Q: “What locally definable problems are locally computable?”

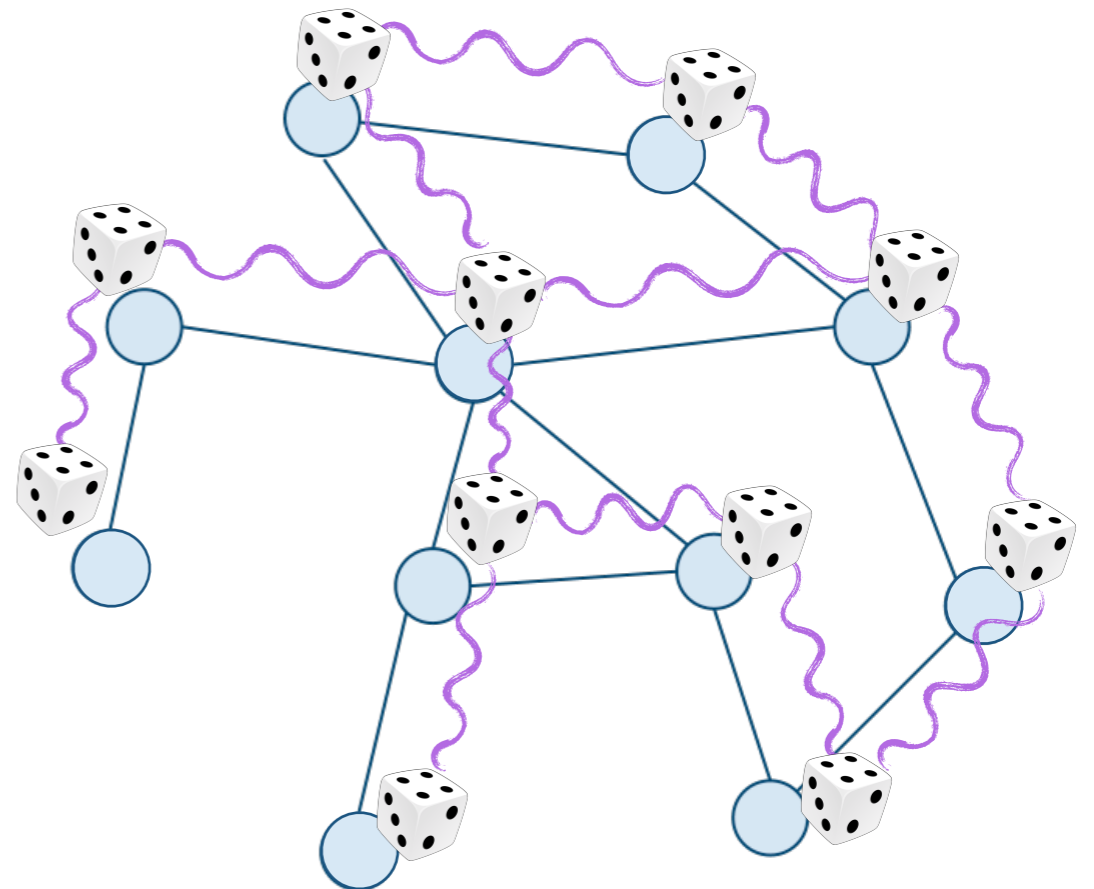
by local constraints

in $O(1)$ rounds
or in small number of rounds

“What can be *sampled* locally?”

- CSP with **local constraints** on the network:
 - proper q -coloring;
 - independent set;
- Sample a **uniform** random feasible solution:
 - distributed algorithms
(in the *LOCAL* model)

network $G(V,E)$:



Q: “What locally definable joint distributions are locally sample-able?”

Markov Random Fields

(MRF)

- Each vertex corresponds to a **variable** with finite domain $[q]$.
- Each edge $e=(u,v)\in E$ imposes a **weighted binary constraint**:

$$A_e : [q]^2 \rightarrow \mathbb{R}_{\geq 0}$$

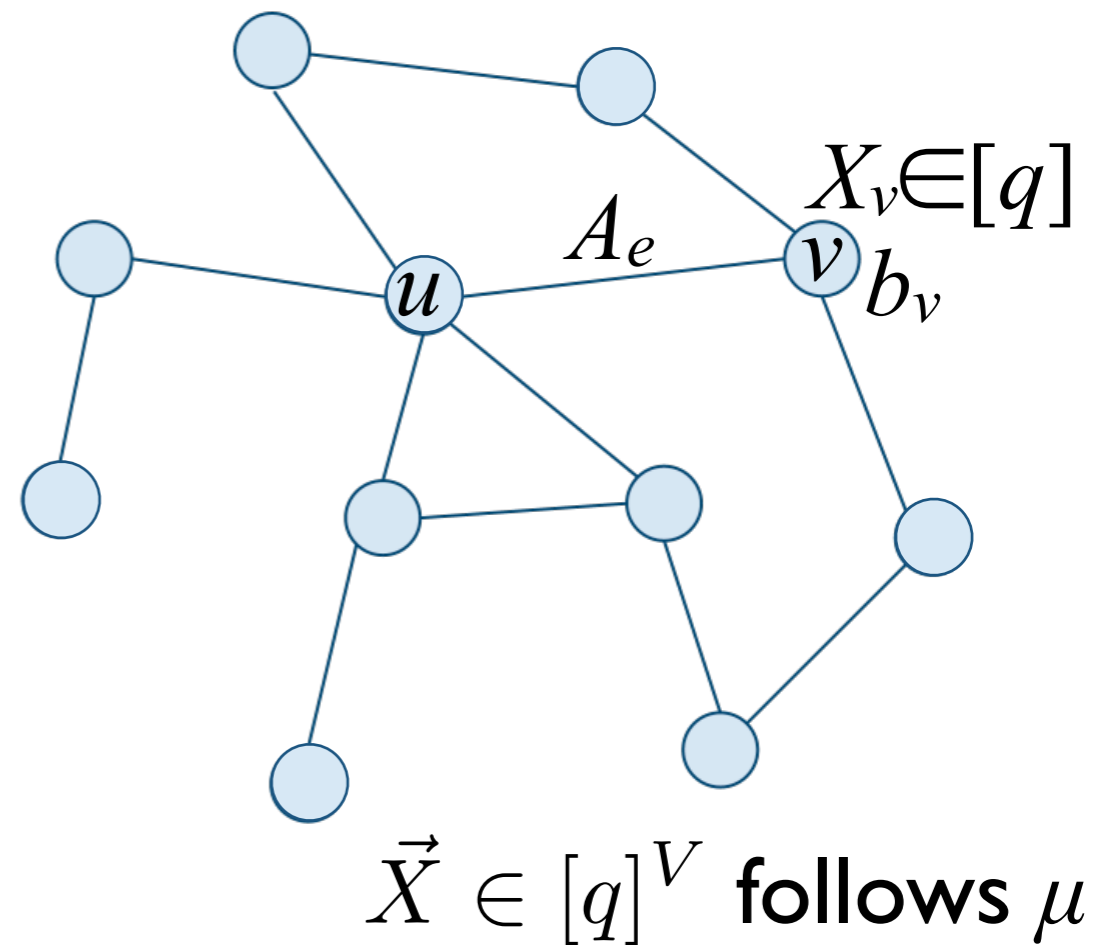
- Each vertex $v\in V$ imposes a **weighted unary constraint**:

$$b_v : [q] \rightarrow \mathbb{R}_{\geq 0}$$

- **Gibbs distribution** $\mu : \forall \sigma \in [q]^V$

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

network $G(V,E)$:



Markov Random Fields

(MRF)

- **Gibbs distribution** μ : $\forall \sigma \in [q]^V$ network $G(V, E)$:

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

- **proper q -coloring:**

$$A_e = \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & 1 & & \\ & 1 & \ddots & & \\ & & & \ddots & 0 \end{bmatrix} \quad b_v = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

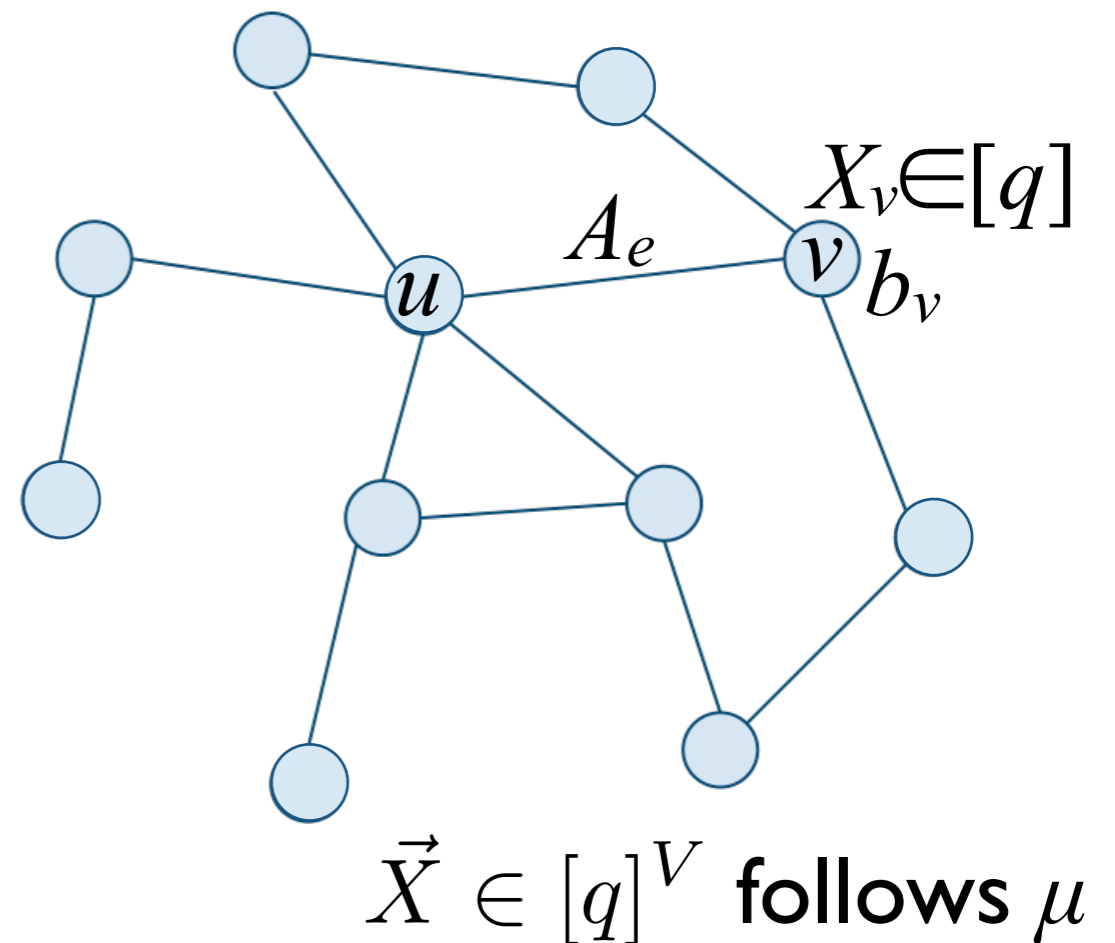
- **independent set:**

$$A_e = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad b_v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- **local conflict colorings:**

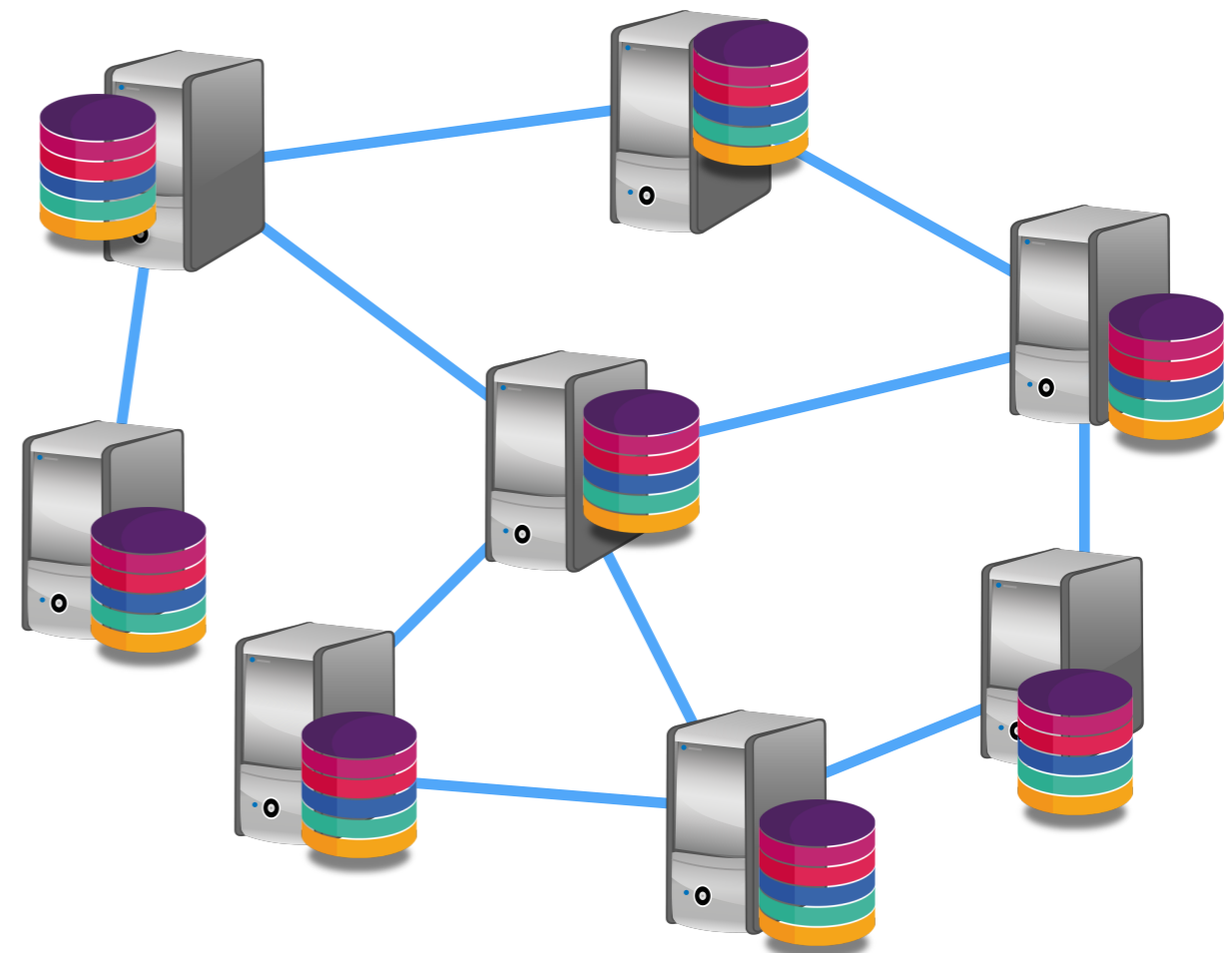
[Fraigniaud, Heinrich, Kosowski FOCS'16]

$$A_e \in \{0, 1\}^{q \times q}, b_v \in \{0, 1\}^q$$



A Motivation: *Distributed Machine Learning*

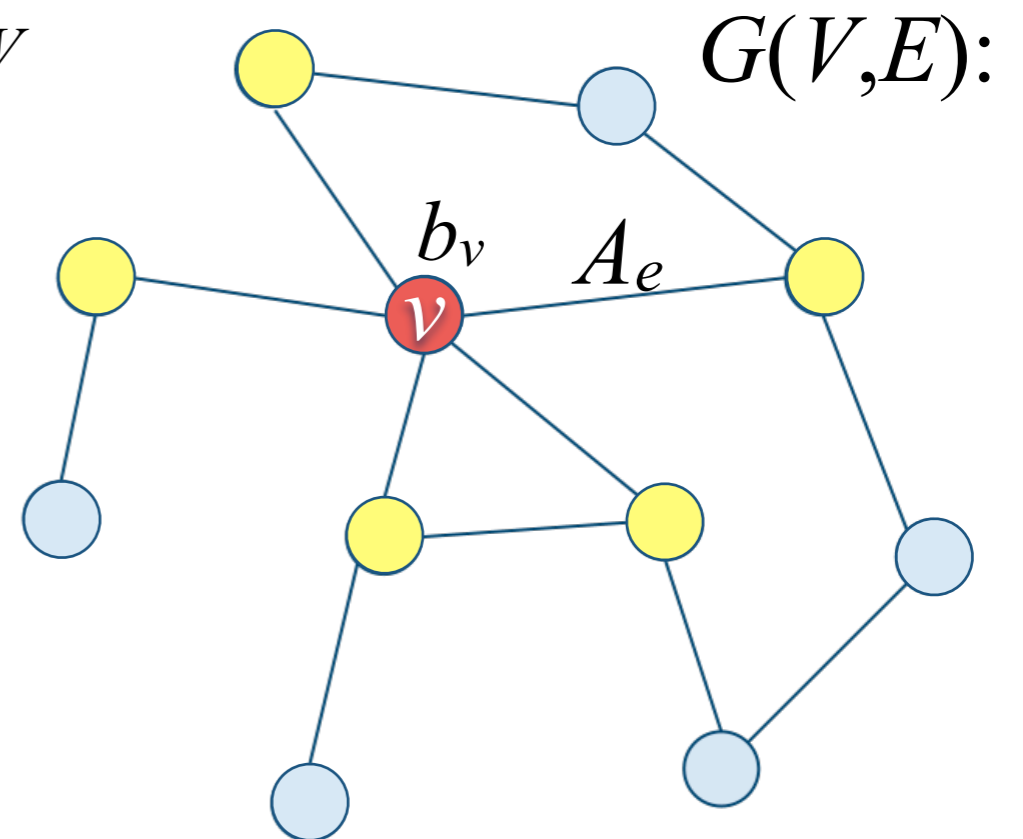
- Data are stored in a distributed system.
- Sampling from a *probabilistic graphical model* (e.g. the *Markov random field*) by distributed algorithms.



Glauber Dynamics

starting from an arbitrary $X_0 \in [q]^V$
transition for $X_t \rightarrow X_{t+1}$:

pick a **uniform random** vertex v ;
resample $X(v)$ according to the
marginal distribution induced by μ at
vertex v **conditioning on** $X_t(N(v))$;



marginal distribution:

$$\Pr[X_v = x \mid X_{N(v)}] = \frac{b_v(x) \prod_{u \in N(v)} A_{(u,v)}(X_u, x)}{\sum_{y \in [q]} b_v(y) \prod_{u \in N(v)} A_{(u,v)}(X_u, y)}$$

stationary distribution: μ

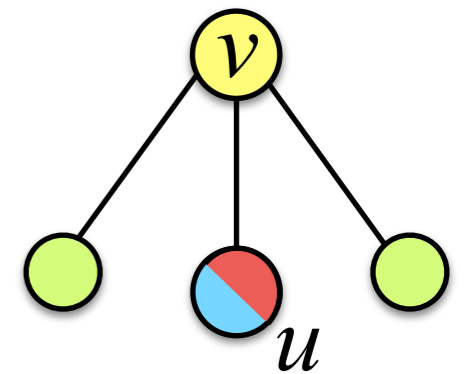
$$\text{MRF: } \forall \sigma \in [q]^V, \\ \mu(\sigma) \propto \prod_{e=(u,v) \in E} A_e(\sigma_u, \sigma_v) \prod_{v \in V} b_v(\sigma_v)$$

mixing time: $\tau_{\text{mix}} = \max_{X_0} \min \left\{ t \mid d_{\text{TV}}(X_t, \mu) \leq \frac{1}{2e} \right\}$

Mixing of Glauber Dynamics

influence matrix $\{\rho_{v,u}\}_{v,u \in V}$:

$\rho_{v,u}$: max discrepancy (in total variation distance) of marginal distributions at v caused by any pair σ, τ of boundary conditions that differ only at u



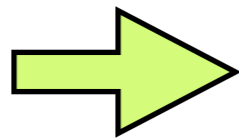
Dobrushin's condition:

$$\|\rho\|_{\infty} = \max_{v \in V} \sum_{u \in V} \rho_{v,u} \leq 1 - \epsilon$$

contraction of **one-step optimal coupling** in the **worst case** w.r.t. **Hamming distance**

Theorem (Dobrushin '70; Salas, Sokal '97):

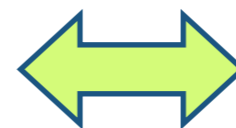
Dobrushin's condition



$\tau_{\text{mix}} = O(n \log n)$
for Glauber dynamics

for **q -coloring**:

Dobrushin's condition



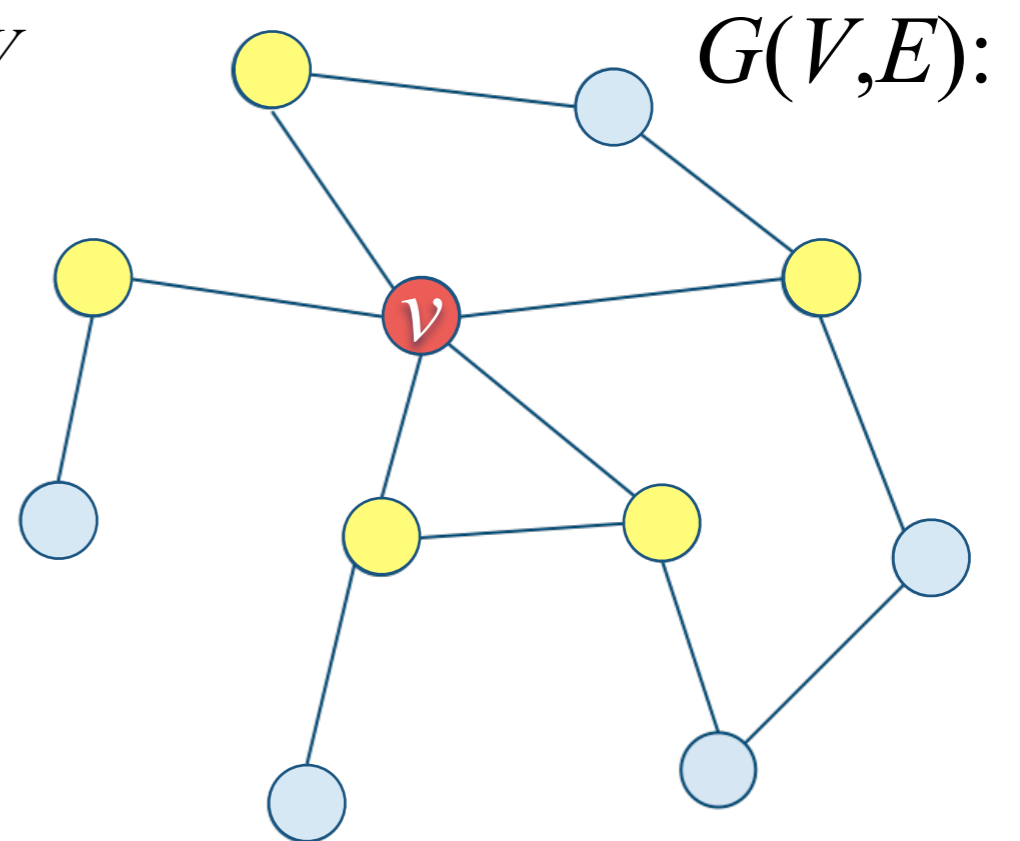
$q \geq (2 + \epsilon)\Delta$
 $\Delta = \text{max-degree}$

Parallelization

Glauber dynamics:

starting from an arbitrary $X_0 \in [q]^V$
transition for $X_t \rightarrow X_{t+1}$:

pick a **uniform random** vertex v ;
resample $X(v)$ according to the
marginal distribution induced by μ at
vertex v **conditioning on** $X_t(N(v))$;



Parallelization:

- **Chromatic scheduler** [folklore] [Gonzalez *et al.*, AISTAT'11]:
Vertices **in the same color class** are updated in parallel.
- **“Hogwild!”** [Niu, Recht, Ré, Wright, NIPS'11][De Sa, Olukotun, Ré, ICML'16]:
All vertices are updated in parallel, ignoring concurrency issues.

Warm-up: When Luby meets Glauber

starting from an arbitrary $X_0 \in [q]^V$

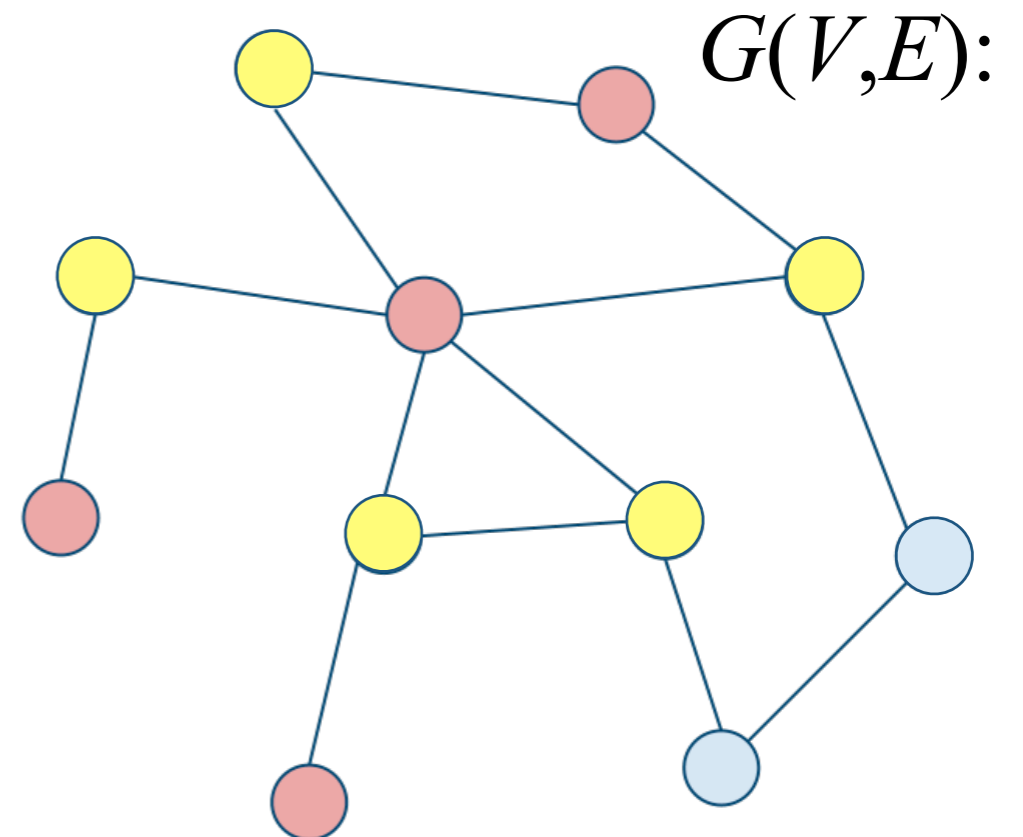
at each step, for each vertex $v \in V$:

Luby
step

independently sample a random number $\beta_v \in [0, 1]$;
if β_v is **locally maximum** among its neighborhood $N(v)$:

Glauber
step

resample $X(v)$ according to the **marginal distribution** induced by μ at vertex v **conditioning on** $X_t(N(v))$;



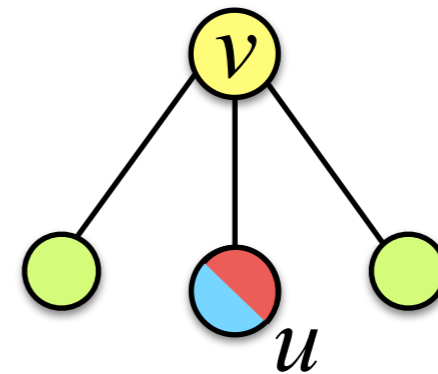
- **Luby step**: Independently sample a random **independent set**.
- **Glauber step**: For independent set vertices, update correctly according to the current marginal distributions.
- Stationary distribution: the Gibbs distribution μ .

Mixing of *LubyGlauber*

influence matrix $\{\rho_{v,u}\}_{v,u \in V}$

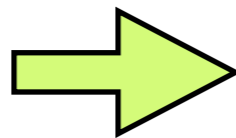
Dobrushin's condition:

$$\|\rho\|_{\infty} = \max_{v \in V} \sum_{u \in V} \rho_{v,u} \leq 1 - \epsilon$$



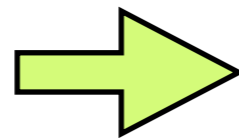
Theorem (Dobrushin '70; Salas, Sokal '97):

Dobrushin's
condition



$\tau_{\text{mix}} = O(n \log n)$
for Glauber dynamics

Dobrushin's
condition

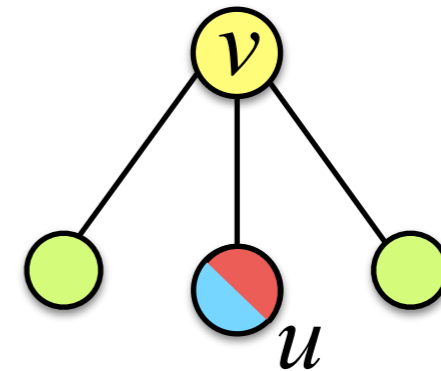


$\tau_{\text{mix}} = O(\Delta \log n)$
for the *LubyGlauber* chain

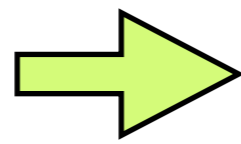
influence matrix $\{\rho_{v,u}\}_{v,u \in V}$

Dobrushin's condition:

$$\|\rho\|_\infty = \max_{v \in V} \sum_{u \in V} \rho_{v,u} \leq 1 - \epsilon$$



Dobrushin's
condition



$\tau_{\text{mix}} = O(\Delta \log n)$
for the *LubyGlauber* chain

Proof (similar to [Hayes'04] [Dyer-Goldberg-Jerrum'06]):

in the **one-step optimal coupling** (X_t, Y_t) , let $p_v^{(t)} = \Pr[X_t(v) \neq Y_t(v)]$

$$\mathbf{p}^{(t+1)} \leq M \mathbf{p}^{(t)}$$

where $M = (I - D) + D\rho$

D is diagonal and

$$D_{v,v} = \Pr[v \text{ is picked in Luby step}]$$

$$\geq \frac{1}{\deg(v) + 1}$$

$$\Pr[X_t \neq Y_t] \leq \|\mathbf{p}^{(t)}\|_1$$

$$\leq n \|\mathbf{p}^{(t)}\|_\infty$$

$$\leq n \|M\|_\infty^t \|\mathbf{p}^{(0)}\|_\infty$$

$$\leq n \left(1 - \frac{\epsilon}{\Delta + 1}\right)^t$$

Crossing the Chromatic # Barrier

Glauber

$O(n \log n)$



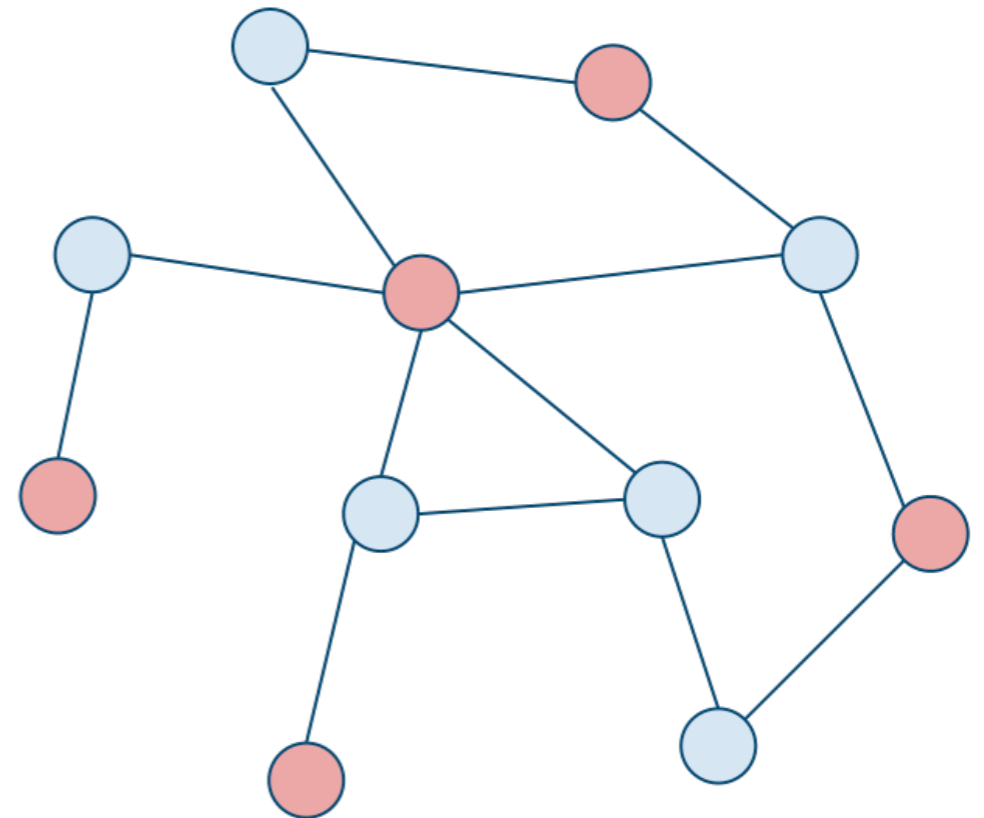
LubyGlauber

$O(\Delta \log n)$

parallel speedup
 $= \theta(n / \Delta)$

Δ = max-degree

χ = chromatic no.



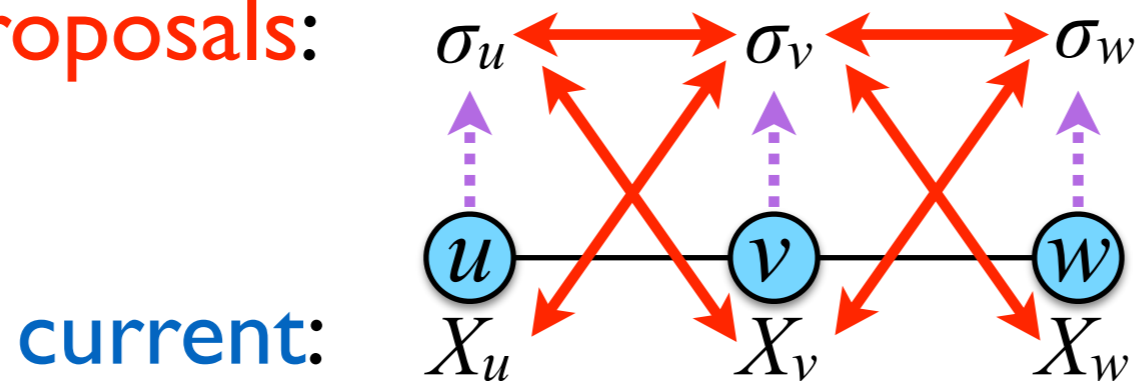
Do not update adjacent vertices simultaneously.

➡ It takes $\geq \chi$ steps to update all vertices at least once.

Q: “How to update all variables simultaneously and still converge to the correct distribution?”

The *LocalMetropolis* Chain

proposals:



current:

starting from an arbitrary $X \in [q]^V$, at each step:

each vertex $v \in V$ **independently proposes** a random $\sigma_v \in [q]$ with probability $b_v(\sigma_v) / \sum_{i \in [q]} b_v(i)$;

each edge $e=(u,v)$ **passes** its **check independently** with prob. $A_e(X_u, \sigma_v) A_e(\sigma_u, X_v) A_e(\sigma_u, \sigma_v) / \max_{i,j \in [q]} (A_e(i,j))^3$;

each vertex $v \in V$ **accepts** its proposal and update X_v to σ_v if **all incident edges pass their checks**;

a collective
coin flipping
made between
 u and v

- [Feng, Sun, Y. '17]: the *LocalMetropolis* chain is **time-reversible** w.r.t. the MRF Gibbs distribution μ .

Detailed Balance Equation:

$$\forall X, Y \in [q]^V, \quad \mu(X)P(X, Y) = \mu(Y)P(Y, X)$$

$\sigma \in [q]^V$: the proposals of all vertices

$\mathcal{C} \in \{0, 1\}^E$: indicates whether each edge $e \in E$ passes its check

$$\Omega_{X \rightarrow Y} \triangleq \{(\sigma, \mathcal{C}) \mid X \rightarrow Y \text{ when the random choice is } (\sigma, \mathcal{C})\}$$

$$\frac{P(X, Y)}{P(Y, X)} = \frac{\sum_{(\sigma, \mathcal{C}) \in \Omega_{X \rightarrow Y}} \Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, X)}{\sum_{(\sigma, \mathcal{C}) \in \Omega_{Y \rightarrow X}} \Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, Y)} = \frac{\mu(Y)}{\mu(X)}$$

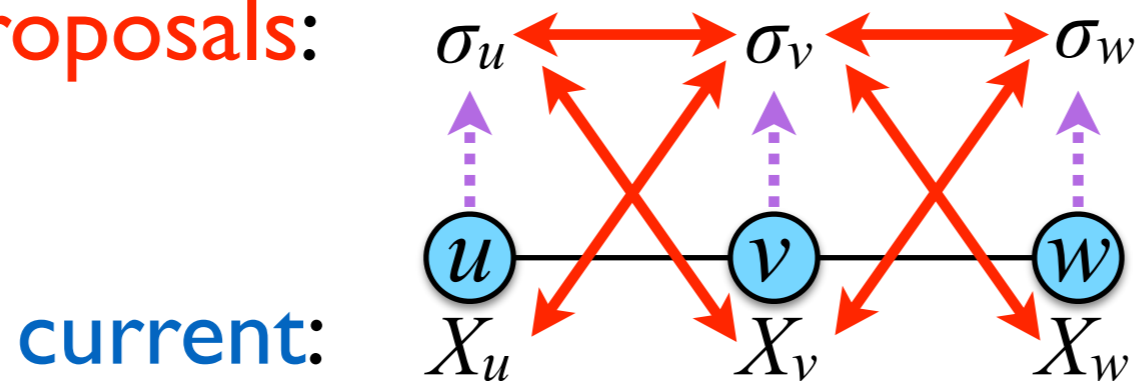
Bijection $\phi_{X, Y} : \Omega_{X \rightarrow Y} \rightarrow \Omega_{Y \rightarrow X}$ is constructed as:

$$(\sigma, \mathcal{C}) \xrightarrow{\phi_{X, Y}} (\sigma', \mathcal{C}') \quad \text{s.t.} \quad \begin{cases} \mathcal{C} = \mathcal{C}' \\ \text{if } \mathcal{C}_e = 1 \text{ for all } e \text{ incident with } v, \text{ then } \sigma'_v = X_v \\ \text{otherwise } \sigma'_v = \sigma_v \end{cases}$$

$$\Rightarrow \frac{\Pr(\sigma) \Pr(\mathcal{C} \mid \sigma, X)}{\Pr(\sigma') \Pr(\mathcal{C}' \mid \sigma', Y)} = \prod_{v \in V} \frac{b_v(Y_v)}{b_v(X_v)} \prod_{e=uv \in E} \frac{A_e(Y_u, Y_v)}{A_e(X_u, X_v)} = \frac{\mu(Y)}{\mu(X)}$$

The *LocalMetropolis* Chain

proposals:



current:

starting from an arbitrary $X \in [q]^V$, at each step:

each vertex $v \in V$ **independently proposes** a random $\sigma_v \in [q]$ with probability $b_v(\sigma_v) / \sum_{i \in [q]} b_v(i)$;

each edge $e=(u,v)$ **passes** its **check independently** with prob. $A_e(X_u, \sigma_v) A_e(\sigma_u, X_v) A_e(\sigma_u, \sigma_v) / \max_{i,j \in [q]} (A_e(i,j))^3$;

each vertex $v \in V$ **accepts** its proposal and update X_v to σ_v if **all incident edges pass their checks**;

a collective
coin flipping
made between
 u and v

- [Feng, Sun, Y. '17]: the *LocalMetropolis* chain is **time-reversible** w.r.t. the MRF Gibbs distribution μ .

LocalMetropolis for Hardcore model

the **hardcore model** on $G(V,E)$ with **fugacity** λ :

$$\forall \text{ independent set } I \text{ in } G: \quad \mu(I) = \frac{\lambda^{|I|}}{\sum_{I: \text{ IS in } G} \lambda^{|I|}}$$

starting from an arbitrary $X \in \{0,1\}^V$, with 1 indicating **occupied**
at each step, **each vertex** $v \in V$:

proposes a random $\sigma_v \in \{0,1\}$ **independently**

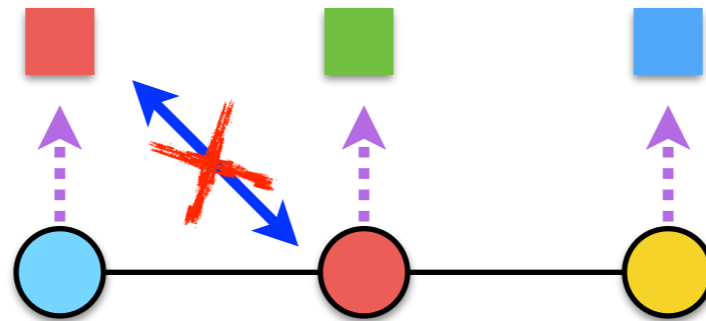
$$\sigma_v = \begin{cases} 1 & \text{with probability } \frac{\lambda}{1+\lambda}, \\ 0 & \text{with probability } \frac{1}{1+\lambda}; \end{cases}$$

accepts the proposal and update X_v to σ_v unless for some neighbor u of v :

$$X_u = \sigma_v = 1 \text{ or } \sigma_u = X_v = 1 \text{ or } \sigma_u = \sigma_v = 1;$$

- $\lambda < 1/\Delta$: $\tau_{\text{mix}} = O(\log n)$, even for unbounded Δ .

LocalMetropolis for q -Coloring



starting from an arbitrary $X \in [q]^V$, at each step, **each vertex** $v \in V$:

proposes a color $\sigma_v \in [q]$ **uniformly and independently** at random;
accepts the proposal and update X_v to σ_v if for all v 's neighbors u :

$$X_u \neq \sigma_v \wedge \sigma_u \neq X_v \wedge \sigma_u \neq \sigma_v;$$

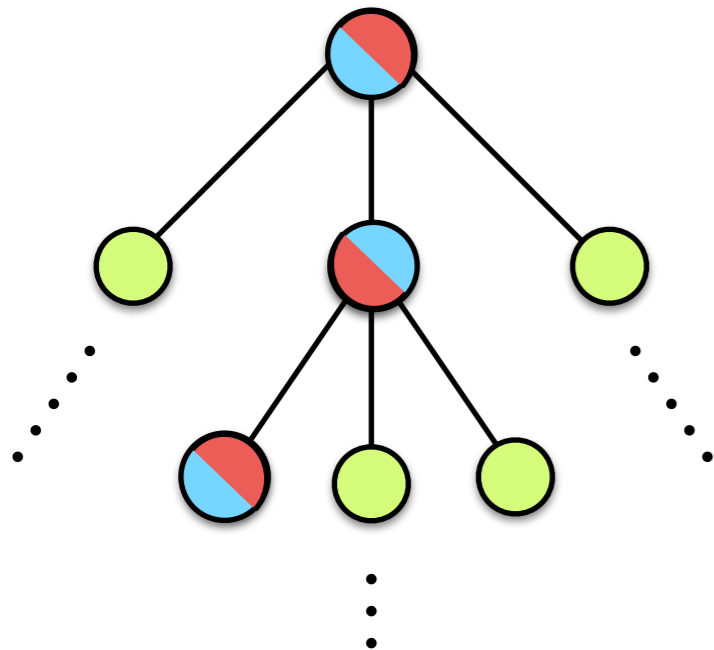
Theorem (Feng, Sun, Y. '17):

$$q \geq (2 + \sqrt{2} + \epsilon)\Delta \quad \Rightarrow \quad \tau_{\text{mix}} = O(\log n)$$

for LocalMetropolis on q -coloring

The $O(\log n)$ mixing time bound holds even for *unbounded* Δ and q .

Δ -regular tree



each v :

proposes a uniform random color $\sigma_v \in [q]$;

update X_v to σ_v if for all v 's neighbors u :

$$X_u \neq \sigma_v \wedge \sigma_u \neq X_v \wedge \sigma_u \neq \sigma_v;$$

$$X_{\text{root}} = \text{red}, \quad Y_{\text{root}} = \text{blue}$$

$$\forall \text{ non-root } v, \quad X_v = Y_v \notin \{\text{red}, \text{blue}\}$$

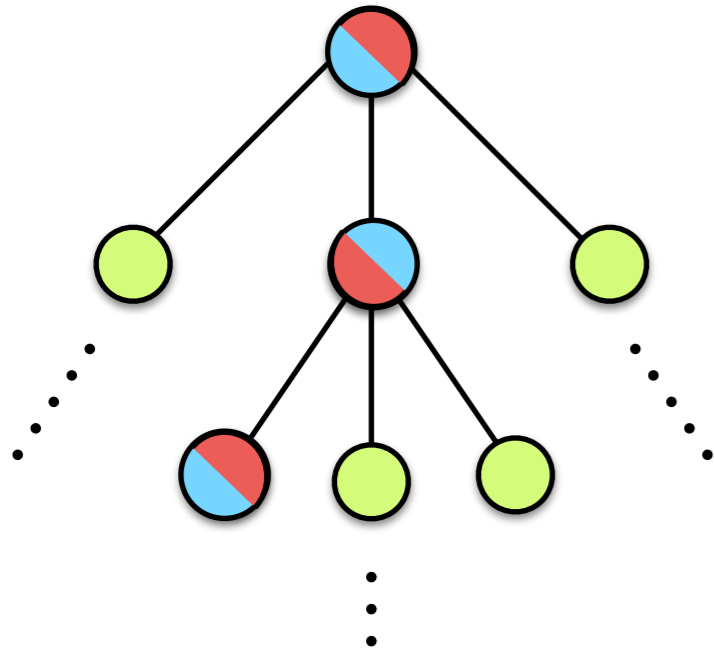
coupling: coupling the proposals (σ^X, σ^Y) so that $(X, Y) \xrightarrow{(\sigma^X, \sigma^Y)} (X', Y')$

vertex v proposes **consistently**: $\sigma_v^X = \sigma_v^Y$

vertex v proposes **bijectionally**:
$$\sigma_v^X = \begin{cases} \text{red} & \text{if } \sigma_v^Y = \text{blue} \\ \text{blue} & \text{if } \sigma_v^Y = \text{red} \\ \sigma_v^Y & \text{otherwise} \end{cases}$$

1. the root proposes **consistently**;
2. each child of the root proposes **bijectionally**;
3. each vertex of depth ≥ 2 proposes **bijectionally** if its parent proposed different colors in the two chains, and proposes **consistently** if otherwise;

Δ -regular tree



each v :

proposes a uniform random color $\sigma_v \in [q]$;

update X_v to σ_v if for all v 's neighbors u :

$$X_u \neq \sigma_v \wedge \sigma_u \neq X_v \wedge \sigma_u \neq \sigma_v;$$

$$X_{\text{root}} = \text{red}, \quad Y_{\text{root}} = \text{blue}$$

$$\forall \text{ non-root } v, \quad X_v = Y_v \notin \{\text{red}, \text{blue}\}$$

coupling: coupling the proposals (σ^X, σ^Y) so that $(X, Y) \xrightarrow{(\sigma^X, \sigma^Y)} (X', Y')$

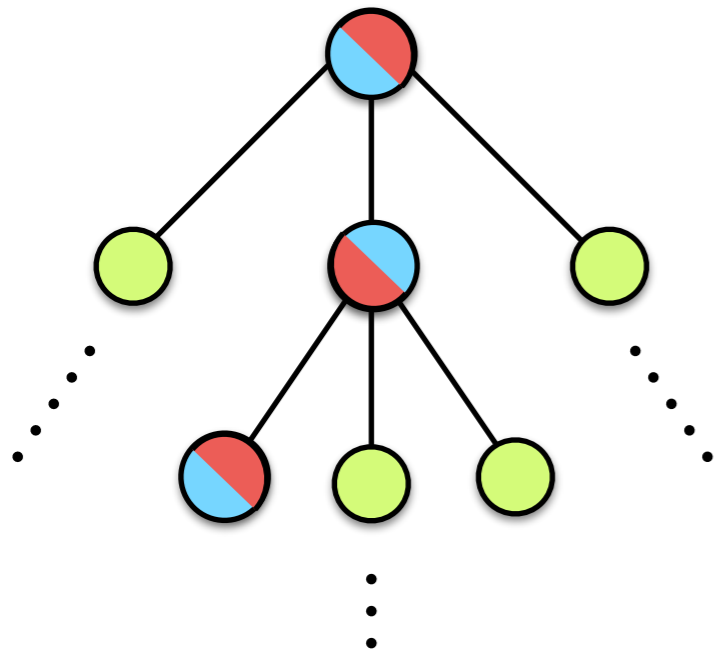
$$\text{root:} \quad \Pr[X'_{\text{root}} \neq Y'_{\text{root}}] \leq 1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^{\Delta}$$

$$\text{non-root } u \text{ at level } l: \quad \Pr[X'_u \neq Y'_u] \leq \frac{1}{q} \left(1 - \frac{2}{q}\right)^{\Delta-1} \left(\frac{2}{q}\right)^{\ell-1}$$

$$\Pr[X'_{\text{root}} \neq Y'_{\text{root}}] + \sum_{\text{non-root } u} \Pr[X'_u \neq Y'_u] \leq 1 - \left(1 - \frac{\Delta}{q}\right) \left(1 - \frac{2}{q}\right)^{\Delta} + \frac{\Delta}{q - 2\Delta} \left(1 - \frac{2}{q}\right)^{\Delta-1}$$

$$(\text{assume } q \geq \alpha\Delta) \quad \leq 1 - e^{-2/\alpha} \left(1 - \frac{1}{\alpha} - \frac{1}{\alpha - 2}\right)$$

Δ -regular tree



each v :

proposes a uniform random color $\sigma_v \in [q]$;

update X_v to σ_v if for all v 's neighbors u :

$$X_u \neq \sigma_v \wedge \sigma_u \neq X_v \wedge \sigma_u \neq \sigma_v;$$

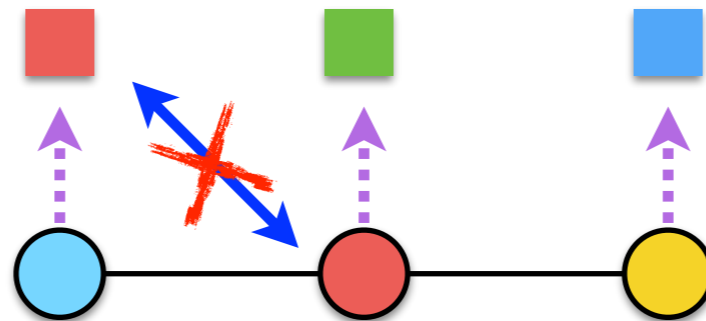
$$X_{\text{root}} = \text{red}, \quad Y_{\text{root}} = \text{blue}$$

$$\forall \text{ non-root } v, \quad X_v = Y_v \notin \{\text{red}, \text{blue}\}$$

for **general graph**:

1. deal with **irregularity** by a **metric** in path coupling;
2. deal with **cycles** by **self-avoiding walks**;
3. deal with **red/blue** non-root vertices by a **monotone argument**;

LocalMetropolis for q -Coloring



starting from an arbitrary $X \in [q]^V$, at each step, **each vertex** $v \in V$:

proposes a color $\sigma_v \in [q]$ **uniformly and independently** at random;
accepts the proposal and update X_v to σ_v if for all v 's neighbors u :

$$X_u \neq \sigma_v \wedge \sigma_u \neq X_v \wedge \sigma_u \neq \sigma_v;$$

$$q \geq (2 + \sqrt{2} + \epsilon) \Delta \quad \Rightarrow \quad \tau_{\text{mix}} = O(\log n)$$

- The mixing time holds even for **unbounded** Δ and q .
- **$q \geq (1 + \epsilon) \Delta$** : each vertex is updated at $\Omega(1)$ rate in *LocalMetropolis*

Lower Bounds

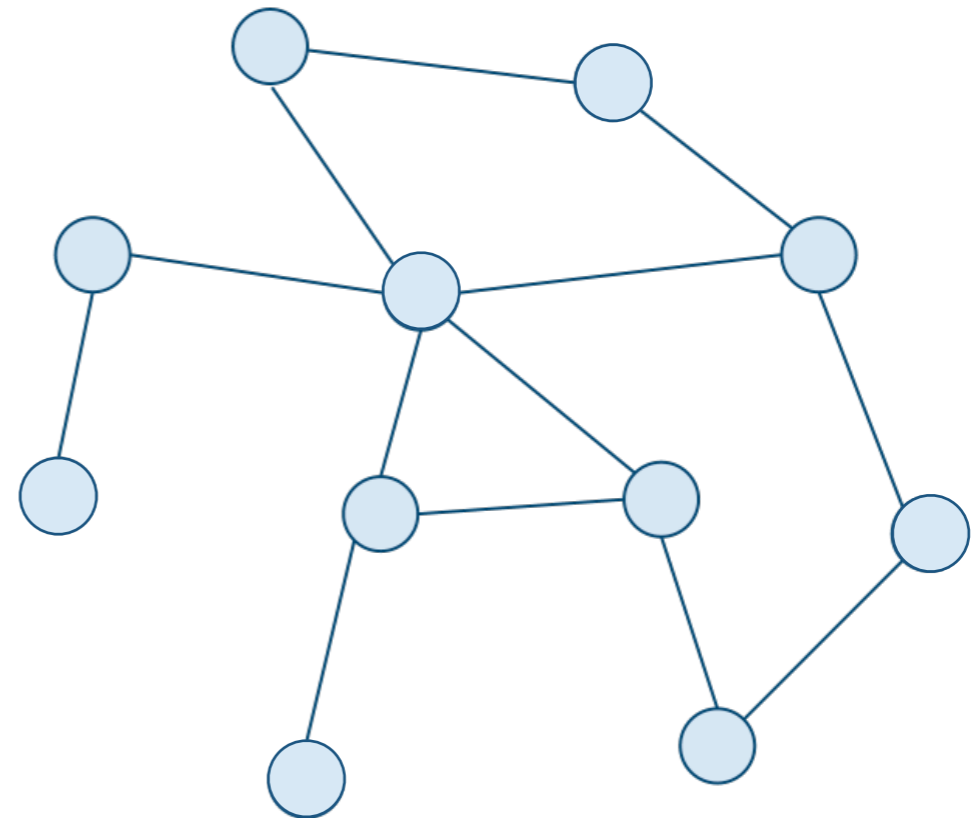
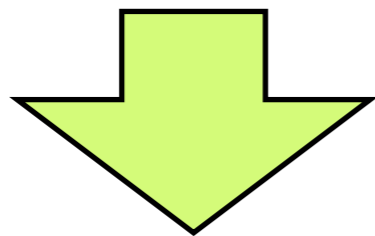
Q: “How local can a distributed sampling algorithm be?”

Q: “What *cannot* be sampled locally?”

The \mathcal{LOCAL} Model

the \mathcal{LOCAL} model:

- In t rounds: each node can collect information up to distance t .



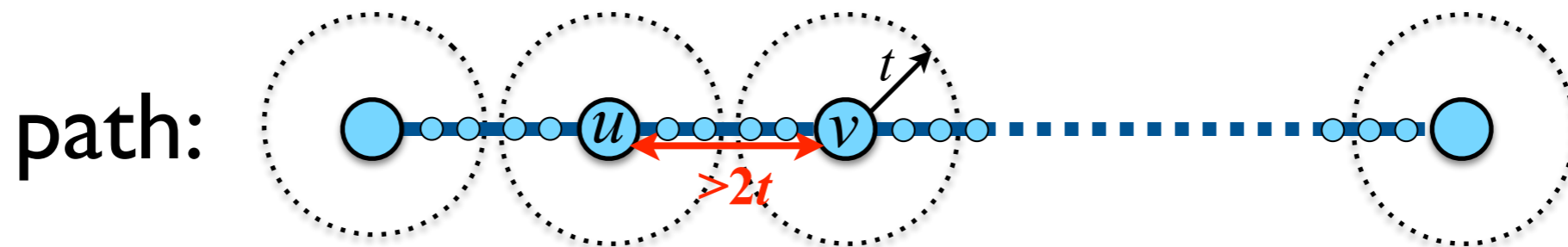
Outputs returned by vertices at distance $>2t$ from each other are **mutually independent**.

$\Omega(\log n)$ Lower Bound for Sampling

Theorem (Feng, Sun, Y. '17):

For any **non-degenerate** MRF, any distributed algorithm that samples from its distribution μ **within bounded total variation distance** requires $\Omega(\log n)$ rounds of communications.

outputs of t -round algorithm: **mutually independent** \tilde{X}_v 's



Gibbs distribution μ : **exponential correlation between** X_v 's

$$\sigma_u \neq \tau_u : \quad \|\mu_v^{\sigma_u} - \mu_v^{\tau_u}\|_{\text{TV}} \geq \exp(-O(t)) > n^{-1/4}$$

for a $t = O(\log n)$

$$d_{\text{TV}}(\mathbf{X}, \widetilde{\mathbf{X}}) > \frac{1}{2e} \quad \text{for **any** product distribution } \widetilde{\mathbf{X}}$$

$\Omega(\log n)$ Lower Bound for Sampling

Theorem (Feng, Sun, Y. '17):

For any **non-degenerate** MRF, any distributed algorithm that samples from its distribution μ **within bounded total variation distance** requires $\Omega(\log n)$ rounds of communications.

- The $\Omega(\log n)$ lower bound holds for all MRFs with **exponential correlation**:
 - non-trivial spin systems with $O(1)$ spin states.
- $O(\log n)$ is the new criteria of “**being local**” for distributed sampling algorithms.

An $\Omega(\textit{diam})$ Lower Bound

Theorem (Feng, Sun, Y. '17):

For any $\Delta \geq 6$, any distributed algorithm that samples **uniform** independent set **within bounded total variation distance** in graphs with max-degree Δ requires $\Omega(\textit{diam})$ rounds of communications.

Sampling almost uniform independent set in graphs with max-degree Δ by **poly-time Turing machines**:

- [Weitz'06] If $\Delta \leq 5$, there are poly-time algorithms.
- [Sly'10] If $\Delta \geq 6$, there is no poly-time algorithm unless **NP=RP**.

The $\Omega(\textit{diam})$ lower bound holds for sampling from the **hardcore model** with **fugacity** $\lambda > \lambda_c(\Delta) = \frac{(\Delta - 1)^{\Delta-1}}{(\Delta - 2)^\Delta}$

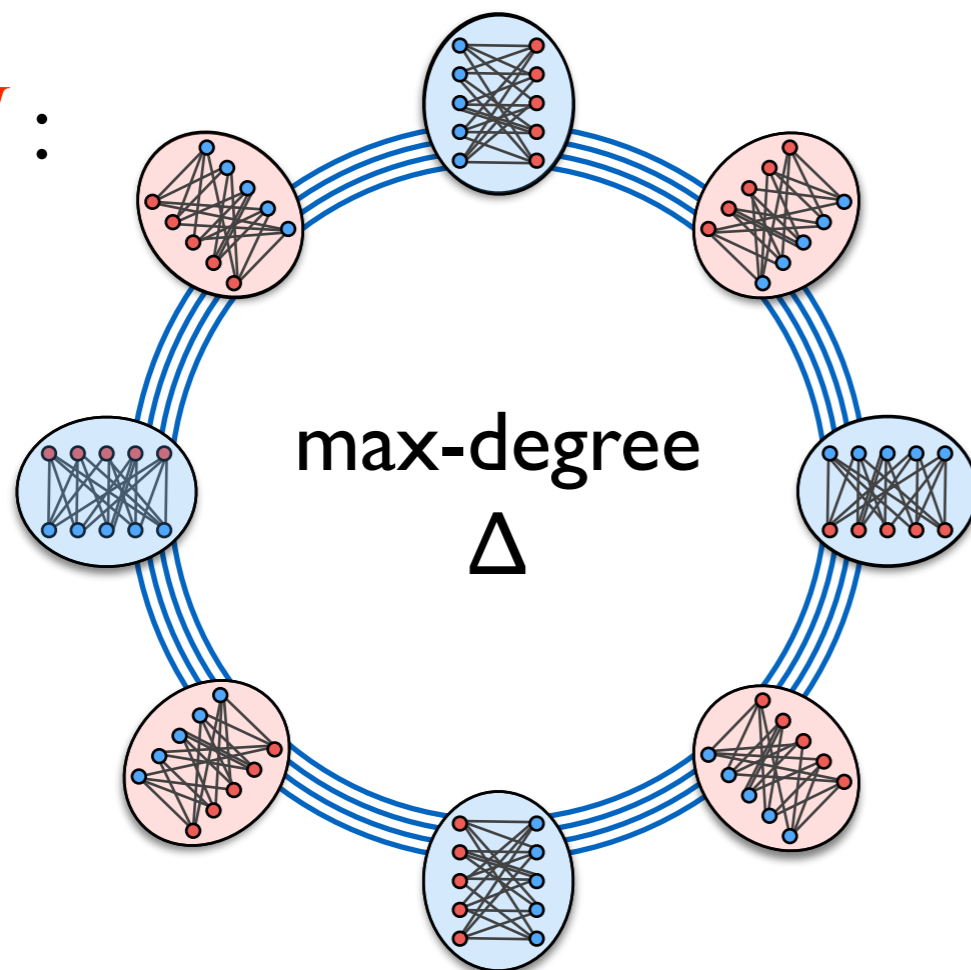
Theorem (Feng, Sun, Y. '17):

For any $\Delta \geq 6$, any distributed algorithm that samples **uniform independent set within bounded total variation distance** in graphs with max-degree Δ requires $\Omega(\text{diam})$ rounds of communications.

G : **even cycle**

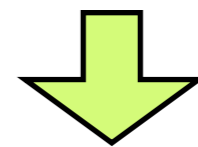
H : random **Δ -regular bipartite gadget**  of [Sly'10]

G^H :



if $\Delta \geq 6$:

sample nearly uniform
independent set in G^H



sample nearly uniform
max-cut in **even cycle** G
(**long-range correlation!**)

Theorem (Feng, Sun, Y. '17):

For any $\Delta \geq 6$, any distributed algorithm that samples **uniform independent set within bounded total variation distance** in graphs with max-degree Δ requires $\Omega(\textit{diam})$ rounds of communications.

A strong separation of **sampling** from other **local computation** tasks:

- Independent set is trivial to construct locally (because \emptyset is an independent set).
- The $\Omega(\textit{diam})$ lower bound for sampling holds even when every vertex knows the entire graph:
 - The lower bound holds not because of the **locality of input information**, but because of the **locality of randomness**.

Summary

- Sampling from locally-defined joint distribution via distributed algorithms:
 - *LubyGlauber*: $O(\Delta \log n)$ rounds under **Dobrushin condition**;
 - *LocalMetropolis*: may achieve $O(\log n)$ rounds;
 - $\Omega(\log n)$ lower bound for sampling from **almost all nontrivial** joint distributions;
 - $\Omega(\text{diam})$ lower bound for sampling from joint distributions exhibiting (**non-uniqueness**) **phase transition property**.

Weiming Feng, Yuxin Sun, Yitong Yin. *What can be sampled locally?*
In **PODC**'17. arxiv: 1702.00142.

Open Problems

- Better analysis of *LocalMetropolis*.
- Distributed sampling of:
 - matchings;
 - ferromagnetic Ising model on graphs of unbounded degree;
 - anti-ferromagnetic 2-spin systems in the uniqueness regime on graphs of unbounded degree;
- Self-reducible sampling in the LOCAL model?
- Complexity hierarchy for distributed sampling?
- New ideas for distributed sampling: e.g. the *LLL* sampler for hardcore model of Guo-Jerrum-Liu.

Thank you!

Any questions?