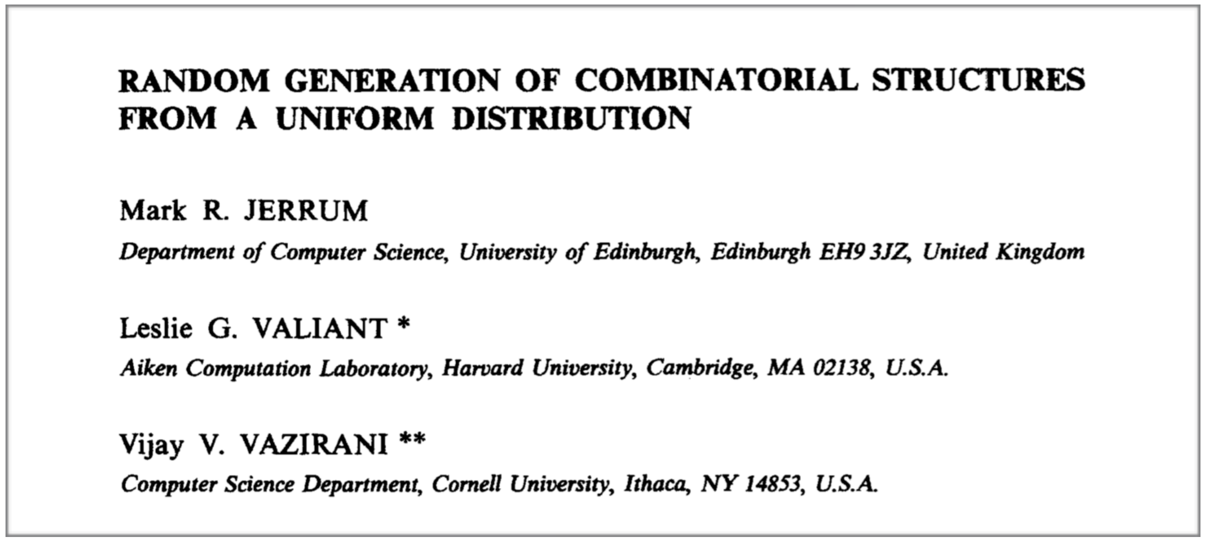
Local Distributed Sampling
from
Locally-Defined Distribution

Yitong Yin
Nanjing University

Counting and Sampling

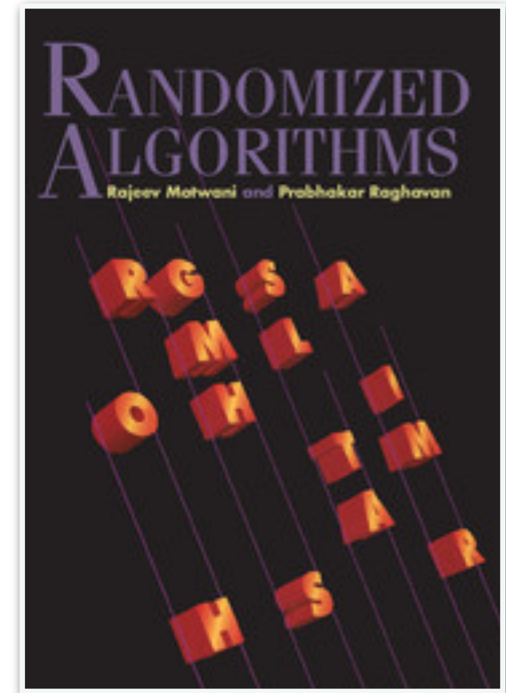


RANDOM GENERATION OF COMBINATORIAL STRUCTURES FROM A UNIFORM DISTRIBUTION

Mark R. JERRUM
Department of Computer Science, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom

Leslie G. VALIANT *
Aiken Computation Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.

Vijay V. VAZIRANI **
Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A.



[Jerrum-Valiant-Vazirani '86]:

(For *self-reducible* problems)

approx. counting is tractable ⟺ **(approx., exact) sampling** is tractable

Computational Phase Transition

Sampling almost-uniform independent set in graphs with maximum degree Δ:

- [Weitz 2006]: If $\Delta \leq 5$, poly-time.
- [Sly 2010]: If $\Delta \geq 6$, no poly-time algorithm unless **NP**=**RP**.

A phase transition occurs when Δ: $5 \to 6$.

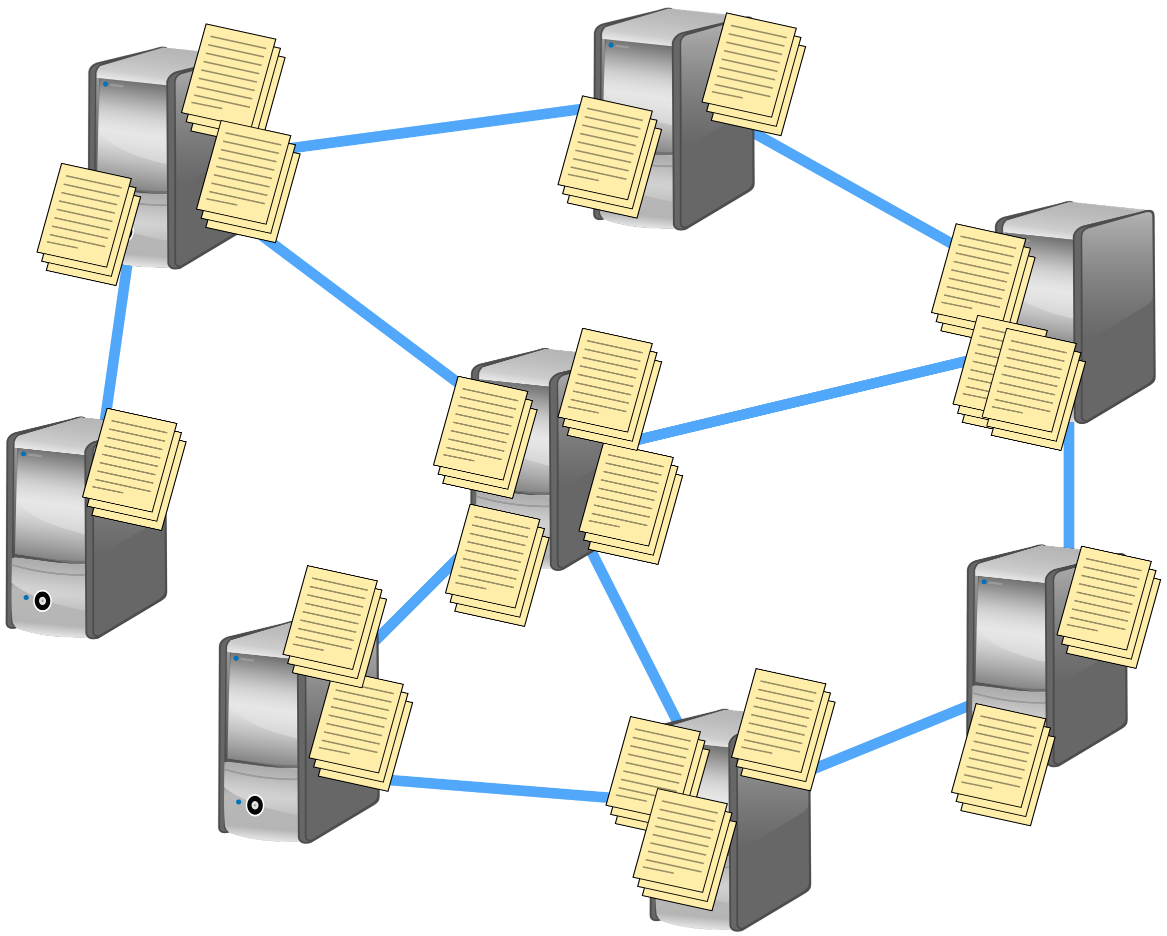
Local Computation?

Local Computation

"*What can be computed locally?*" [Naor, Stockmeyer '93]

the *LOCAL* model [Linial '87]:

- Communications are synchronized.
- In each round: each node can exchange unbounded messages with all neighbors, perform unbounded local computation, and read/write to unbounded local memory.
- Complexity: # of rounds to terminate in the worst case.

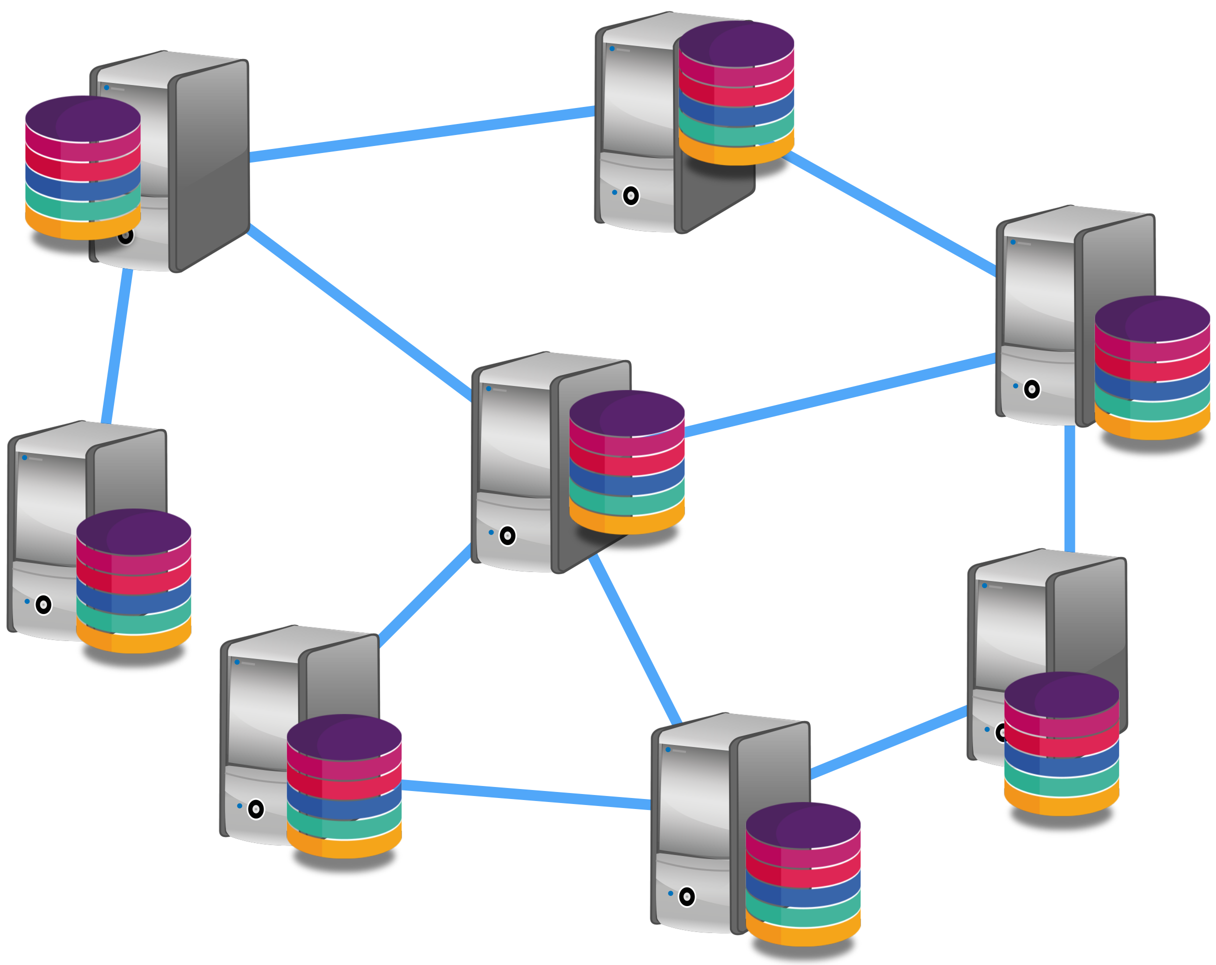


- In t rounds: each node can collect information up to distance t.

PLOCAL: $t = \text{polylog}(n)$

A Motivation: *Distributed Machine Learning*

- Data are stored in a distributed system.
- Distributed algorithms for:
  - sampling from a *joint distribution* (specified by a *probabilistic graphical model*);
  - inferring according to a *probabilistic graphical model*.

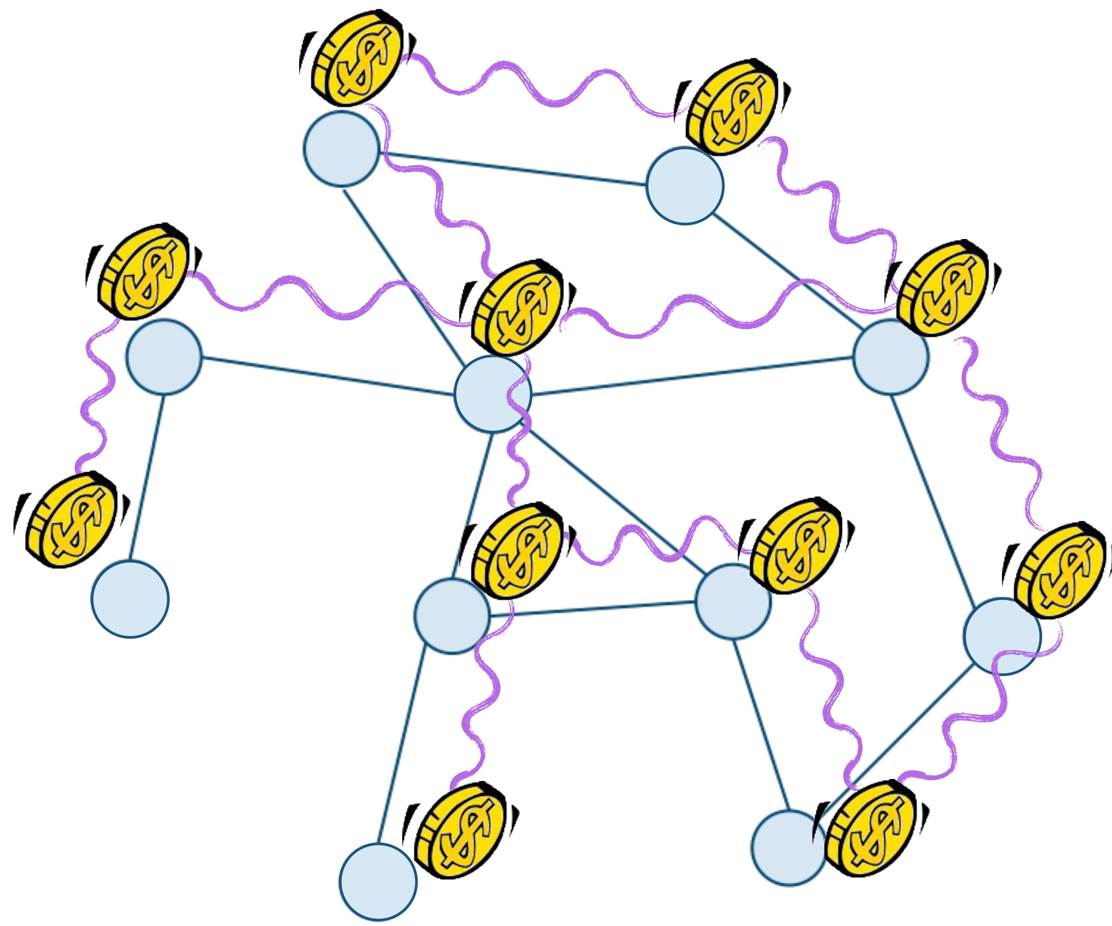


Example: Sample Independent Set

μ: uniform distribution of independent sets in G.

$\boldsymbol{Y} \in \{0,1\}^V$ indicates an independent set

- Each $v \in V$ returns a $Y_v \in \{0,1\}$, such that $\boldsymbol{Y} = (Y_v)_{v \in V} \sim \mu$
- Or: $d_{\mathrm{TV}}(\boldsymbol{Y}, \mu) < 1/\mathrm{poly}(n)$



network $G(V,E)$

Inference (Local Counting)

μ: uniform distribution of independent sets in G.

μ_v^σ: marginal distribution at v conditioning on $\sigma \in \{0,1\}^S$.

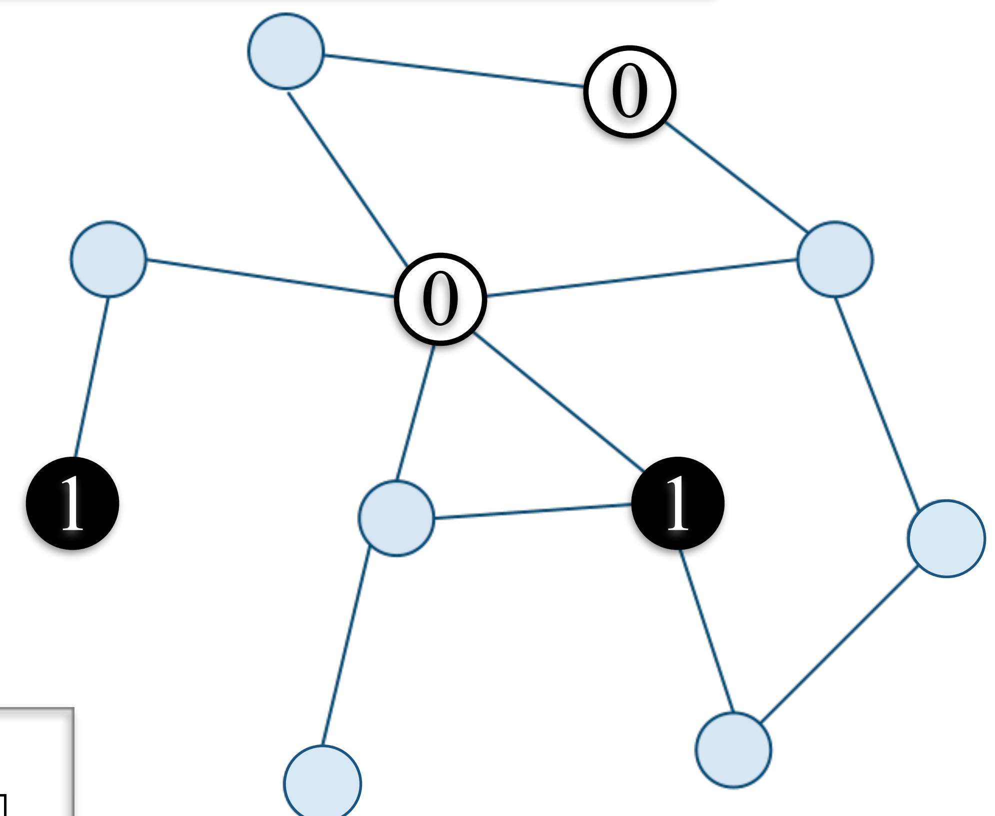
$$\forall y \in \{0,1\}: \quad \mu_v^\sigma(y) = \Pr_{\boldsymbol{Y}\sim\mu}[Y_v = y \mid Y_S = \sigma]$$

- Each $v \in S$ receives σ_v as input.
- Each $v \in V$ returns a marginal distribution $\hat{\mu}_v^\sigma$ such that:

$$d_{\mathrm{TV}}(\hat{\mu}_v^\sigma, \mu_v^\sigma) \le \frac{1}{\mathrm{poly}(n)}$$

$$\frac{1}{Z} = \mu(\emptyset) = \prod_{i=1}^{n} \Pr_{\boldsymbol{Y}\sim\mu}[Y_{v_i} = 0 \mid \forall j < i : Y_{v_j} = 0]$$

Z: # of independent sets



network $G(V,E)$

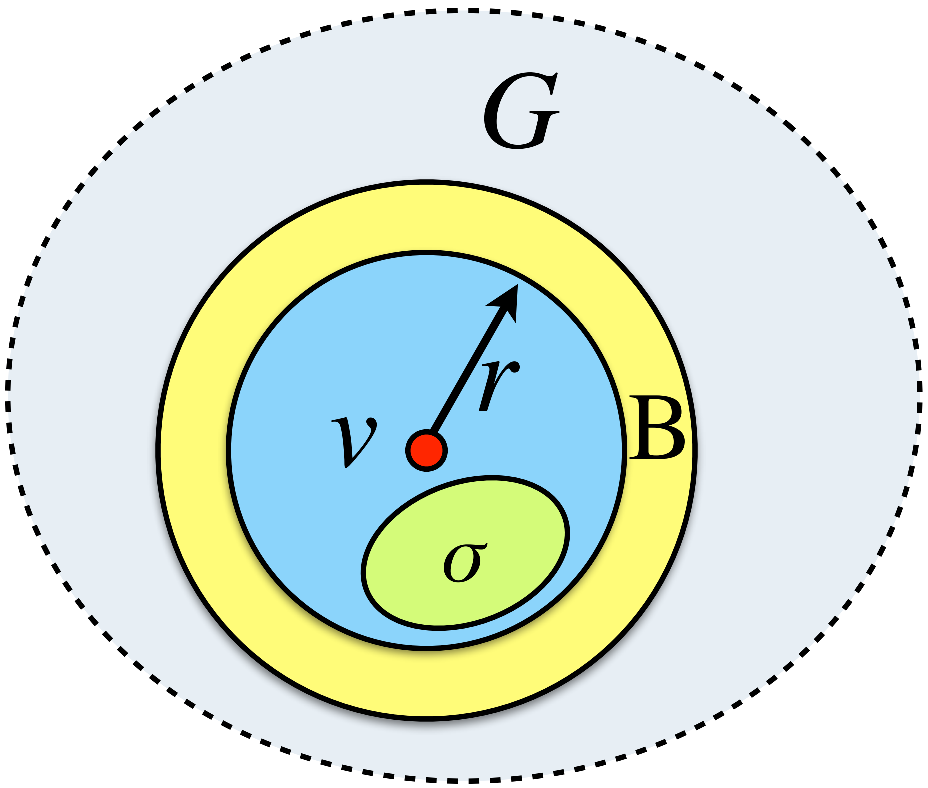
Decay of Correlation

μ_v^σ: marginal distribution at v conditioning on $\sigma \in \{0,1\}^S$.

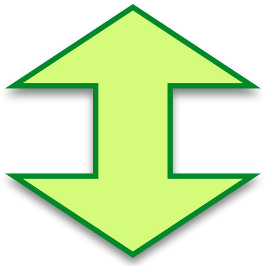
strong spatial mixing (SSM):

$\forall$ boundary condition $B \in \{0,1\}^{r\text{-sphere}(v)}$:

$$d_{\mathrm{TV}}(\mu_v^\sigma, \mu_v^{\sigma,B}) \leq \mathrm{poly}(n) \cdot \exp(-\Omega(r))$$



SSM (iff $\Delta \leq 5$ when μ is uniform distribution of ind. sets)



approx. inference is solvable in $O(\log n)$ rounds in the $\mathcal{LOCAL}$ model

Gibbs Distribution

(with pairwise interactions)

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

$$A_e: [q] \times [q] \to [0,1]$$

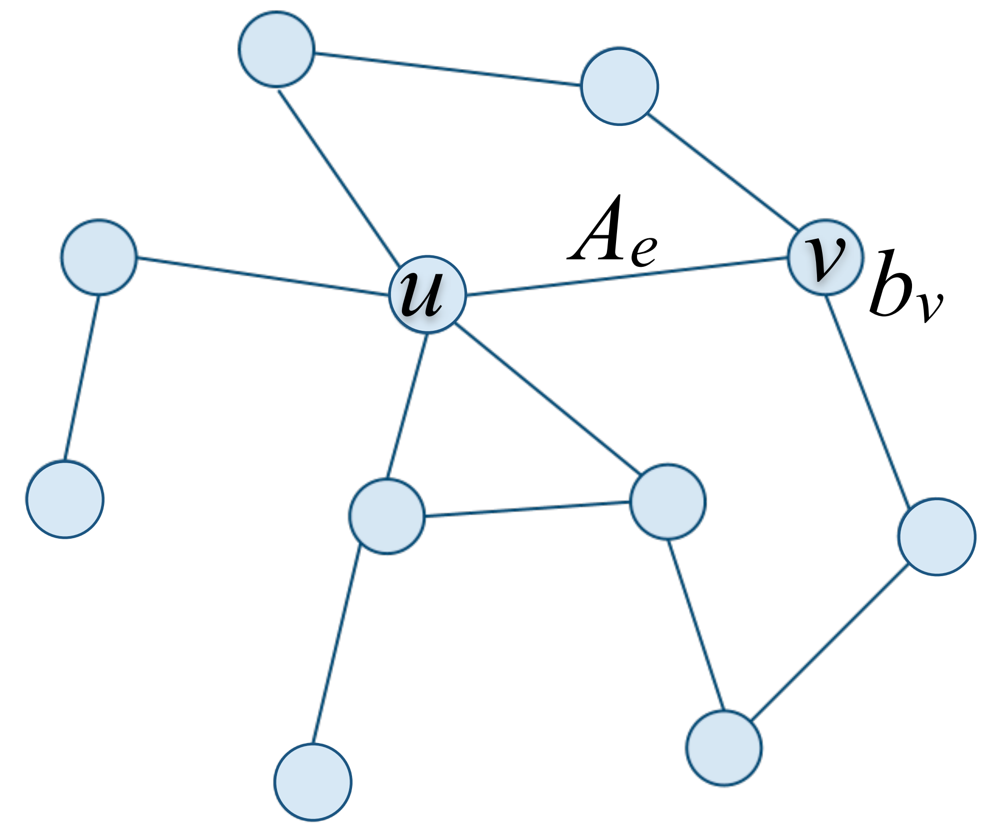
- Each vertex $v\in V$ has a vector (unary constraint):

$$b_v: [q] \to [0,1]$$

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

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

network $G(V,E)$:



Gibbs Distribution

(with pairwise interactions)

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

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

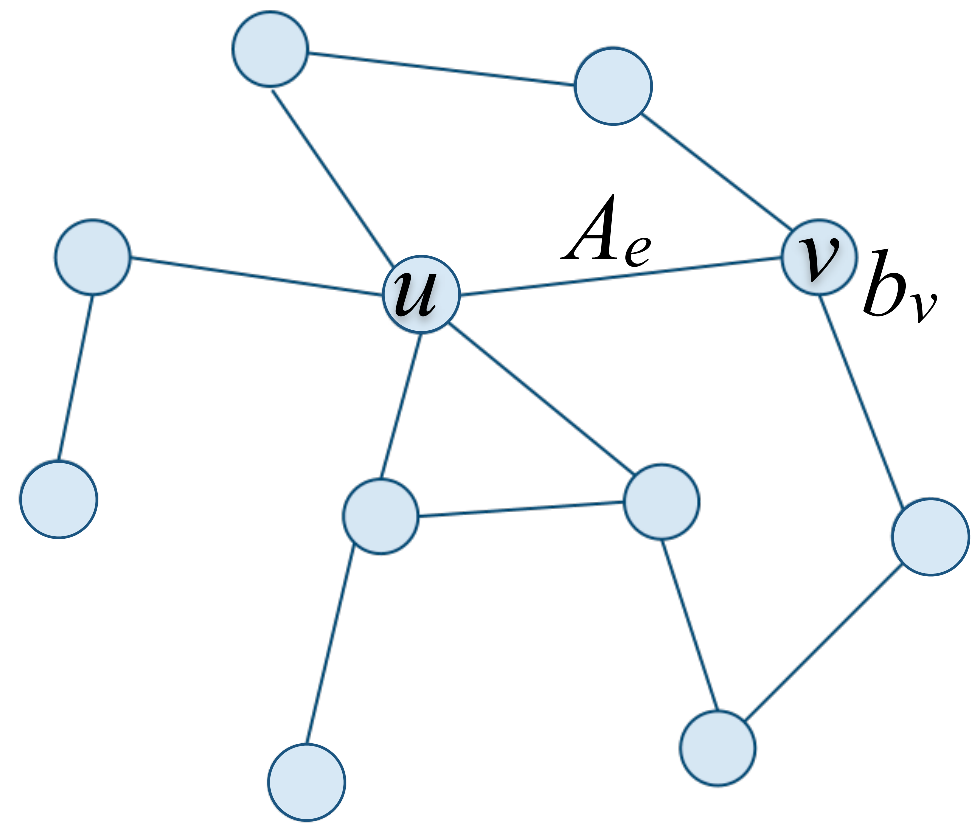
- independent set:

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

- coloring:

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

network $G(V,E)$:



A_e: $[q] \times [q] \to [0,1]$

b_v: $[q] \to [0,1]$

Gibbs Distribution

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

$$\mu(\sigma) \propto \prod_{(f,S)\in\mathcal{F}} f(\sigma_S)$$

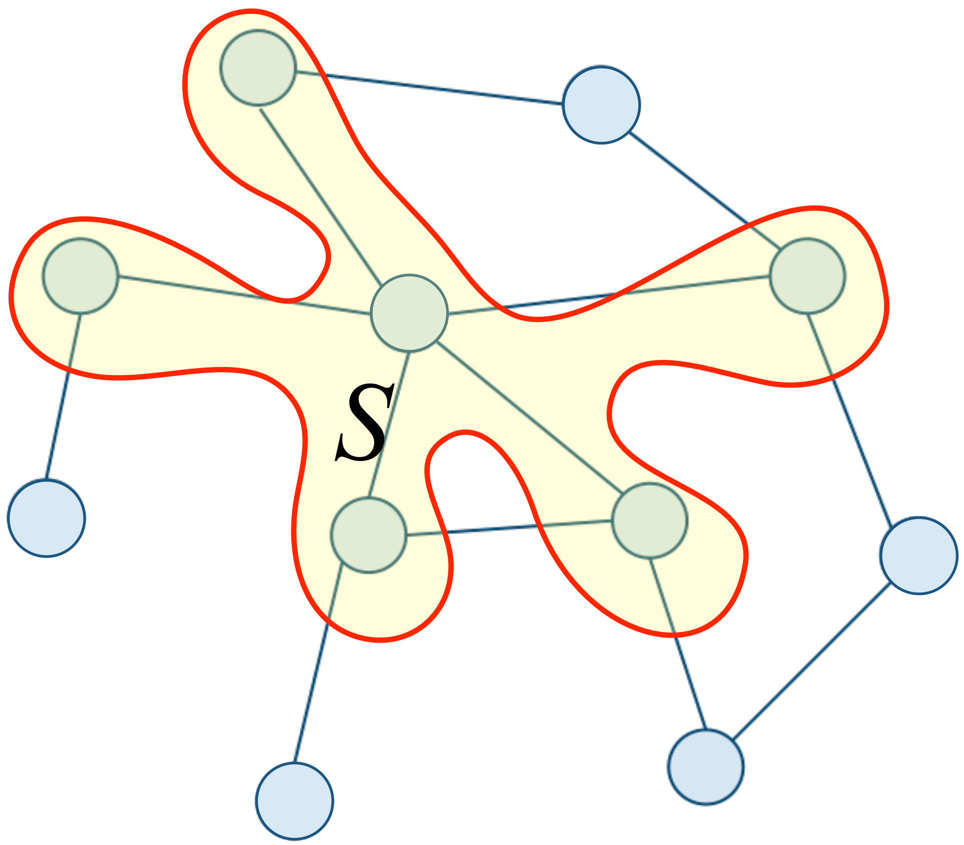
each $(f, S) \in \mathcal{F}$

is a *local* constraints (factors):

$$f : [q]^S \to \mathbb{R}_{\geq 0}$$

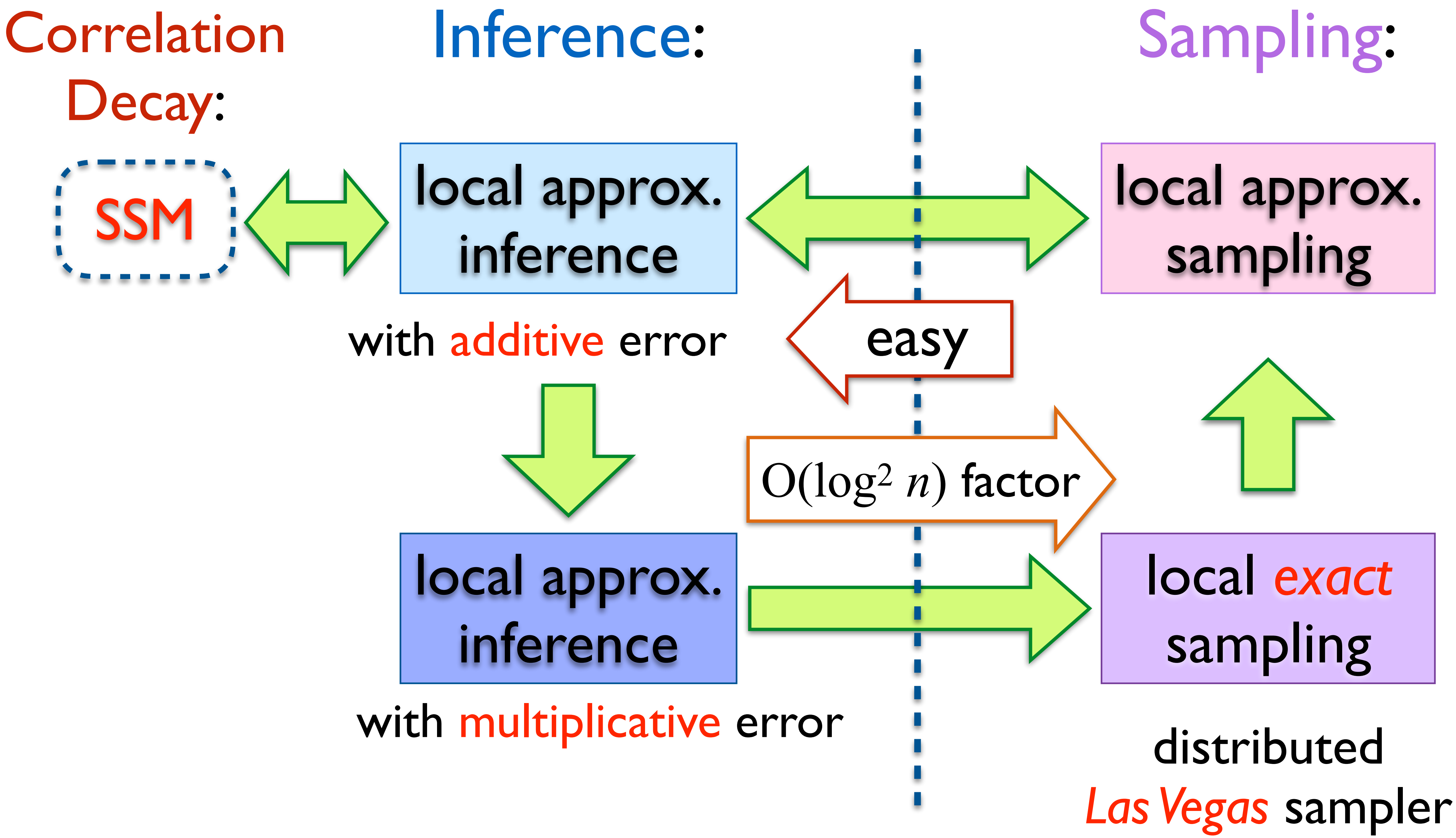
$S \subseteq V$ with $diam_G(S) = O(1)$

network $G(V,E)$:



Locality of Counting & Sampling

For Gibbs distributions (defined by *local* factors):



Locality of Sampling

Correlation Decay: SSM ⟷ **Inference:** local approx. inference ⟶ **Sampling:** local approx. sampling

Inference:

each v can compute a $\hat{\mu}_v^\sigma$ within $O(\log n)$-ball

s.t. $d_{\mathrm{TV}}\left(\hat{\mu}_v^\sigma, \mu_v^\sigma\right) \le \frac{1}{\mathrm{poly}(n)}$

Sampling:

return a random $\boldsymbol{Y} = (Y_v)_{v\in V}$ whose distribution $\hat{\mu} \approx \mu$

$d_{\mathrm{TV}}\left(\hat{\mu}, \mu\right) \le \frac{1}{\mathrm{poly}(n)}$

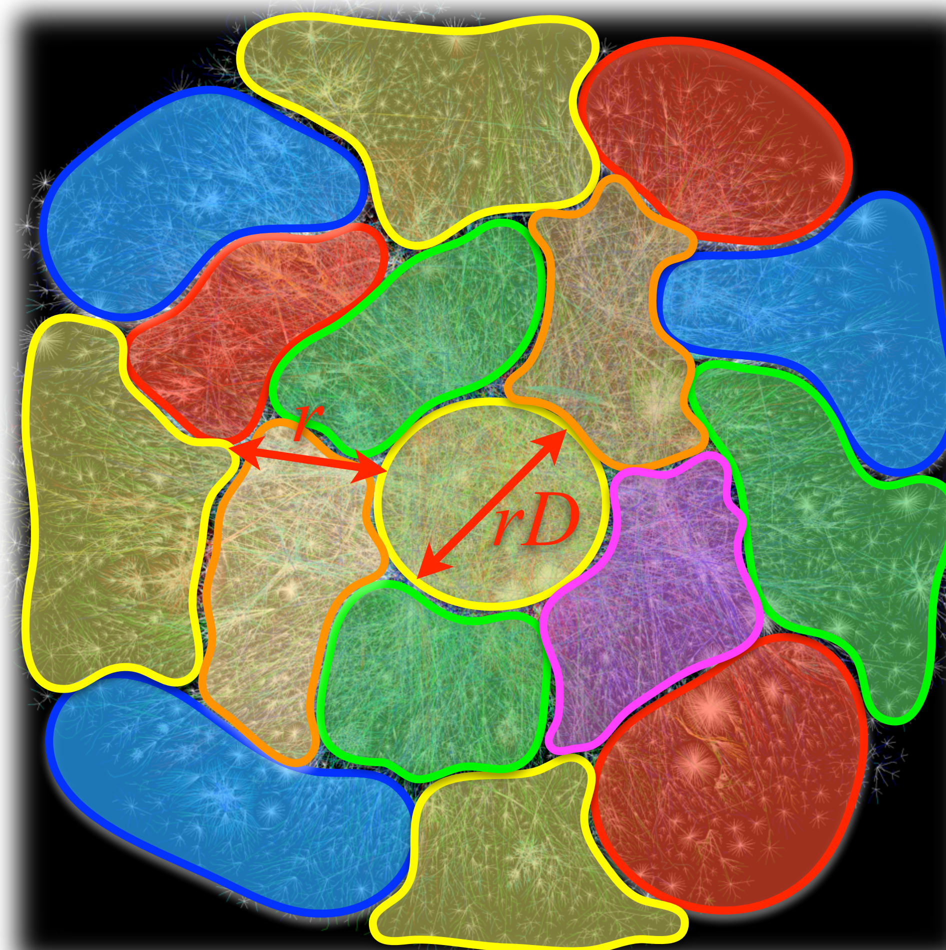
sequential $O(\log n)$-local procedure:

- scan vertices in V in an arbitrary order $v_1, v_2, \ldots, v_n$
- for i=1,2, …, n: sample Y_{v_i} according to $\hat{\mu}_{v_i}^{Y_{v_1},\ldots,Y_{v_{i-1}}}$

Network Decomposition

(C,D) -network-decomposition of G:

- classifies vertices into clusters;
- assign each cluster a color in $[C]$;
- each cluster has diameter $\leq D$;
- clusters are properly colored.



$(C,D)^r$-ND: (C,D)-ND of G^r

Given a $(C,D)^r$- ND:

sequential r-local procedure: $r = O(\log n)$

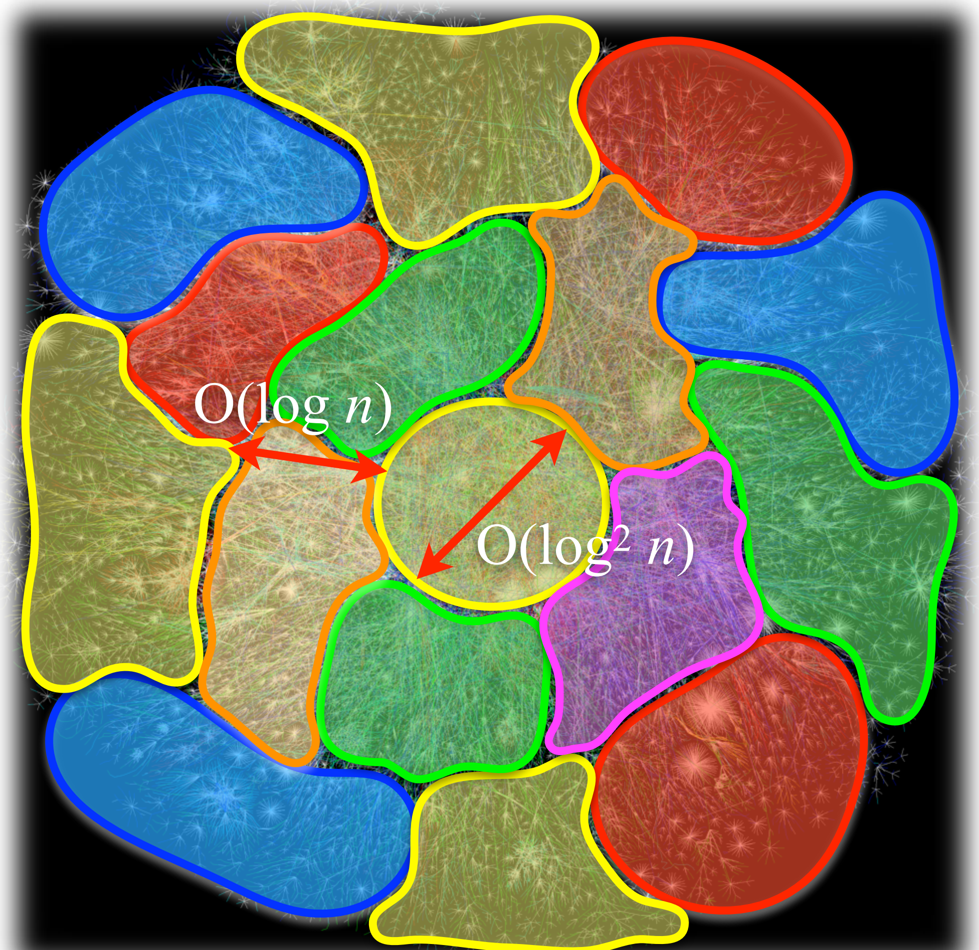
- scan vertices in V in an arbitrary order $v_1, v_2, \ldots, v_n$
- for i=1,2, …, n: sample Y_{v_i} according to $\hat{\mu}_{v_i}^{Y_{v_1},\ldots,Y_{v_{i-1}}}$

can be simulated in $O(CDr)$ rounds in $\mathcal{LOCAL}$ model

Network Decomposition

(C,D) -network-decomposition of G:

- classifies vertices into clusters;
- assign each cluster a color in $[C]$;
- each cluster has diameter $\leq D$;
- clusters are properly colored.



$(C,D)^r$-ND: (C,D)-ND of G^r

$(O(\log n), O(\log n))^r$-ND can be constructed in $O(r \log^2 n)$ rounds *w.h.p.*

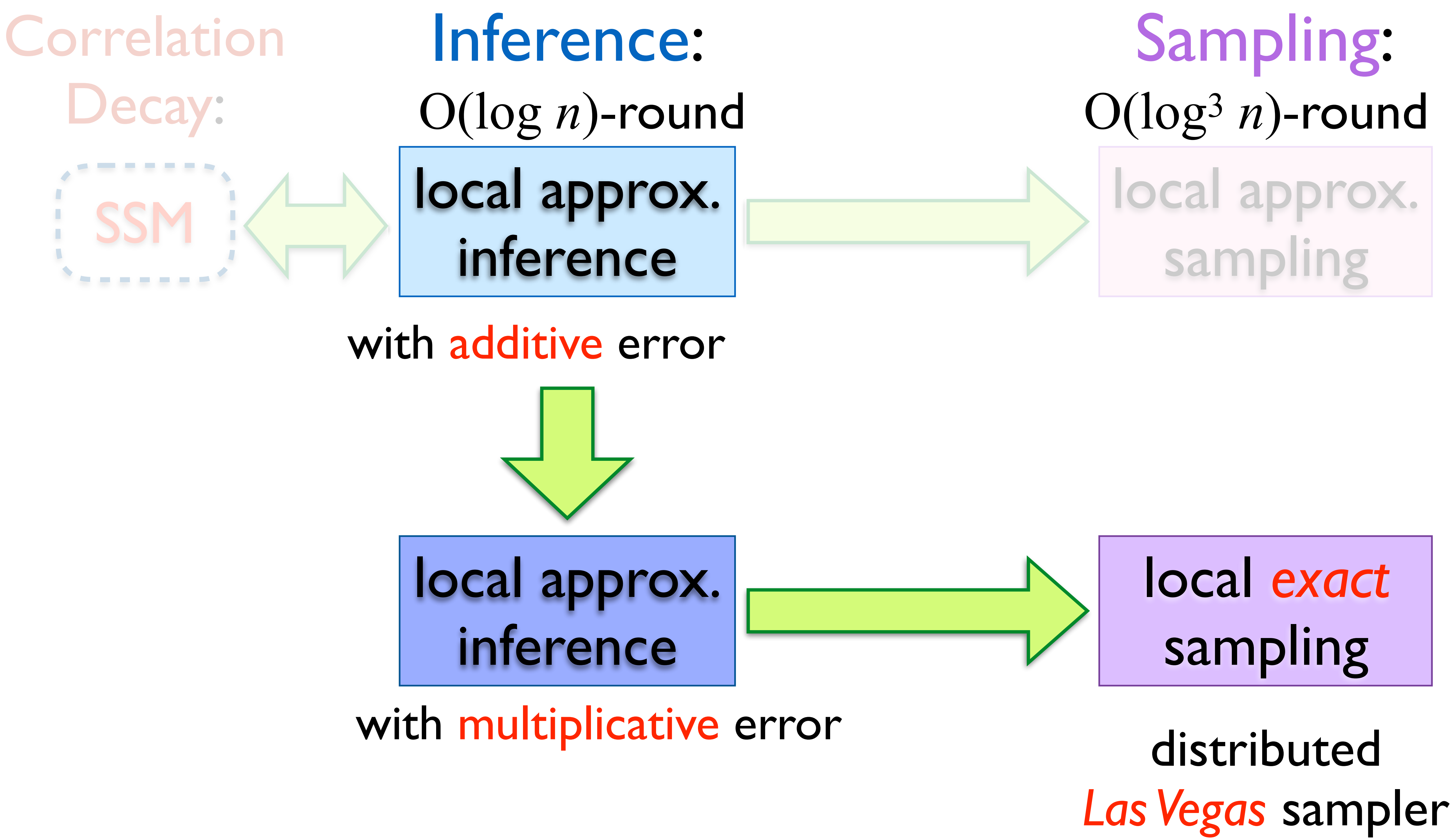
[Linial, Saks, 1993] — [Ghaffari, Kuhn, Maus, 2017]:

r-local SLOCAL algorithm:
$\forall$ ordering $\pi=(v_1, v_2, \ldots, v_n)$,
returns random vector $Y^{(\pi)}$

ND ⇒

$O(r\log^2 n)$-round LOCAL alg.:
returns *w.h.p.* the $Y^{(\pi)}$
for some ordering π

Locality of Sampling



An *LLL*-like Framework

independent random variables: $X_1, \ldots, X_n$ with domain Ω

$\mathcal{A}$: a set of *bad events*

each $A \in \mathcal{A}$ is associated with
- variable set $\mathsf{vbl}(A) \subseteq [n]$
- function $q_A : \Omega^{\mathsf{vbl}(A)} \to \{0, 1\}$

variable-framework Lovász local lemma

Rejection sampling: (with conditionally mutually independent filters)

- $X_1, \ldots, X_n$ are drawn independently;
- each $A \in \mathcal{A}$ occurs independently with prob. $1 - q_A\left(X_{\mathsf{vbl}(A)}\right)$;
- the sample is accepted if none of $A \in \mathcal{A}$ occurs.

Target distribution D^*: $X_1, \ldots, X_n$ conditioned on accepted

Partial rejection sampling [Guo-Jerrum-Liu'17]: resample not all variables

Resample variables local to the errors? (Moser-Tardos)

Local Rejection Sampling

- draw independent samples of $\boldsymbol{X} = (X_1, \ldots, X_n)$;
- each $A \in \mathcal{A}$ occurs (*violated*) ind. with $\Pr[A]=1-q_A(X_{\mathsf{vbl}(A)})$;
- while there is a violated bad event $A \in \mathcal{A}$: $\boldsymbol{X}^{\mathsf{old}} \leftarrow \text{current } \boldsymbol{X}$
  - resample all variables in vbl(A) for violated A;
  - for violated A: violate A with $\Pr[A] = 1-q_A(X_{\mathsf{vbl}(A)})$;
  - for non-violated A that shares variables with violated event: violate A with $\Pr[A] = 1 - q_A^* \cdot q_A\left(X_{\mathsf{vbl}(A)}\right) / q_A\left(X^{\mathtt{old}}_{\mathsf{vbl}(A)}\right)$

where $q_A{}^*$ is a worst-case lower bound for $q_A(\ )$:

$$\forall X_{\mathsf{vbl}(A)} : \ q_A\left(X_{\mathsf{vbl}(A)}\right) \geq q_A^*$$

soft filters: $\forall A \in \mathcal{A},\ q_A^* > 0$ ⇒ $(X_1, \ldots, X_n) \sim D^*$ (target distribution) upon termination

Only the variables **local** to the violated events are resampled.

(work even for **dynamic** filters)

By a resampling table argument.

Local *Ising* Sampler

$0 < \beta < 1$

$\lambda > 0$

ferro: $A = \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}$ anti-ferro: $A = \begin{bmatrix} \beta & 1 \\ 1 & \beta \end{bmatrix}$ external field $b = \begin{bmatrix} \lambda \\ 1 \end{bmatrix}$

- each vertex $v \in V$ ind. samples a spin state $\sigma_v \in \{0,1\} \propto \boldsymbol{b}$;
- each edge $e=(u,v) \in E$ fails ind. with prob. $1\text{-}A(\sigma_u,\sigma_v)$;
- while there is a failed edge: $\boldsymbol{\sigma}^{\text{old}} \leftarrow$ current $\boldsymbol{\sigma}$
  - resample σ_v for all vertices v involved in failed edges;
  - each failed $e=(u,v)$ is revived ind. with prob. $A(\sigma_u,\sigma_v)$;
  - each non-failed $e=(u,v)$ that is incident to a failed edge, fails ind. with prob. $1 - \beta\cdot A(\sigma_u,\sigma_v)/A(\sigma_u^{\text{old}},\sigma_v^{\text{old}})$;

Pros:

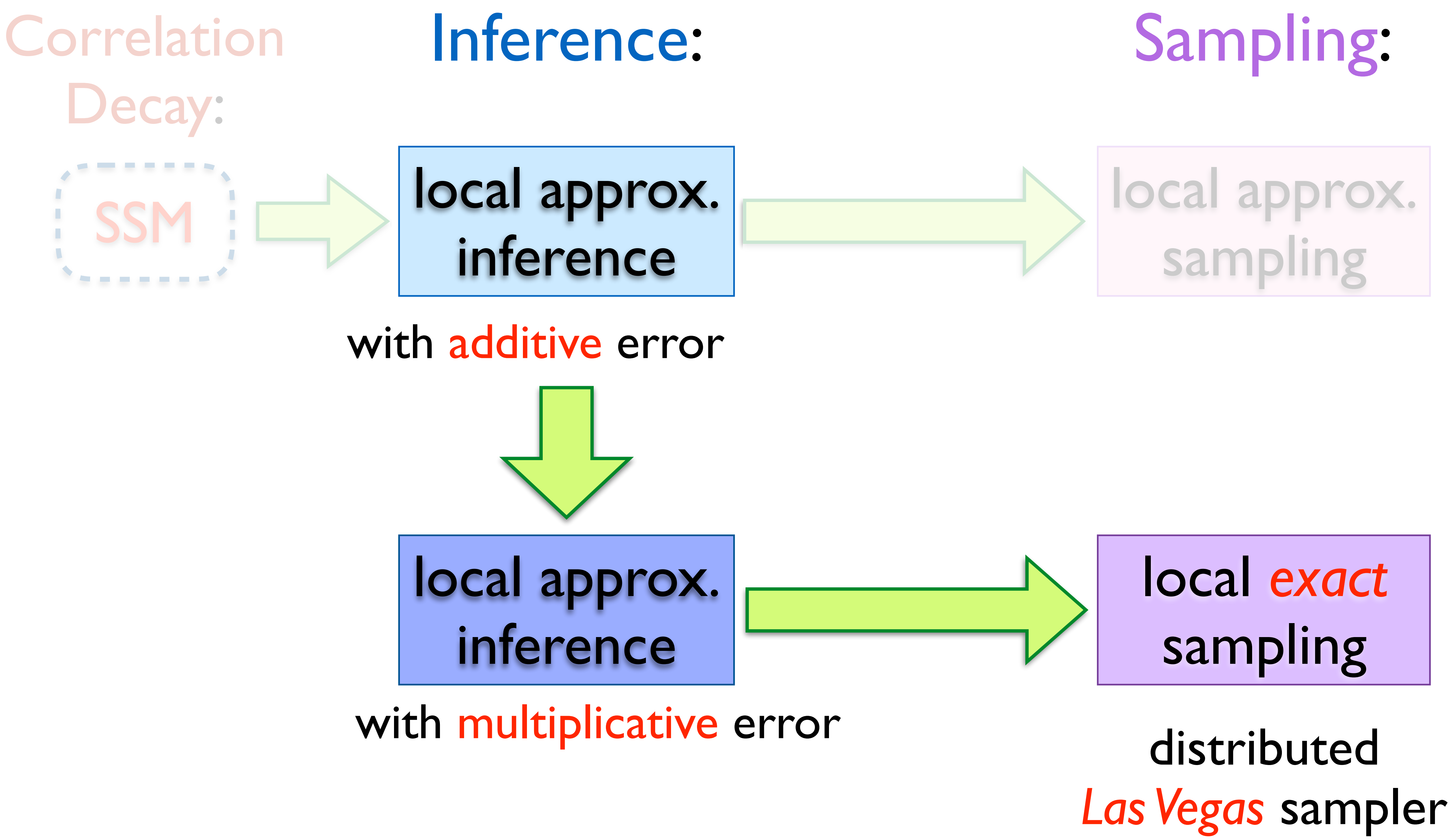
- local & parallel
- dynamic graph
- exact sampler
- certifiable termination

Cons:

- convergence is hard to analyze
- regime is not tight $\beta > 1 - \Theta(\frac{1}{\Delta})$
- soft constraints

Locality of Sampling

For Gibbs distributions (distributions defined by *local* factors):



Jerrum-Valiant-Vazirani Sampler

[**J**errum-**V**aliant-**V**azirani '86]

$\exists$ an efficient algorithm that samples from $\hat{\mu}$

and evaluates $\hat{\mu}(\sigma)$ given any $\sigma \in \{0,1\}^V$

multiplicative error: $\forall \sigma \in \{0,1\}^V$: $\quad e^{-1/n^2} \le \frac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le e^{1/n^2}$

<u>Self-reduction:</u>

$$\mu(\sigma) = \prod_{i=1}^{n} \mu_{v_i}^{\sigma_1,\ldots,\sigma_{i-1}}(\sigma_i) = \prod_{i=1}^{n} \frac{Z(\sigma_1,\ldots,\sigma_i)}{Z(\sigma_1,\ldots,\sigma_{i-1})}$$

let $\hat{\mu}_{v_i}^{\sigma_1,\ldots,\sigma_{i-1}}(\sigma_i) = \frac{\hat{Z}(\sigma_1,\ldots,\sigma_i)}{\hat{Z}(\sigma_1,\ldots,\sigma_{i-1})} \approx e^{\pm 1/n^3} \cdot \mu_{v_i}^{\sigma_1,\ldots,\sigma_{i-1}}(\sigma_i)$

where $e^{-1/2n^3} \le \frac{\hat{Z}(\cdots)}{Z(\cdots)} \le e^{1/2n^3}$ by approx. counting

Jerrum-Valiant-Vazirani Sampler

[**J**errum-**V**aliant-**V**azirani '86]

∃ an efficient algorithm that samples from $\hat{\mu}$

and evaluates $\hat{\mu}(\sigma)$ given any $\sigma \in \{0,1\}^V$

multiplicative error: $\forall \sigma \in \{0,1\}^V: \quad e^{-1/n^2} \le \frac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le e^{1/n^2}$

Sample a random $\boldsymbol{Y} \sim \hat{\mu}$;
pick $\boldsymbol{Y}_0 = \emptyset$;
accept $\boldsymbol{Y}$ with prob.: $q = \frac{\hat{\mu}(\boldsymbol{Y}_0)}{\hat{\mu}(\boldsymbol{Y})} \cdot e^{-\frac{3}{n^2}} \in \left[e^{-5/n^2}, 1\right]$
fail if otherwise;

$\forall \sigma \in \{0,1\}^V$:

$$\Pr[\boldsymbol{Y} = \sigma \wedge \text{ accept}] = \hat{\mu}(\sigma) \cdot \frac{\hat{\mu}(\emptyset)}{\hat{\mu}(\sigma)} \cdot e^{-\frac{3}{n^2}} \propto \begin{cases} 1 & \sigma \text{ is ind. set} \\ 0 & \text{otherwise} \end{cases}$$

Boosting Local Inference

SSM → local approx. inference

each v computes a $\hat{\mu}_v^\sigma$ within r-ball

additive error:

$$d_{\mathrm{TV}}\left(\hat{\mu}_v^\sigma, \mu_v^\sigma\right) \le \frac{1}{\mathrm{poly}(n)}$$

multiplicative error:

$$\frac{\hat{\mu}_v^\sigma(0)}{\mu_v^\sigma(0)}, \frac{\hat{\mu}_v^\sigma(1)}{\mu_v^\sigma(1)} \in \left[\mathrm{e}^{-1/\mathrm{poly}(n)}, \mathrm{e}^{1/\mathrm{poly}(n)}\right]$$

SSM —*local self-reduction*→ both are achievable with $r = O(\log n)$

boosted sequential r-local sampler: $r = O(\log n)$

- scan vertices in V in an arbitrary order $v_1, v_2, \ldots, v_n$
- for i=1,2, …, n: sample Y_{v_i} according to $\hat{\mu}_{v_i}^{Y_{v_1},\ldots,Y_{v_{i-1}}}$

multiplicative error: $\forall \sigma \in \{0,1\}^V : \ \mathrm{e}^{-1/n^2} \le \frac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le \mathrm{e}^{1/n^2}$

SLOCAL JVV

Scan vertices in V in an arbitrary order $v_1, v_2, \ldots, v_n$:

pass 1: sample $\boldsymbol{Y} \in \{0,1\}^V$ by *boosted* sequential r-local sampler $\hat{\mu}$;

$$\forall \sigma \in [q]^V: \quad e^{-1/n^2} \le \frac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le e^{1/n^2}$$

$r = O(\log n)$

pass 1': construct a sequence of ind. sets $\emptyset = \boldsymbol{Y}_0, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n = \boldsymbol{Y}$;

s.t. $\forall\ 0 \le i \le n$:
- $\boldsymbol{Y}_i$ agrees with $\boldsymbol{Y}$ over $v_1, \ldots, v_i$
- $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_{i-1}$ differ only at v_i

each v_i: bad event A_{v_i} occurs independently with $\Pr[A_{v_i}] = 1 - q_{v_i}$

where $q_{v_i} = \dfrac{\hat{\mu}(\boldsymbol{Y}_{i-1})}{\hat{\mu}(\boldsymbol{Y}_i)} \cdot e^{-3/n^2} \in [e^{-5/n^2}, 1]$

$O(\log n)$-local to compute

$\boldsymbol{Y} = (Y_v)_{v \in V}$ is accepted if no bad event occurs

Scan vertices in V in an arbitrary order $v_1, v_2, \ldots, v_n$:

pass 1: sample $Y \in \{0,1\}^V$ by *boosted* sequential r-local sampler $\hat{\mu}$;

$$\forall \sigma \in [q]^V: \quad e^{-1/n^2} \le \frac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le e^{1/n^2}$$

$r = O(\log n)$

pass 1': construct a sequence of ind. sets $\emptyset = Y_0, Y_1, \ldots, Y_n = Y$;

s.t. $\forall\ 0 \le i \le n$:
- Y_i agrees with Y over $v_1, \ldots, v_i$
- Y_i and Y_{i-1} differ only at v_i

each v_i: bad event A_{v_i} occurs independently with $\Pr[A_{v_i}] = 1 - q_{v_i}$

where $q_{v_i} = \dfrac{\hat{\mu}(\boldsymbol{Y}_{i-1})}{\hat{\mu}(\boldsymbol{Y}_i)} \cdot e^{-3/n^2} \quad \in [e^{-5/n^2}, 1]$

$$\forall \sigma \in \{0,1\}^V:$$

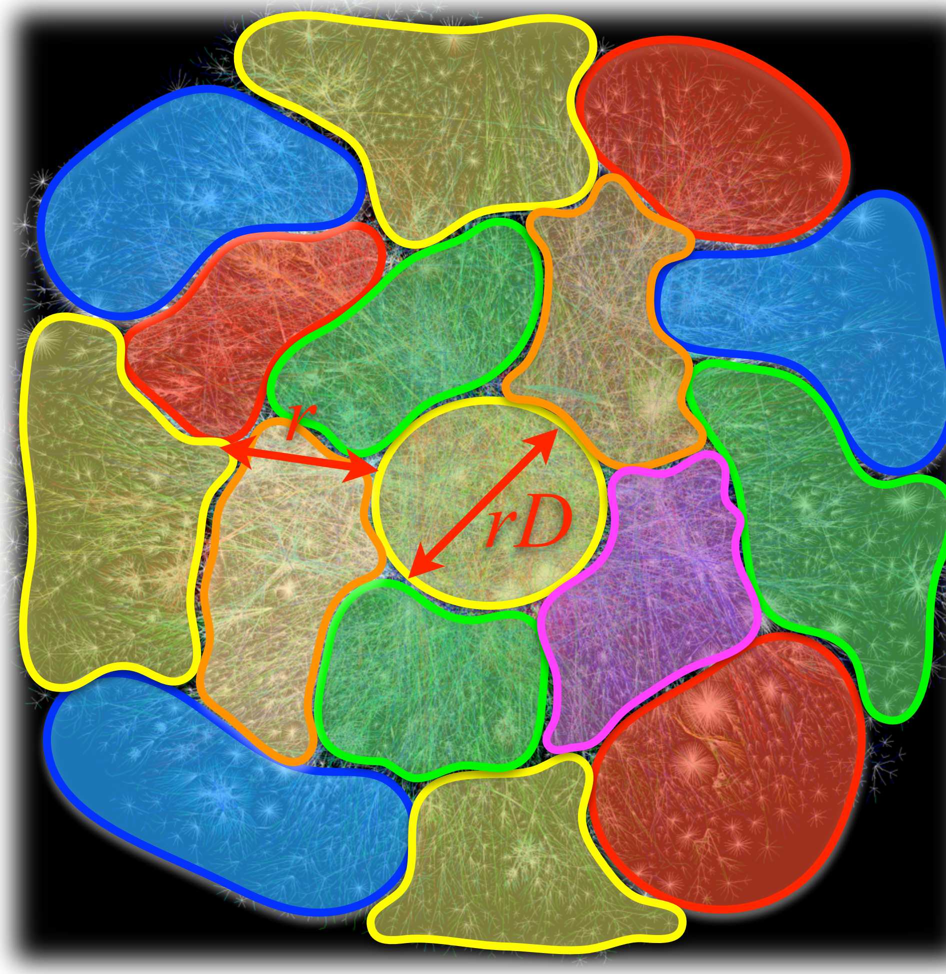
$$\Pr[\boldsymbol{Y} = \sigma \wedge \text{accept}] = \hat{\mu}(\sigma) \prod_{i=1}^{n} q_{v_i} = \hat{\mu}(\sigma) \prod_{i=1}^{n} \left(\frac{\hat{\mu}(\boldsymbol{Y}_{i-1})}{\hat{\mu}(\boldsymbol{Y}_i)} \cdot e^{-3/n^2} \right) \Bigg|_{\boldsymbol{Y}_n = \boldsymbol{Y} = \sigma}$$

$$= \hat{\mu}(\sigma) \cdot \frac{\hat{\mu}(\emptyset)}{\hat{\mu}(\sigma)} \cdot e^{-\frac{3}{n}} \quad \propto \begin{cases} 1 & \sigma \text{ is ind. set} \\ 0 & \text{otherwise} \end{cases}$$

$(C,D)^r$ -network-decomposition of G:

- classifies vertices into clusters;
- assign each cluster a color in $[C]$;
- each cluster has diameter $\leq D$ in G^r;
- clusters with same color are $>r$ distance away from each other.

C colors



Given a $(C,D)^r$- ND:

- each vertex v has an ind. local random source X_v;
- each v assigned with color c in ND can compute in $O(rcD)$ rounds:
  - a random indicator $Y_v \in \{0,1\}$
  - the local function q_v to determine bad event A_v

even with access only to the part of ND with colors $\leq c$

$\boldsymbol{Y}$ conditioned on no A_v's occurring follows Gibbs distribution μ.

An *LLL*-like Framework

Each v holds: an ind. random variable X_v with domain Ω

a bad event A_v

each A_v is associated with
- variable set $\mathsf{vbl}(v) \subseteq [n]$
- function $q_v : \Omega^{\mathsf{vbl}(v)} \to [0,1]$

Each v maps random sources $X_{\mathsf{vbl}(v)}$ to final output Y_v by a local function.

Rejection sampling:

- Each v draws an ind. sample of X_v and maps $X_{\mathsf{vbl}(v)}$ to Y_v;
- each A_v occurs independently with prob. $1 - q_v(X_{\mathsf{vbl}(v)})$;
- the sample $\boldsymbol{Y} = (Y_v)_{v \in V}$ is accepted if no A_v occurs.

Target distribution μ^*: $\boldsymbol{Y}$ conditioned on accepted

Local Rejection Sampling

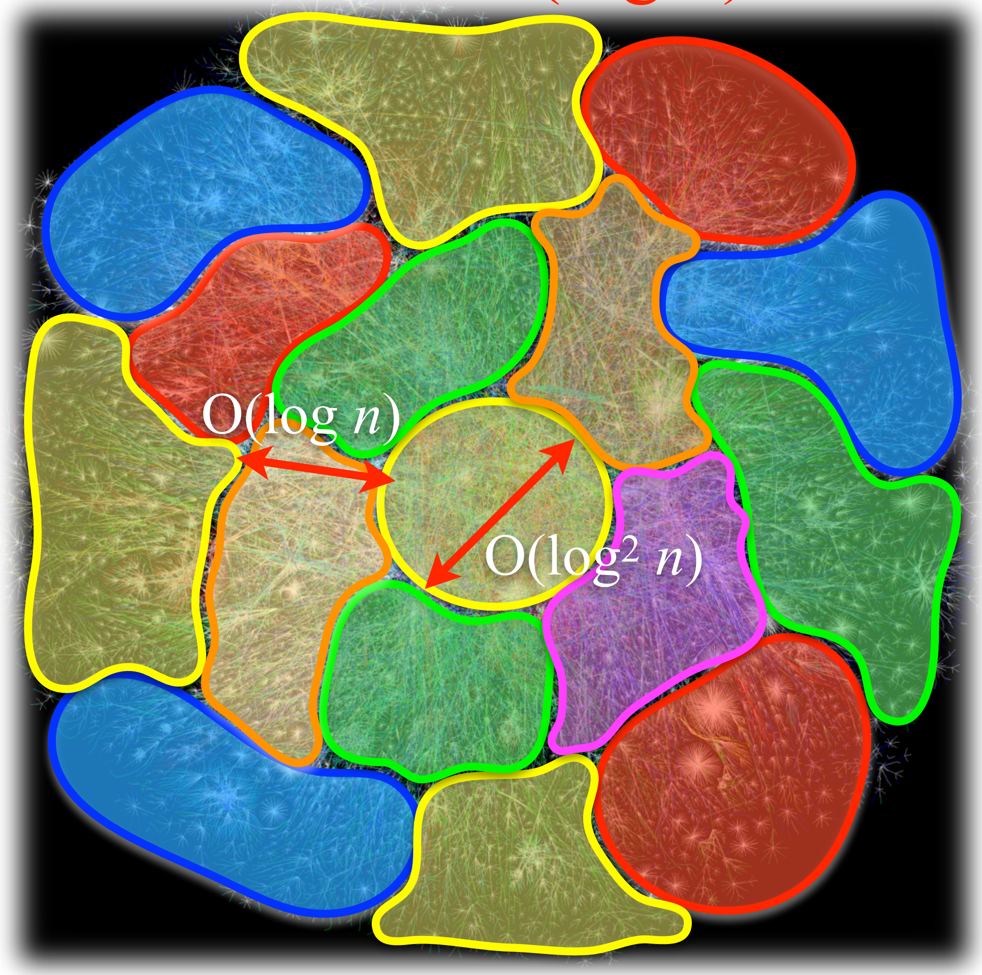
- Each v draws ind. sample of X_v and computes Y_v from $X_{\mathsf{vbl}(v)}$.
- Each v violates A_v ind. with $\Pr[A_v]=1-q_v(X_{\mathsf{vbl}(v)})$.
- In each iteration: for each v with A_v violated:
  - resample all variables in vbl(v) and update Y_v;
  - resample A_v with $\Pr[A_v] = 1-q_v(X_{\mathsf{vbl}(v)})$;
  - for non-violated A_u that shares variables with A_v: resample A_u with $\Pr[A_u] = 1 - \mathrm{e}^{-5/n^2} \cdot q_u\left(X_{\mathsf{vbl}(u)}\right)/q_u\left(X^{\mathtt{old}}_{\mathsf{vbl}(u)}\right)$.

Given a $(C,D)^r$- ND: $r = O(\log n)$ determined by SSM decay rate

- each iteration costs $O(rCD)$ rounds in $\mathcal{LOCAL}$ model;
- terminates in $O(1)$ iterations *w.h.p.*;
- upon termination: $\boldsymbol{Y} \sim \mu$.

$(C,D)^r$-network-decomposition of G:

- classifies vertices into clusters;
- assign each cluster a color in $[C]$;
- each cluster has diameter $\leq D$ in G^r;
- clusters with same color are $>r$ distance away from each other.



[Linial, Saks, 1993]

$(C, D)^r$- ND constructed in $O(rCD)$ rounds by a Las Vegas process

with fixed $D=O(\log n)$ and random $C=O(\log n)$ *w.h.p.*

- each vertex v has an ind. local random source X_v;
- each v assigned with color c in ND can compute in $O(rcD)$ rounds:
  - a random indicator $Y_v \in \{0,1\}$
  - the local function q_v to determine bad event A_v

even with access only to the part of ND with colors $\leq c$

Local Rejection Sampling

- Each v draws ind. sample of X_v and computes Y_v from $X_{\mathsf{vbl}(v)}$.
- Each v violates A_v ind. with $\Pr[A_v]=1-q_v(X_{\mathsf{vbl}(v)})$.
- In each iteration: for each v with A_v violated:
  - resample all variables in vbl(v) and update Y_v;
  - resample A_v with $\Pr[A_v] = 1-q_v(X_{\mathsf{vbl}(v)})$;
  - for non-violated A_u that shares variables with A_v:
    resample A_u with $\Pr[A_u] = 1 - \mathrm{e}^{-5/n^2} \cdot q_u\left(X_{\mathsf{vbl}(u)}\right)/q_u\left(X^{\mathtt{old}}_{\mathsf{vbl}(u)}\right)$.

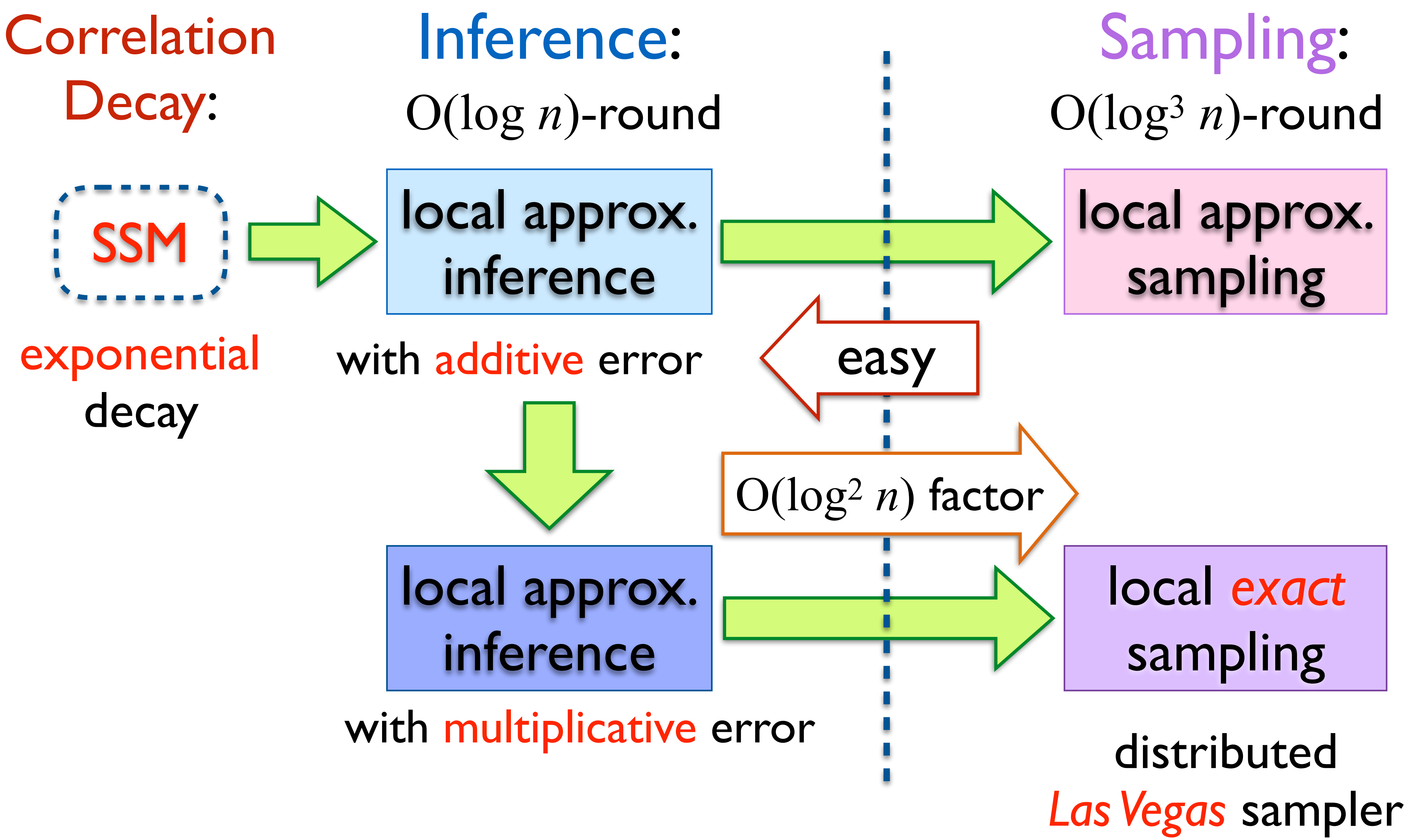
work even for **dynamically** incoming bad events

$(O(\log n), O(\log n))^{O(\log n)}$- ND is constructed: one color c at a time

- each iteration costs $O(c \log^2 n)$ rounds in $\mathcal{LOCAL}$ model;
- terminates in $O(1)$ iterations *w.h.p.*; $O(\log^3 n)$ rounds *w.h.p.*
- upon termination: $\boldsymbol{Y} \sim \mu$.

Locality of Sampling

For Gibbs distributions (distributions defined by *local* factors):



Algorithmic Implications

(due to the state-of-the-arts of strong spatial mixing)

- $O(\sqrt{\Delta}\log^3 n)$ -round distributed algorithm for sampling matchings in graphs with max-degree Δ;
- $O(\log^3 n)$ -round distributed algorithms for sampling:
  - hardcore model (weighted independent set) in the uniqueness regime;
  - antiferromagnetic Ising model in the uniqueness regimes;
  - antiferromagnetic 2-spin systems in the uniqueness regimes;
  - weighted hypergraph matchings in the uniqueness regimes;
  - uniform q-coloring/list-coloring when $q>1.763\ldots\Delta$ in triangle-free graphs with max-degree Δ;
  - $\ldots$ $\ldots$

Local Exact Sampler

Uniform sampling ind. set in graphs with max-degree Δ:

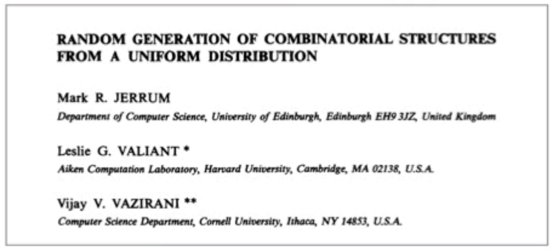
When $\Delta \leq 5$:

- SSM holds;
- $\exists$ $O(\log^3 n)$-round distributed Las Vegas sampler.

[Feng, Sun, Y., PODC'17]:

If $\Delta \geq 6$, there is an infinite sequence of graphs G with $diam(G) = n^{\Omega(1)}$ such that even approx. sampling ind. set requires $\Omega(diam)$ rounds.

Counting and Sampling

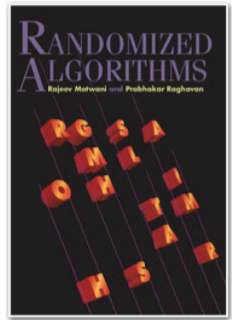


RANDOM GENERATION OF COMBINATORIAL STRUCTURES FROM A UNIFORM DISTRIBUTION

Mark R. JERRUM
Department of Computer Science, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom

Leslie G. VALIANT *
Aiken Computation Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.

Vijay V. VAZIRANI **
Computer Science Department, Cornell University, Ithaca, NY 14853, U.S.A.



[Jerrum-Valiant-Vazirani '86]:

(For *self-reducible* problems)

approx. counting is tractable ⟺ (approx., exact) sampling is tractable

Computational Phase Transition

Sampling almost-uniform independent set in graphs with maximum degree Δ:

- [Weitz, STOC'06]: If $\Delta \leq 5$, poly-time.
- [Sly, FOCS'10]: If $\Delta \geq 6$, no poly-time algorithm unless **NP**=**RP**.

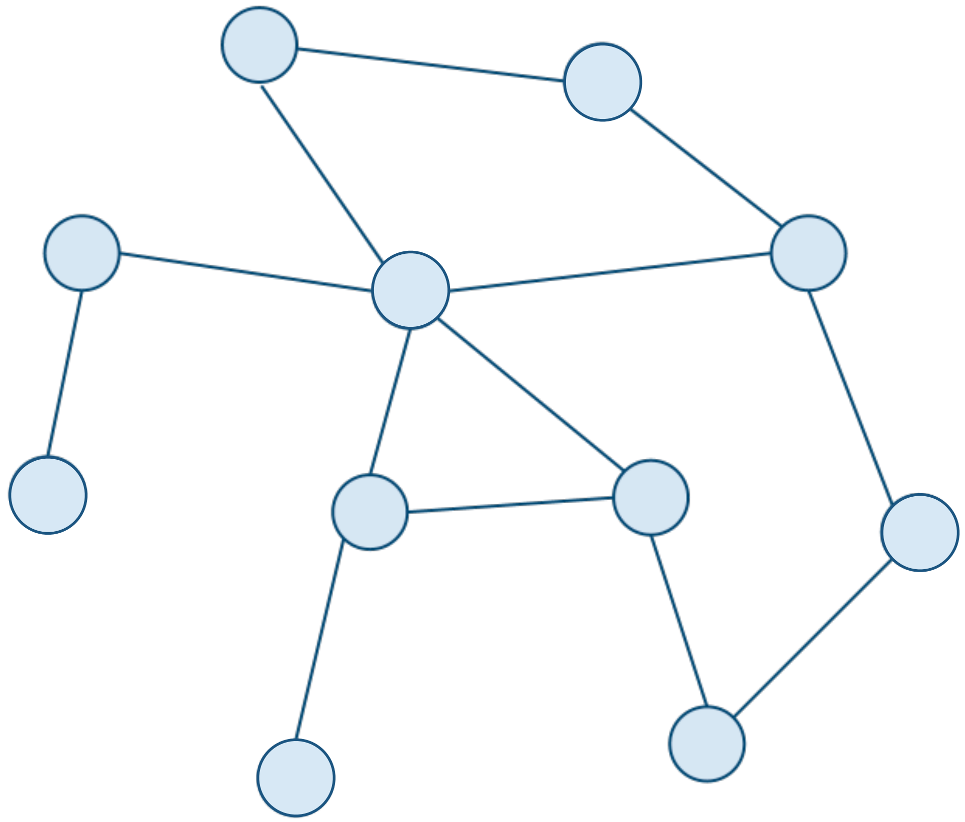
A phase transition occurs when Δ: $5 \to 6$.

Hold for Local Computation!

Message-Passing Algorithms

($\mathcal{LOCAL}$ model with bounded memory/communication)

- Communications are synchronized.
- Each node v has an independent random source X_v.
- In each round, each node can:
  - exchange messages with neighbors
  - perform local computation
  - read/write to local memory
- msg/memory size $= O(\log n)$ or even $O(1)$ bits.



Distributed Gibbs Samplers that may work in practice

- Parallelization of Glauber dynamics:
  - "Hogwild!" —— biased
  - chromatic scheduler —— $\Omega(\Delta \log n)$ rounds
- (lazy) Local Metropolis. —— approximate
- Local Rejection Sampling. —— exact, dynamic

$O(\log n)$ rounds (for Local Metropolis and Local Rejection Sampling)

Local Rejection Sampling

$A_e : [q] \times [q] \to [\beta, 1]$ $\qquad$ $b_v : [q] \to \mathbb{R}_{\geq 0}$

- each vertex $v \in V$ ind. samples a spin state $\sigma_v \in [q] \propto \boldsymbol{b}_v$;
- each edge $e=(u,v) \in E$ fails ind. with prob. $1\text{-}A_e(\sigma_u,\sigma_v)$;
- while there is a failed edge: $\boldsymbol{\sigma}^{\text{old}} \leftarrow$ current $\boldsymbol{\sigma}$
  - resample σ_v for all vertices v involved in failed edges;
  - each failed $e=(u,v)$ is revived ind. with prob. $A_e(\sigma_u,\sigma_v)$;
  - each non-failed $e=(u,v)$ that is incident to a failed edge, fails ind. with prob. $1 - \beta \cdot A_e(\sigma_u,\sigma_v)/A_e(\sigma_u^{\text{old}},\sigma_v^{\text{old}})$;

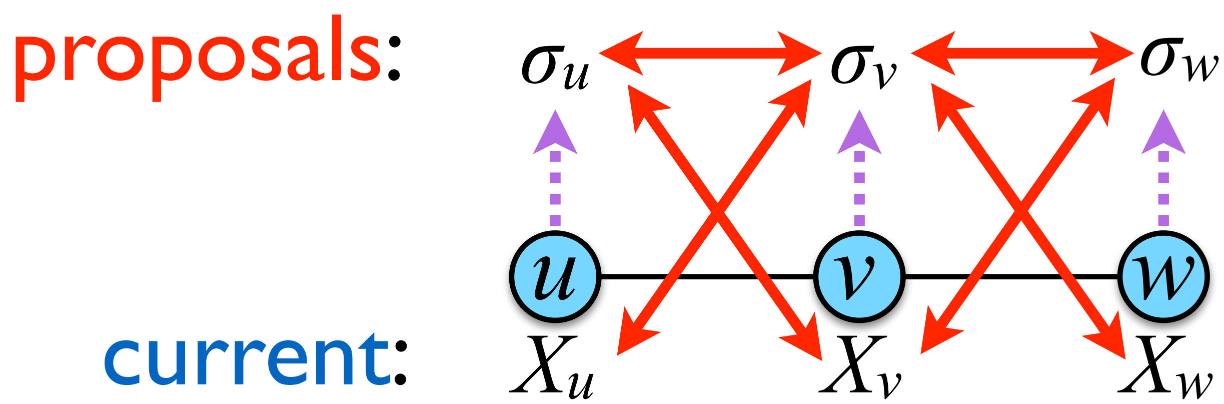
Pros:
- local & parallel
- dynamic graph
- exact sampler
- certifiable termination

Cons:
- convergence is hard to analyze
- regime is not tight
- soft constraints

Local Metropolis

[Feng, Sun, Y. '17] [Feng, Y. '18]

A_e: $[q] \times [q] \to [0,1]$ $\qquad$ b_v: $[q] \to [0,1]$



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

- each vertex $v \in V$ ind. proposes a spin state $\sigma_v \in [q] \propto \boldsymbol{b}_v$;
- each edge fails ind. with prob. $1\text{-}A_e(X_u,\sigma_v)A_e(\sigma_u,X_v)A_e(\sigma_u,\sigma_v)$;
- each vertex $v \in V$ accepts its proposal and update X_v to σ_v if *none of its edge fails*.

Thank you!

Feng, Liu, Y. ***Local rejection sampling with soft filters.*** arxiv: 1807.06481.

Feng, Hayes, Y. ***Distributed Symmetry Breaking in Sampling (Optimal Distributed Randomly Coloring with Fewer Colors).*** arxiv: 1802.06953.

Feng, Y. ***On local distributed sampling and counting.***
In PODC'18. arxiv: 1802.06686.

Feng, Sun, Y. ***What can be sampled locally?*** In PODC'17. arxiv: 1702.00142.