On Local Distributed Sampling and Counting

Yitong Yin
Nanjing University

Joint work with Weiming Feng (Nanjing University)

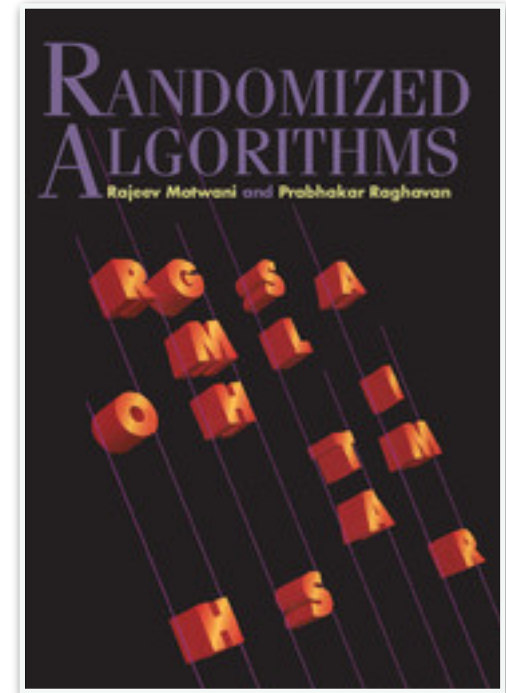
Counting and Sampling

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

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

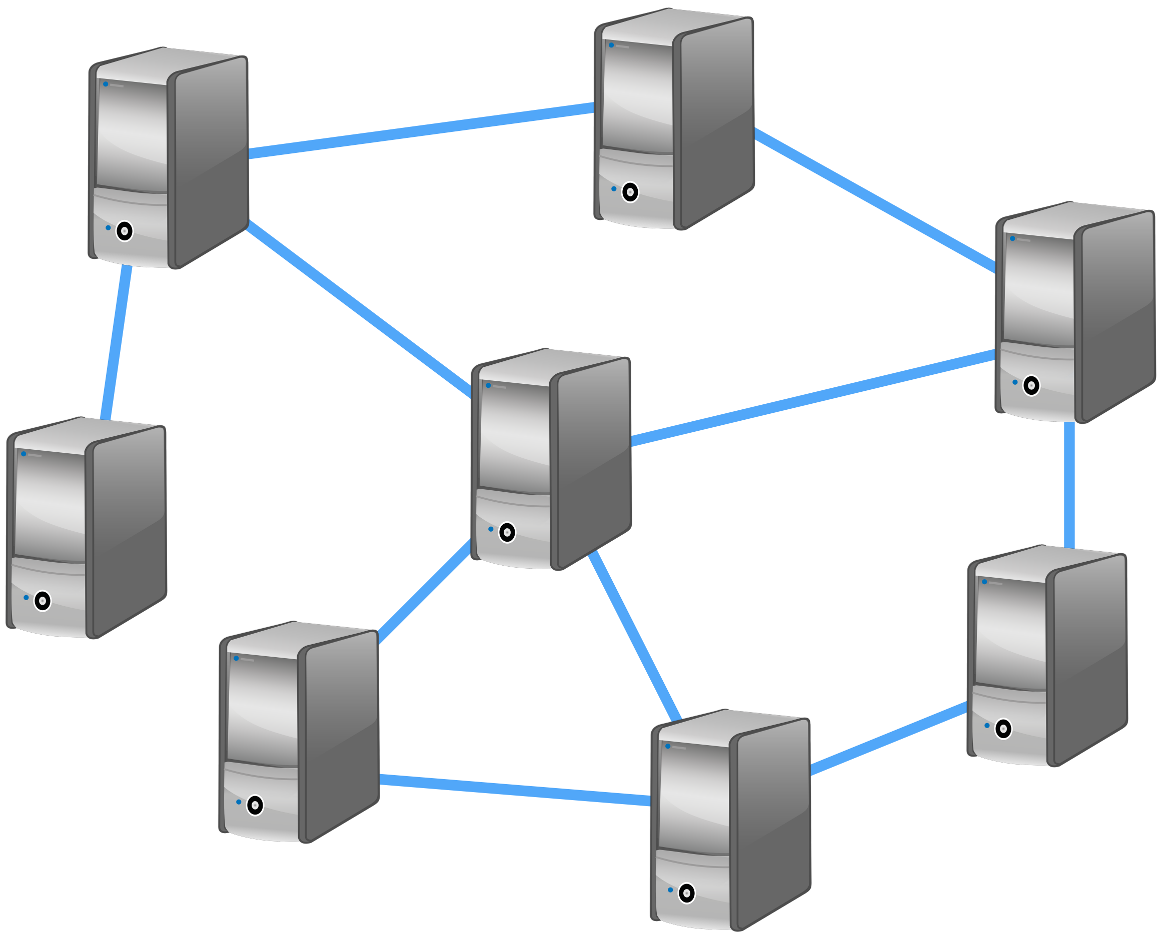
Local Computation?

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
  - read/write to unbounded local memory.



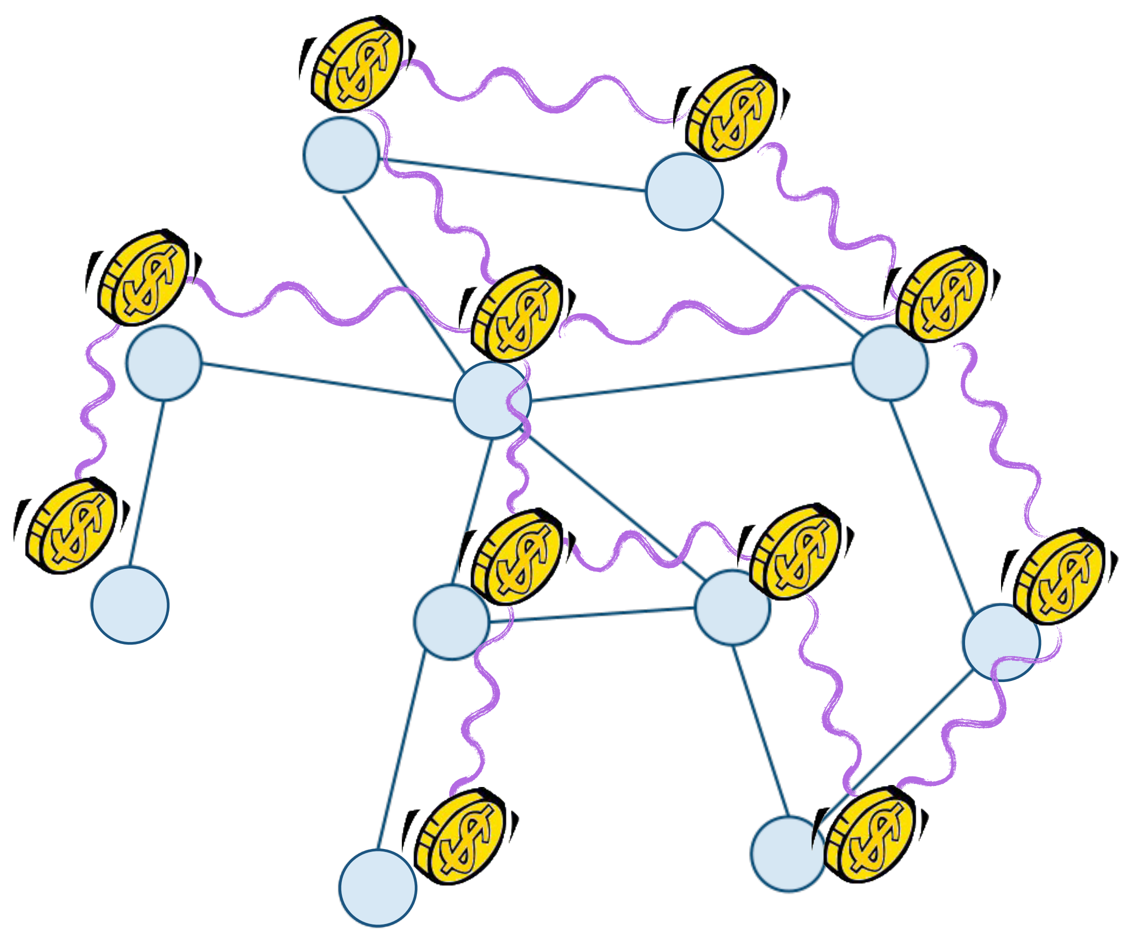
- In t rounds: each node can collect information up to distance t.

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 $\sigma \in \{0,1\}^S$.

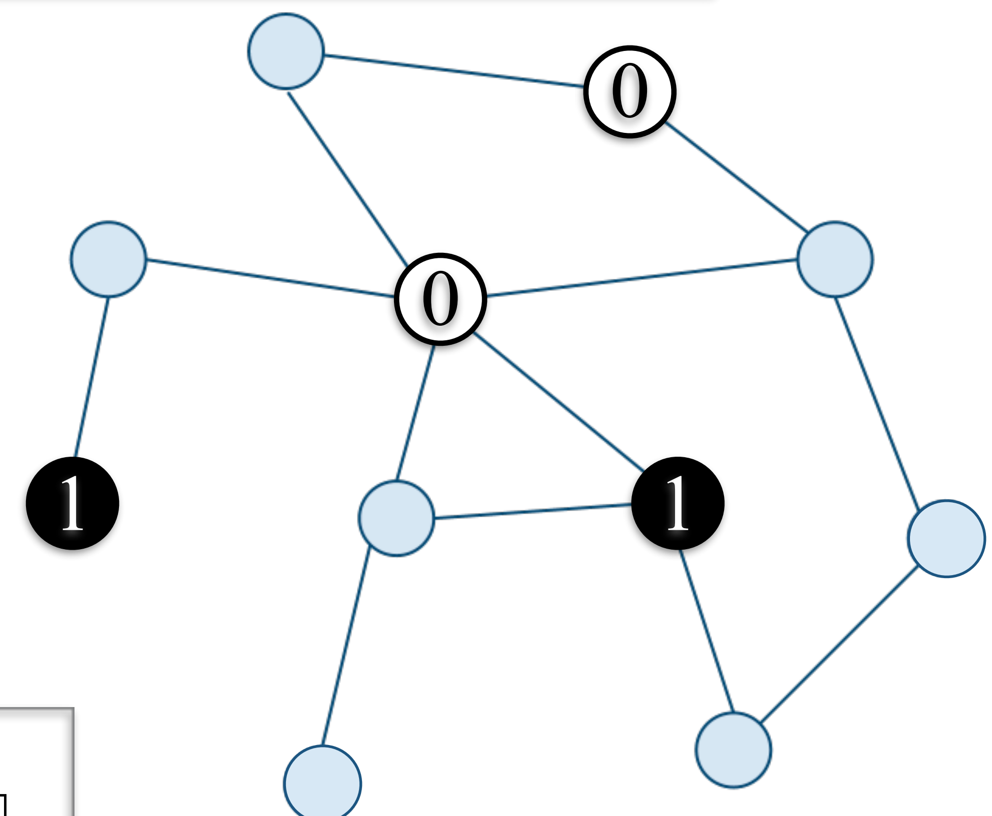
$$\forall y \in \{0,1\}: \quad \mu_v^\sigma(y) = \Pr_{\boldsymbol{Y}\sim\mu}[Y_v = y \mid Y_S = \sigma]$$

- 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 $G(V,E)$

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):

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

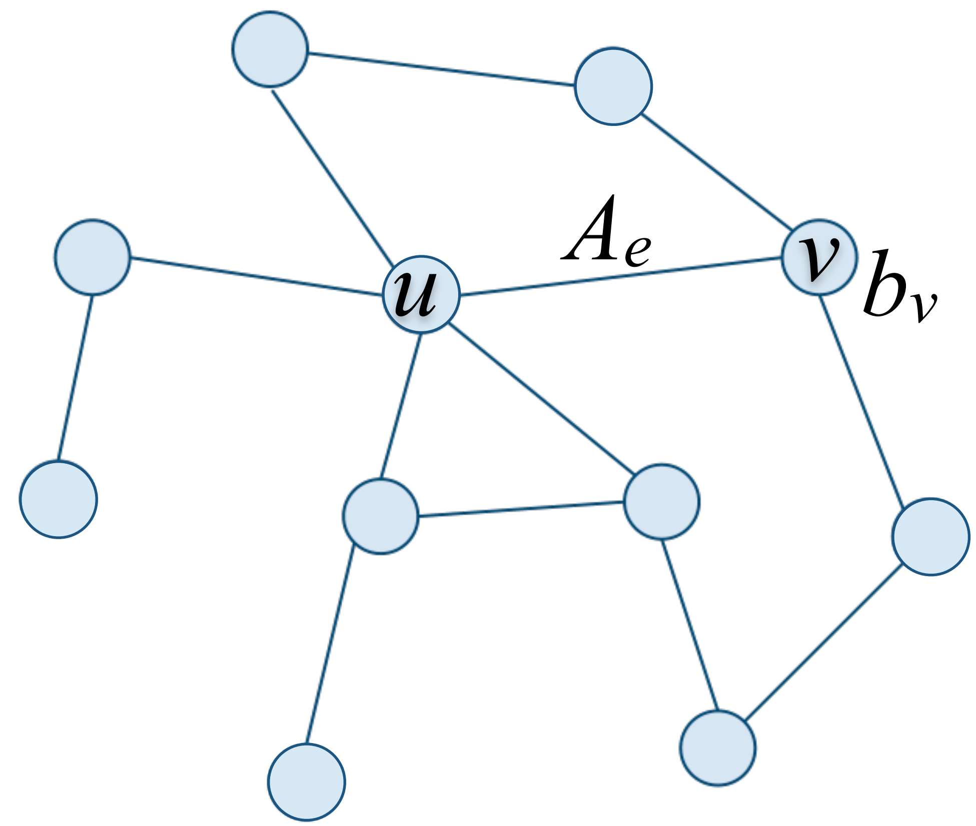
- Each vertex $v\in V$ has a vector (unary constraint):

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

- 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 $G(V,E)$:



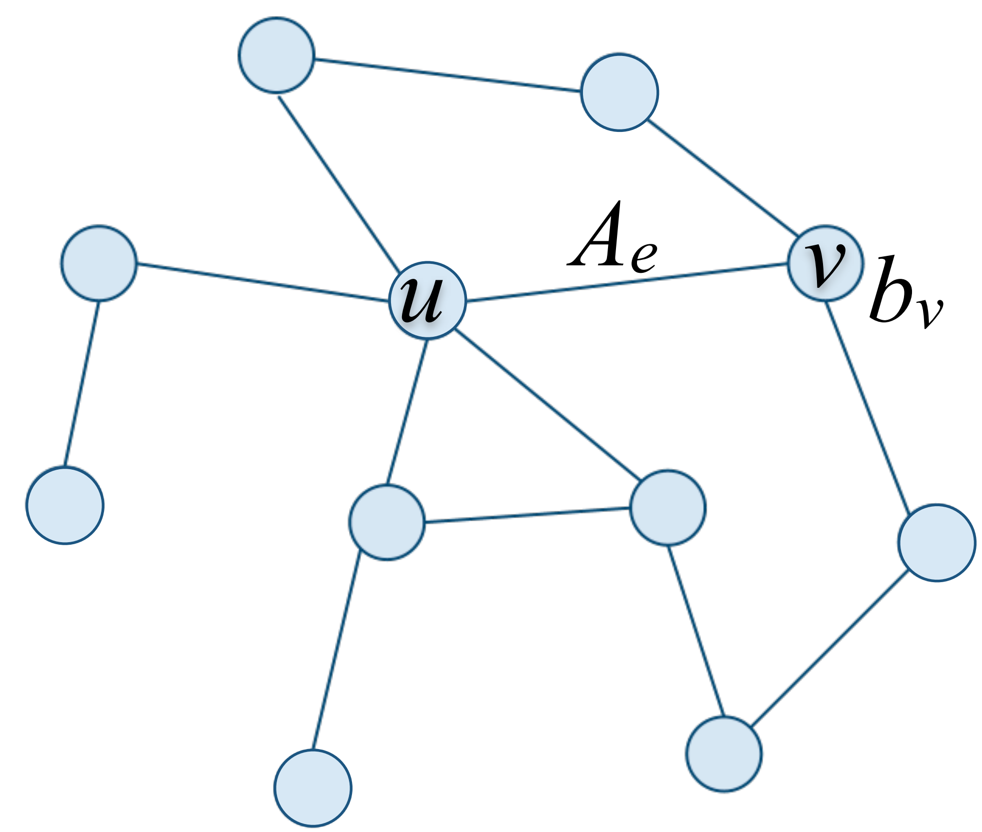
Gibbs Distribution

(with pairwise interactions)

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

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

network $G(V,E)$:



- 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]

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

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

Gibbs Distribution

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

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

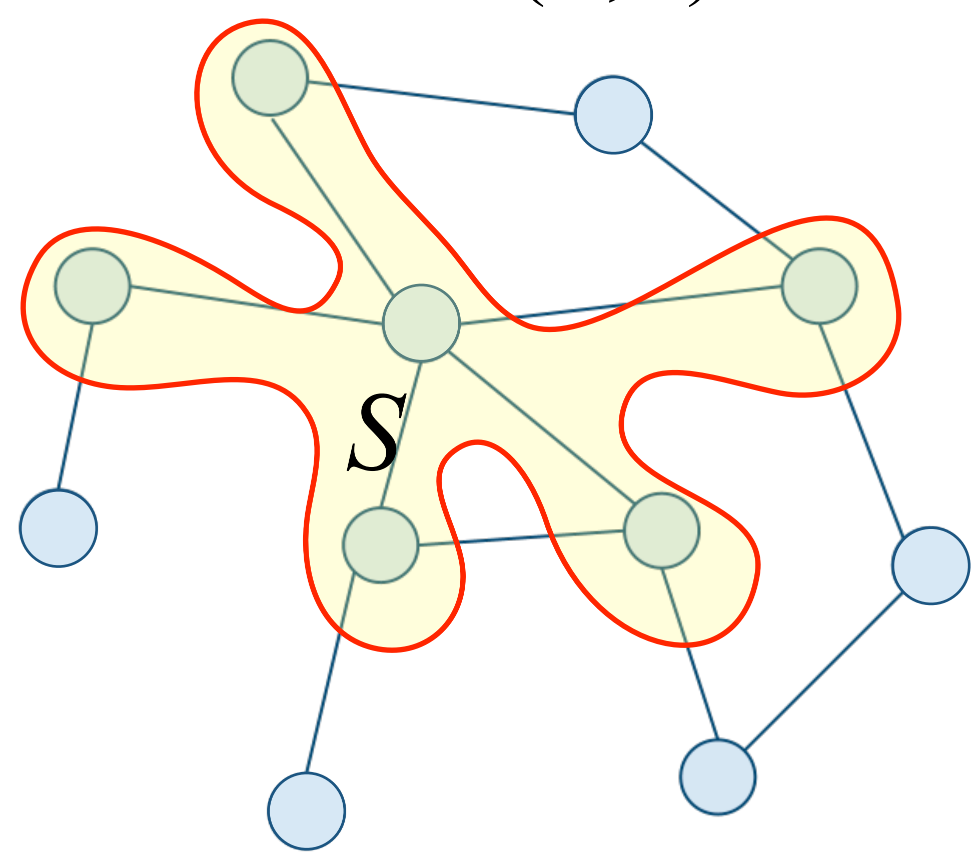
each $(f, S) \in \mathcal{F}$

is a *local* constraints (factors):

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

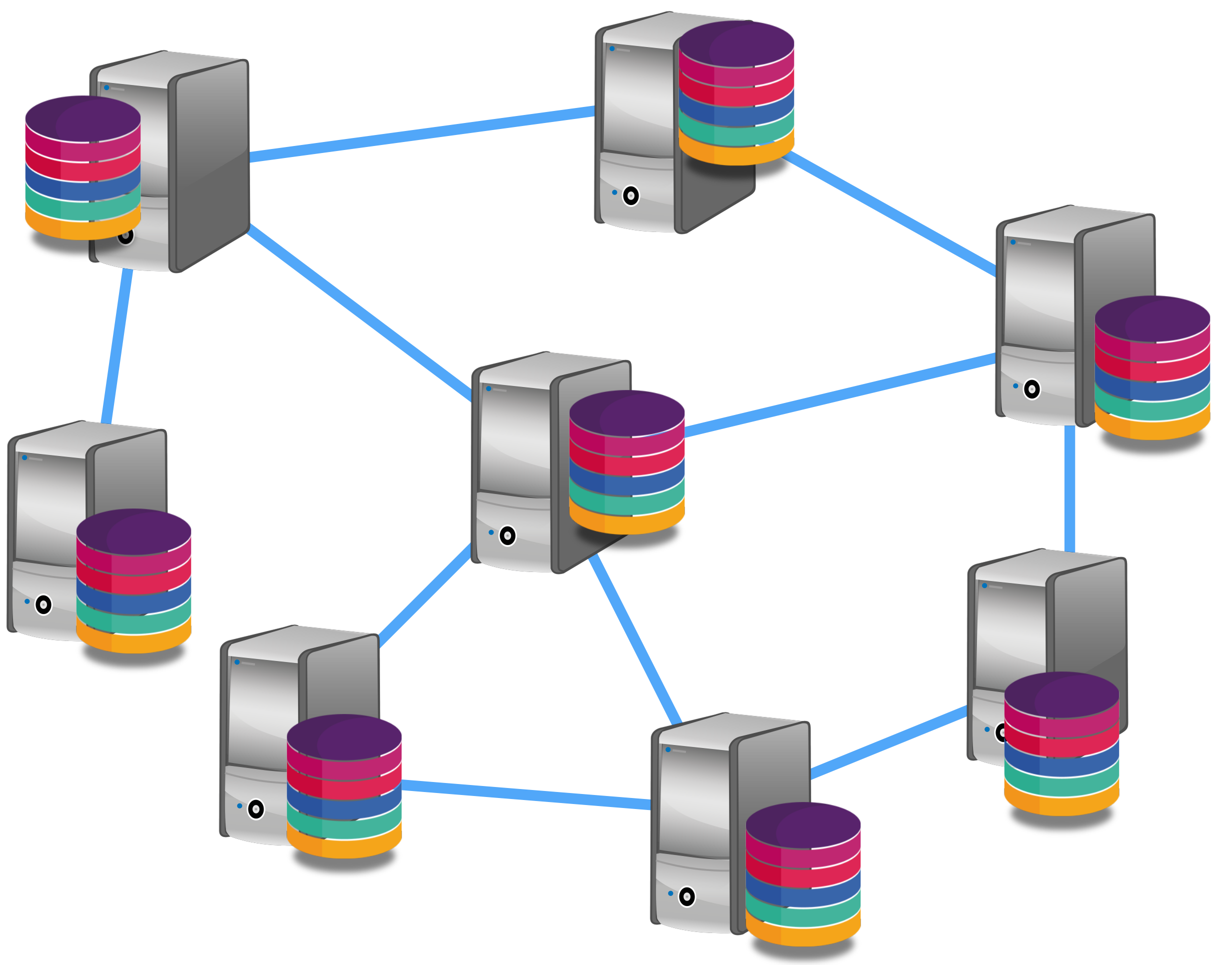
$S \subseteq V$ with $diam_G(S) = O(1)$

network $G(V,E)$:



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*.



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: $5 \rightarrow 6$.

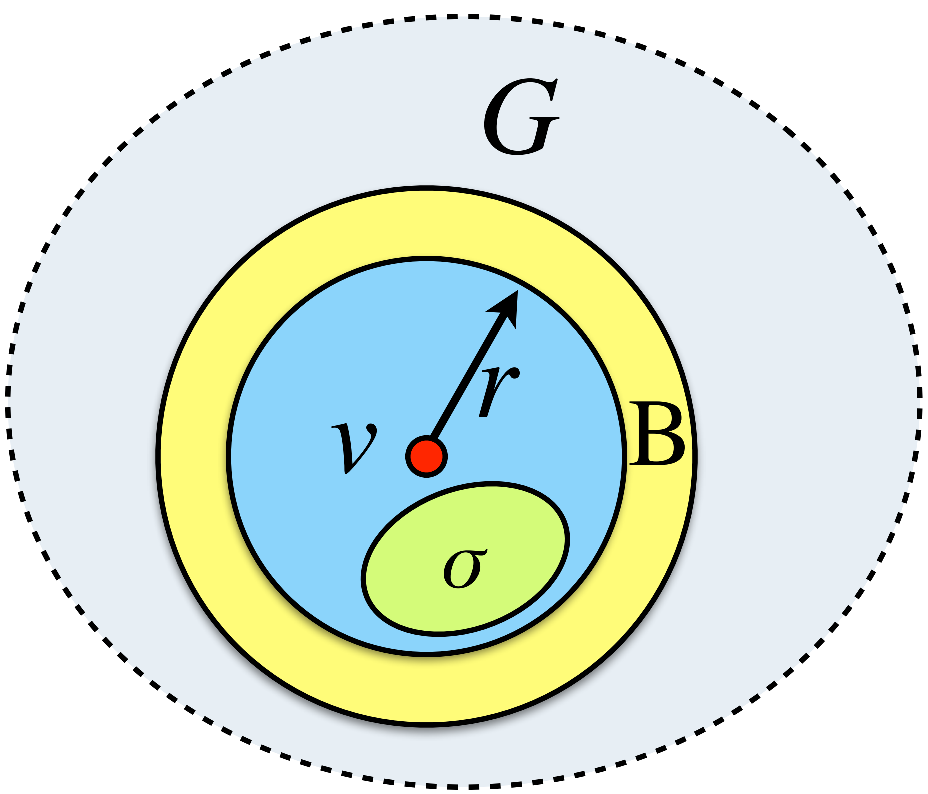
Decay of Correlation

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

strong spatial mixing (SSM):

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

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



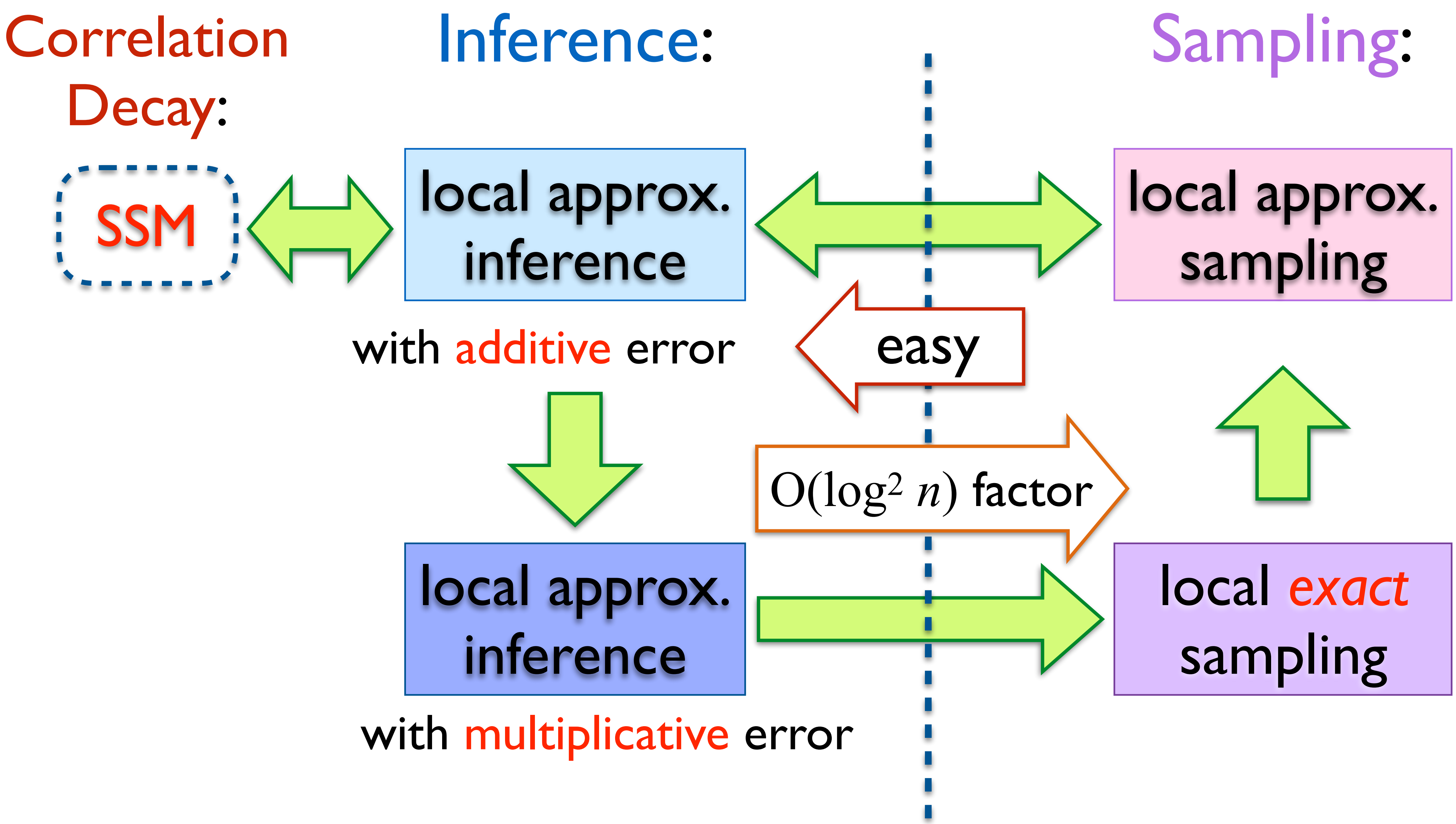
SSM (iff $\Delta \leq 5$ when μ is uniform distribution of ind. sets)

$\Updownarrow$

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

Locality of Counting & Sampling

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



Locality of Sampling

Correlation Decay:

SSM $\Longleftrightarrow$

Inference:

local approx. inference

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

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

$\Longrightarrow$

Sampling:

local approx. sampling

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

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

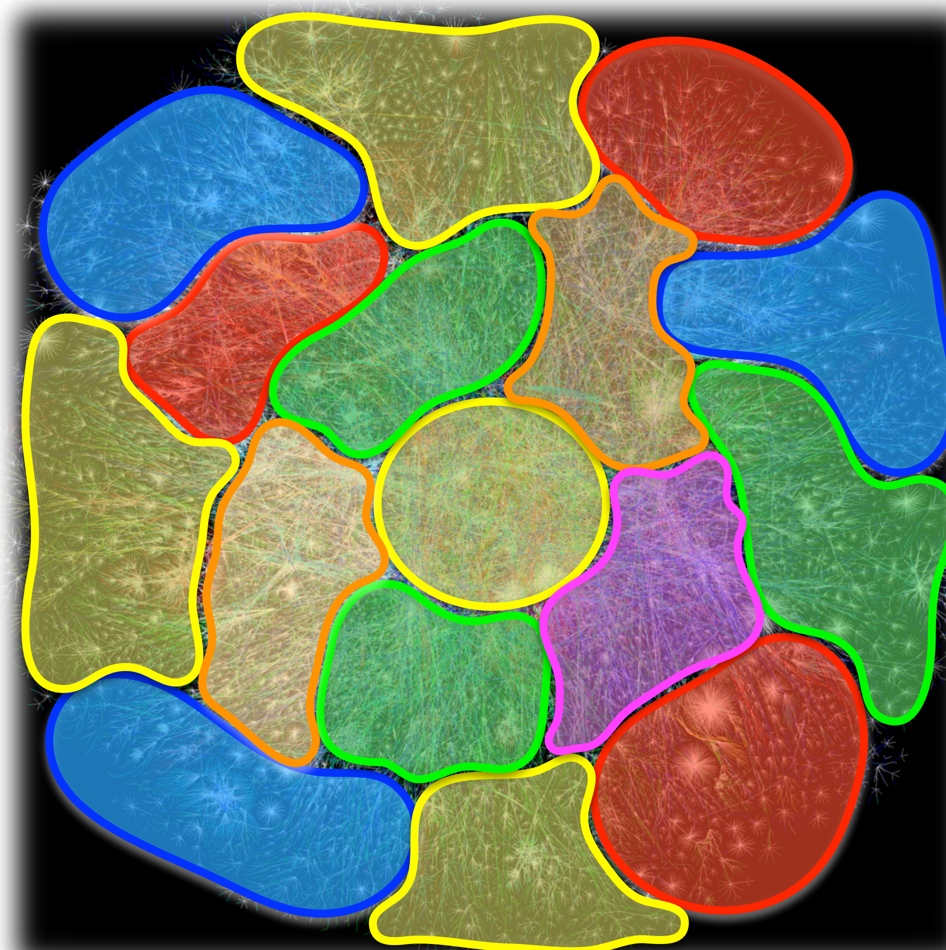
sequential O(log n)-local procedure:

- 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 $\le D$;
- 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.*

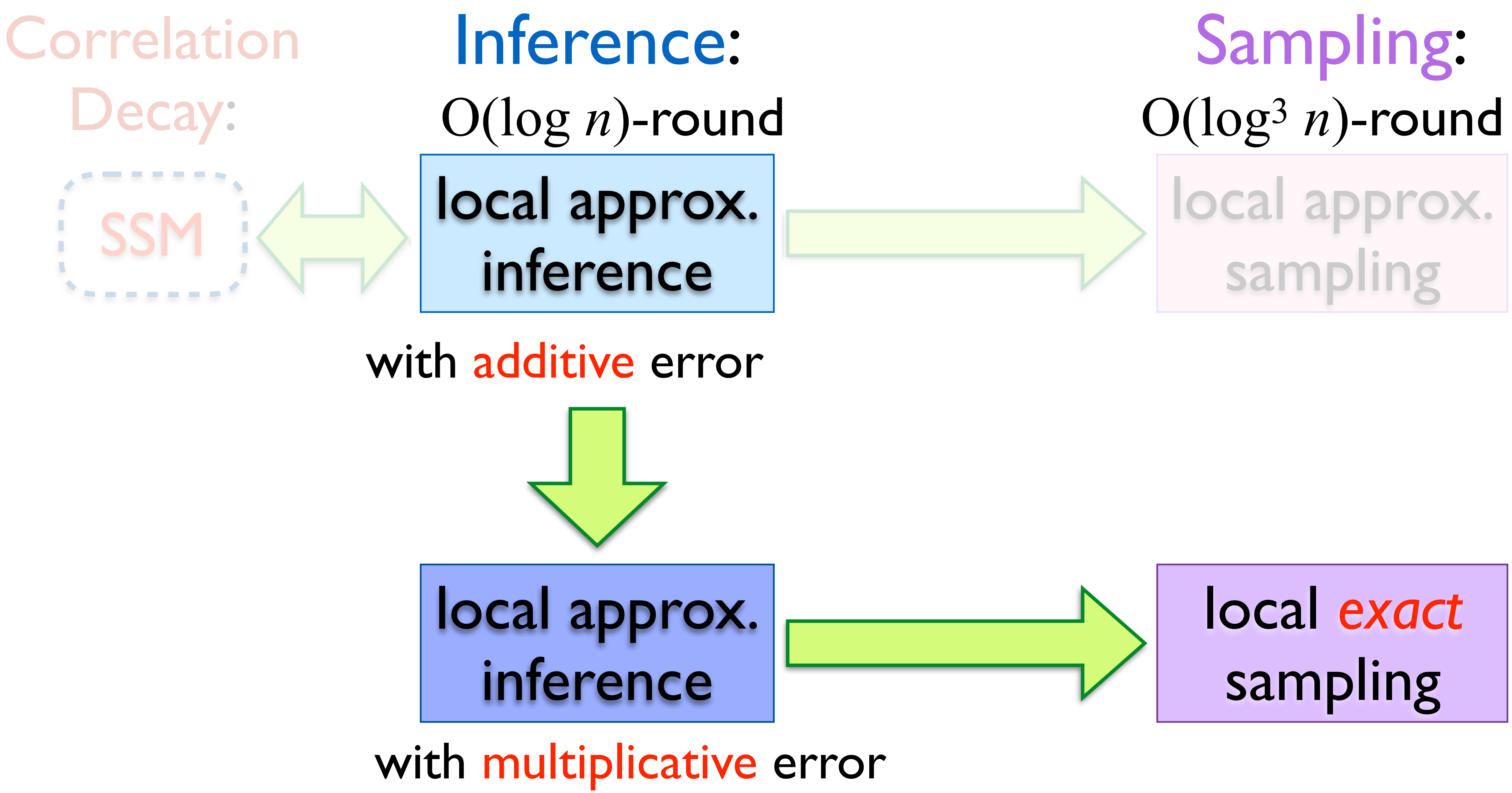
[Linial, Saks, 1993] — [Ghaffari, Kuhn, Maus, 2017]:

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



Local Exact Sampler

In $\mathcal{LOCAL}$ model:

- Each $v \in V$ returns within **fixed** $t(n)$ rounds:
  - local output $Y_v \in \{0,1\}$;
  - local failure $F_v \in \{0,1\}$.
- Succeeds w.h.p.: $\sum_{v \in V} \mathbf{E}[F_v] = \mathrm{O}(1/n)$.
- Correctness: conditioning on success, $\boldsymbol{Y} \sim \mu$.

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

Self-reduction:

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

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 $\xrightarrow{\text{local self-reduction}}$ 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: $\forall \sigma \in \{0,1\}^V : \ \mathrm{e}^{-1/n^2} \le \dfrac{\hat{\mu}(\sigma)}{\mu(\sigma)} \le \mathrm{e}^{1/n^2}$

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

v_i samples $F_{v_i} \in \{0,1\}$ independently with $\Pr[F_{v_i} = 0] = q_{v_i}$

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

Each $v \in V$ returns:

- $Y_v \in \{0,1\}$ to indicate the ind. set;
- $F_v \in \{0,1\}$ indicate failure at v.

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

v_i samples $F_{v_i} \in \{0,1\}$ independently with $\Pr[F_{v_i} = 0] = q_{v_i}$

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

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

$$\Pr[\boldsymbol{Y} = \sigma \wedge \forall i : F_{v_i} = 0] = \hat{\mu}(\sigma) \prod_{i=1}^{n} q_{v_i} = \hat{\mu}(\sigma) \prod_{i=1}^{n} \left(\frac{\hat{\mu}(\boldsymbol{Y}_{i-1})}{\hat{\mu}(\boldsymbol{Y}_i)} \cdot e^{-3/n^2} \right) \Bigg|_{\boldsymbol{Y}_n = \boldsymbol{Y} = \sigma}$$

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

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 $\Rightarrow$

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

Local Exact Sampler

Uniform sampling ind. set in graphs with max-degree $\Delta \leq 5$:

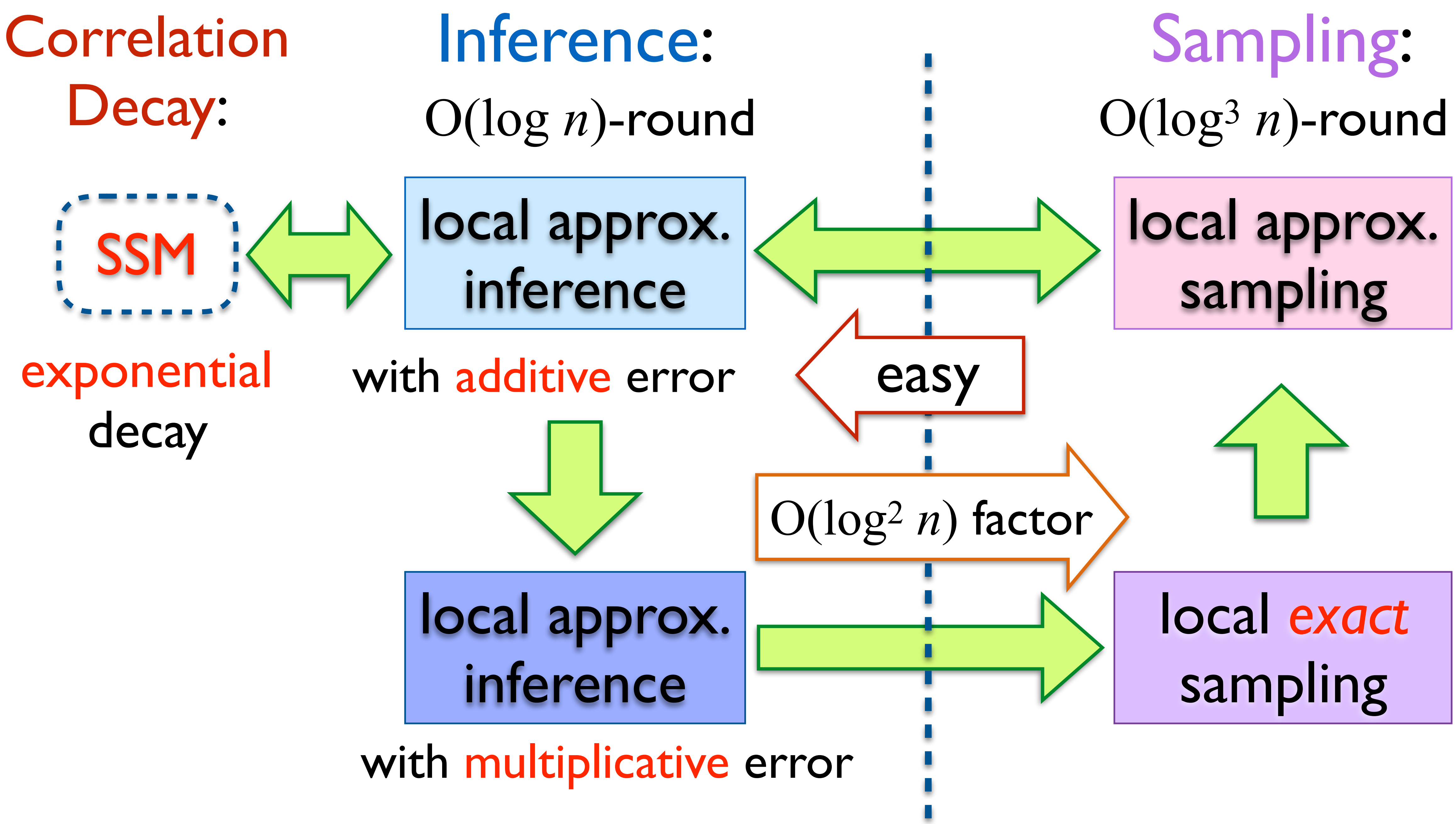
- Each $v \in V$ returns in $O(\log^3 n)$ rounds:
  - local output $Y_v \in \{0,1\}$;
  - local failure $F_v \in \{0,1\}$.
- Succeeds w.h.p.: $\sum_{v \in V} \mathbf{E}[F_v] = O(1/n)$.
- Correctness: conditioning on success, $\boldsymbol{Y} \sim \mu$.

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

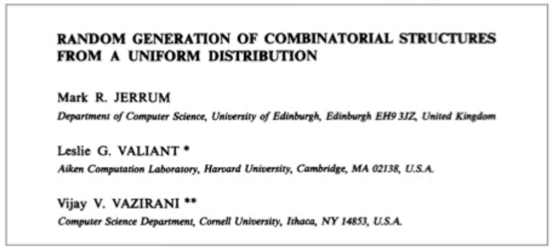
If $\Delta \geq 6$, there is an infinite sequence of graphs G with $diam(G) = n^{\Omega(1)}$ such that even approx. sampling ind. set requires $\Omega(diam)$ rounds.

Locality of Sampling

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



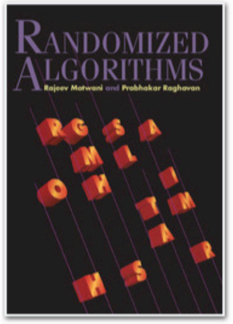
Counting and Sampling


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 $\Longleftrightarrow$ (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: $5 \rightarrow 6$.

Hold for Local Computation!

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;
  - $\ldots$ $\ldots$

Thank you!