Log-Concave IV: Near-Optimal Sampling of Forests

Nima Anari   Kuikui Liu   Shayan Oveis Gharan
Cynthia Vinzant   Thuy-Duong Vuong

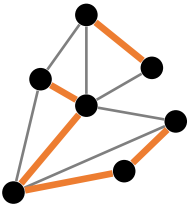
Stanford

November 19, 2022

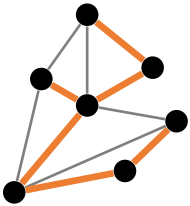
Tree vs. Forest

Given graph $G = (V, E), k = |V|, n = |E|$.

$F \subseteq E$ is a forest if F doesn't contain any cycle.



$T \subseteq E$ is a tree if it is a forest and $|T| = |V| - 1$.



Background

Markov chain to sample uniformly from set of Spanning Trees:

Background

Markov chain to sample uniformly from set of Spanning Trees:

- [Broder'89]; [Aldous-Broder'90]

Background

Markov chain to sample uniformly from set of Spanning Trees:

- [Broder'89]; [Aldous-Broder'90]
- [Colbourn-Myrvold-Neufeld'96]; [Kelner-Madry'09]; [Madry-Straszak-Tarnawski'15]: $\tilde{O}(\min\left\{n\sqrt{k}, k^{\omega}, n^{4/3}\right\})$

Background

Markov chain to sample uniformly from set of Spanning Trees:

- [Broder'89]; [Aldous-Broder'90]
- [Colbourn-Myrvold-Neufeld'96]; [Kelner-Madry'09]; [Madry-Straszak-Tarnawski'15]: $\tilde{O}(\min\left\{n\sqrt{k}, k^{\omega}, n^{4/3}\right\})$
- [Durfee-Kyung-Peebles-Rao-Sachdeva'18]: $\tilde{O}(k^{4/3}n^{1/2} + k^2)$

Background

Markov chain to sample uniformly from set of Spanning Trees:

- [Broder'89]; [Aldous-Broder'90]
- [Colbourn-Myrvold-Neufeld'96]; [Kelner-Madry'09]; [Madry-Straszak-Tarnawski'15]: $\tilde{O}(\min\left\{n\sqrt{k}, k^{\omega}, n^{4/3}\right\})$
- [Durfee-Kyung-Peebles-Rao-Sachdeva'18]: $\tilde{O}(k^{4/3}n^{1/2} + k^2)$
- [Schild18]: $O(n^{1+o(1)})$

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
  $F \sim \{\text{spanning trees}\}$ ☺

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
  $F \sim \{\text{spanning trees}\}$ ☺
  $F \sim \{\text{forests}\}$?

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
- [Feder-Mihail'92]
  Monotonicity/Negative correlation $\Rightarrow$ Efficient sampling

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
- [Feder-Mihail'92]
- [Anari-Liu-OveisGharan-Vinzant'19;Cryan-Guo-Mousa'19]
  Fast sampling from $\mu \sim \{\text{size-}k \text{ forests}\}$

Background

Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
- [Feder-Mihail'92]
- [Anari-Liu-OveisGharan-Vinzant'19;Cryan-Guo-Mousa'19]
  Fast sampling from $\mu \sim \{\text{size-}k \text{ forests}\}$
  Sampling $\Leftrightarrow$ Counting

Background

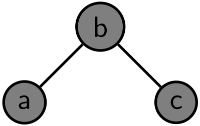
Markov chain to sample uniformly from set of Forests:

- Monotonicity Conjecture: $\mathbb{P}[j \in F \mid i \in F] \leq \mathbb{P}[j \in F]$
- [Feder-Mihail'92]
- [Anari-Liu-OveisGharan-Vinzant'19;Cryan-Guo-Mousa'19]
  Fast sampling from $\mu \sim \{\text{size-}k \text{ forests}\}$
  Sampling $\Leftrightarrow$ Counting
  $\Rightarrow$ Sample/count all-size forests

Background

Application:

- [Goel-Khanna-Raghvendra-Zhang'14]:
  RF-connectivity/liquidity:
  $\text{RF}(u, v) = \mathbb{P}_F[u, v \text{ connected by } F]$



Background

Application:

- [Goel-Khanna-Raghvendra-Zhang'14]:
- [Goel-Ramseyer'20]: liquidity in credit network

Matroid

$\mathcal{B} \subseteq \binom{[n]}{k}$ is set of bases of a matroid $\mathcal{M}$ iff
$\forall B_1, B_2 \in \mathcal{B}, i \in B_1 \setminus B_2, \exists j \in B_2 \setminus B_1$ s.t. $B_1 - i + j \in \mathcal{B}$

Matroid

$\mathcal{B} \subseteq \binom{[n]}{k}$ is set of bases of a matroid $\mathcal{M}$ iff
$\forall B_1, B_2 \in \mathcal{B}, i \in B_1 \setminus B_2, \exists j \in B_2 \setminus B_1$ s.t. $B_1 - i + j \in \mathcal{B}$
$S \subseteq [n]$ is an independent set $\Leftrightarrow \exists B \in \mathcal{B}$ s.t. $S \subseteq B$.

Matroid

$\mathcal{B} \subseteq \binom{[n]}{k}$ is set of bases of a matroid $\mathcal{M}$ iff
$\forall B_1, B_2 \in \mathcal{B}, i \in B_1 \setminus B_2, \exists j \in B_2 \setminus B_1$ s.t. $B_1 - i + j \in \mathcal{B}$
$S \subseteq [n]$ is an independent set $\Leftrightarrow \exists B \in \mathcal{B}$ s.t. $S \subseteq B$.
Fact: $\{S \in I(\mathcal{M}) : |S| = r\}$ form bases of a matroid.

Matroid

Example: for graph $G = (V, E)$,
the spanning trees are the bases of the *graphic matroid*, of rank $|V| - 1$.

Matroid

Example: for graph $G = (V, E)$,
the spanning trees are the bases of the *graphic matroid*, of rank $|V| - 1$.
The forests are independent sets.

Matroid

Example: for graph $G = (V, E)$,
the spanning trees are the bases of the *graphic matroid*, of rank $|V| - 1$.
The forests are independent sets.
The forests of size-r are bases of a matroid

Our results

Given $G = (V, E)$ with $E = [n]$,

Our results

Given $G = (V, E)$ with $E = [n]$,
in $O(n \log^2 n)$ time can approximately sample from

Our results

Given $G = (V, E)$ with $E = [n]$,
in $O(n \log^2 n)$ time can approximately sample from the uniform distribution over spanning trees

Our results

Given $G = (V, E)$ with $E = [n]$,
in $O(n \log^2 n)$ time can approximately sample from
the uniform distribution over spanning trees
the uniform distribution over (all sizes) forests

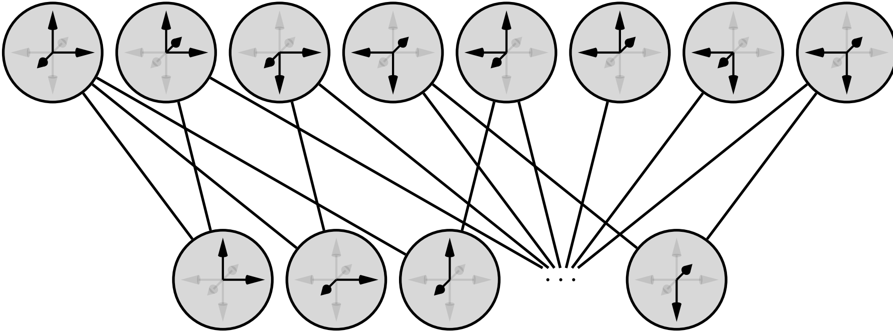
Our results

Given $G = (V, E)$ with $E = [n]$, and weights $q, w_1, \ldots, w_n \geq 0$, in $O(n \log^2 n)$ time can approximately sample from

Our results

Given $G = (V, E)$ with $E = [n]$, and weights $q, w_1, \ldots, w_n \geq 0$, in $O(n \log^2 n)$ time can approximately sample from $\mu(F) = q^{|V|-1-|F|} \prod_{i \in F} w_i$ when F is a forest, and 0 else

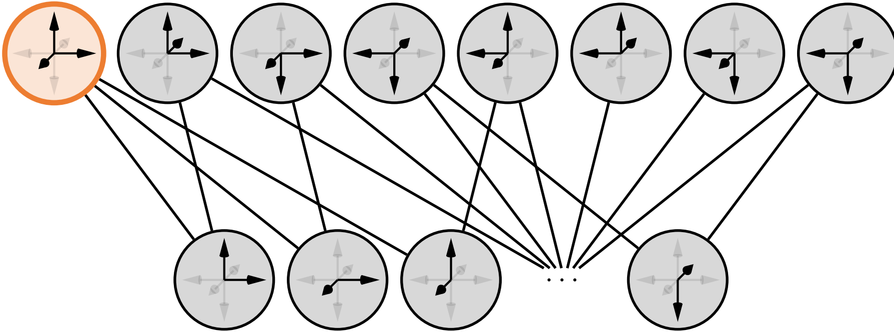
Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.

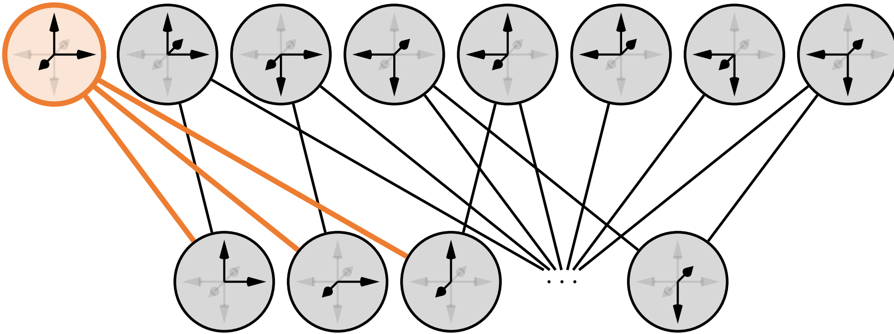
Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.

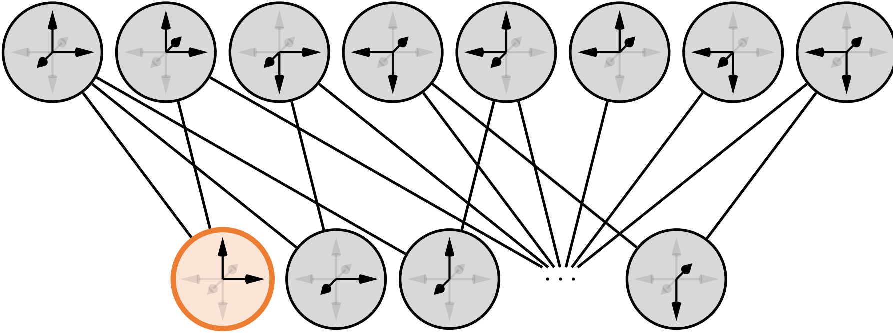
Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.

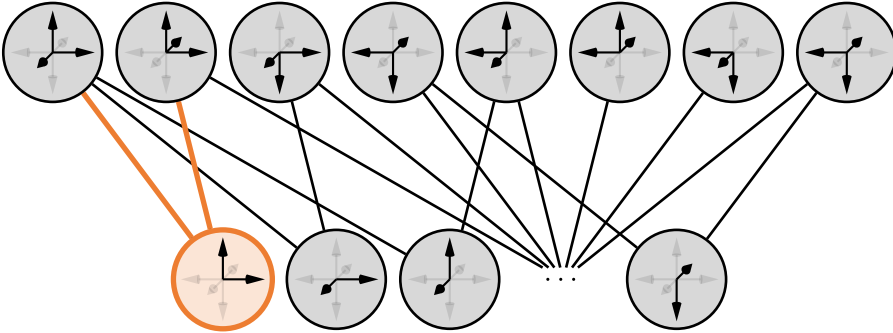
Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.

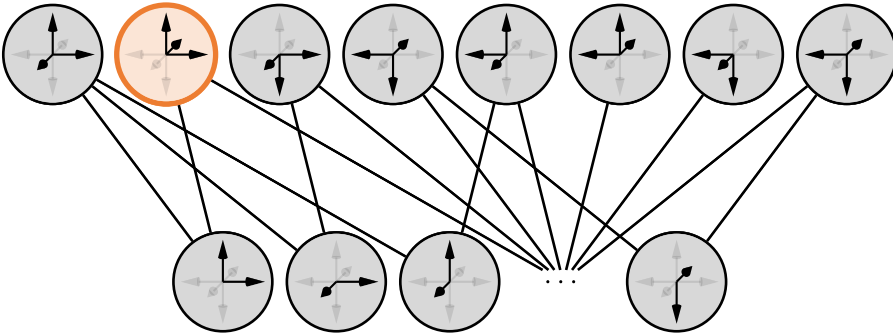
Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.
2. Add an element with probability $\propto \mu$(resulting set).

Down-up walk



Random walk to sample from distribution μ over $\binom{[n]}{k}$

1. Drop an element uniformly at random.
2. Add an element with probability $\propto \mu(\text{resulting set})$.

Down-up walk

For $\mu \sim \mathcal{B}(\mathcal{M})$ with rank$(\mathcal{M}) = k$, the down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$

Down-up walk

For $\mu \sim \mathcal{B}(\mathcal{M})$ with $\text{rank}(\mathcal{M}) = k$, the down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k} \to t_{\text{mix}} \leq k^2 \log n$.

Down-up walk

For $\mu \sim \mathcal{B}(\mathcal{M})$ with $\text{rank}(\mathcal{M}) = k$, the down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$
- [CGM'19]: has MLSI constant $\geq \frac{1}{k}$

Down-up walk

For $\mu \sim \mathcal{B}(\mathcal{M})$ with $\text{rank}(\mathcal{M}) = k$, the down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$
- [CGM'19]: has MLSI constant $\geq \frac{1}{k}$
  $\rightarrow t_{\text{mix}} \leq k(\log k + \log\log n)$.
- [ALOV**V**'21] $t_{\text{mix}} \leq k \log k$

Down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$
- [CGM'19]: has MLSI constant $\geq \frac{1}{k}$
- [ALOV**V**'21] $t_{\text{mix}} \leq k \log k$
  new!: exchange inequality

Down-up walk

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$
- [CGM'19]: has MLSI constant $\geq \frac{1}{k}$
- [ALOV**V**'21] $t_{\text{mix}} \leq k \log k$
  No dependent on $\pi_0 \to$ application for determinantal dist.

Down-up walk

Generalization: μ with $f_\mu = \sum_S \mu(S) \prod_{i \in S} z_i$ log-concave [Gurvits;ALOV'19;Branden-Huh'19]

- [ALOV'19]: has spectral gap $\geq \frac{1}{k}$
- [CGM'19]: has MLSI constant $\geq \frac{1}{k}$
- [ALOV**V**'21] $t_{\text{mix}} \leq k \log k$

Warm-up: sampling trees

Sample from μ uniform over trees:

- can naively use down-up walk on μ.

Warm-up: sampling trees

Sample from μ uniform over trees:

- can naively use down-up walk on μ. But, up-step may take $O(|E|)$ time to implement ☹

Warm-up: sampling trees

Sample from μ uniform over trees:

- can naively use down-up walk on μ. But, up-step may take $O(|E|)$ time to implement ☹
- Instead, use down-up walk on $\bar{\mu}$, $\bar{\mu}(S) = 1[E \setminus S \text{ is tree}]$

Warm-up: sampling trees

Sample from μ uniform over trees:

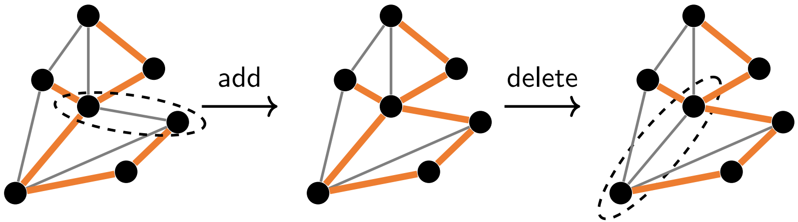
- can naively use down-up walk on μ. But, up-step may take $O(|E|)$ time to implement ☹
- Instead, use down-up walk on $\bar{\mu}$, $\bar{\mu}(S) = 1[E \setminus S \text{ is tree}]$
  Down-step = add edge. Up-step = remove edge from cycle.

Warm-up: sampling trees

Sample from μ uniform over trees:

- can naively use down-up walk on μ. But, up-step may take $O(|E|)$ time to implement ☹
- Instead, use down-up walk on $\bar{\mu}$, $\bar{\mu}(S) = 1[E \setminus S \text{ is tree}]$
  Down-step = add edge. Up-step = remove edge from cycle.
  Implementable in amortized $O(\log|E|)$ via link-cut tree [Russo-Teixeira-Francisco'18] ☺

Figure

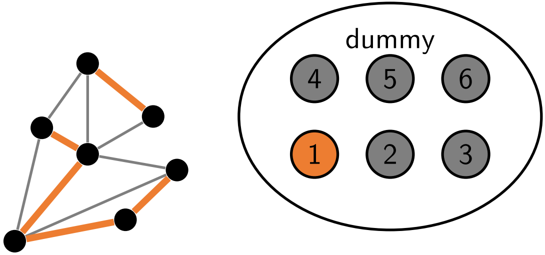


Sampling forests

- Sample complements of forest

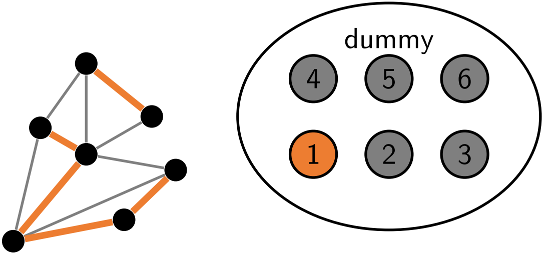
Sampling forests

- Sample complements of forest
- Homogenizing: add dummy variables $y_1, \ldots, y_k$ with $k = |V| - 1$



Sampling forests

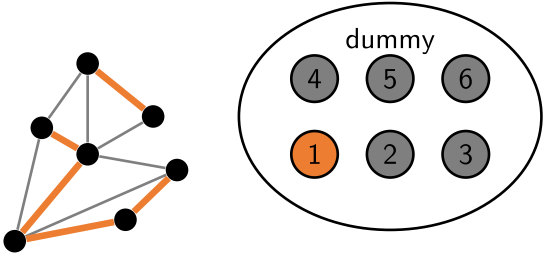
- Sample complements of forest
- Homogenizing: add dummy variables $y_1, \ldots, y_k$ with $k = |V| - 1$



- $\#\{|, \bullet\} = |E|$.

Sampling forests

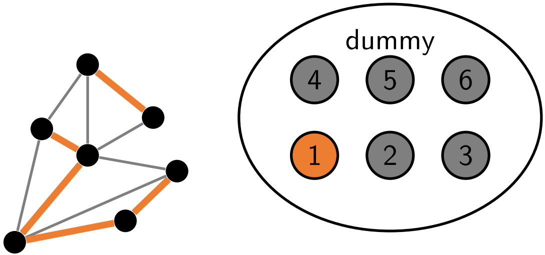
- Sample complements of forest
- Homogenizing: add dummy variables $y_1, \ldots, y_k$ with $k = |V| - 1$



- $\#\{|, \bullet\} = |E|$. $\#\{|, \bullet\} = k$.

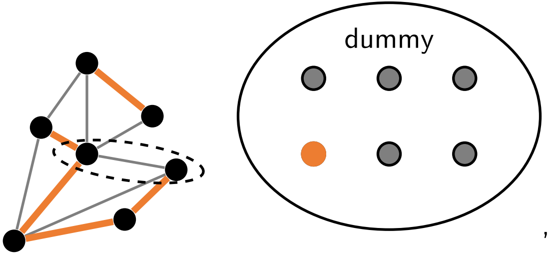
Sampling forests

- Sample complements of forest
- Homogenizing: add dummy variables $y_1, \ldots, y_k$ with $k = |V| - 1$



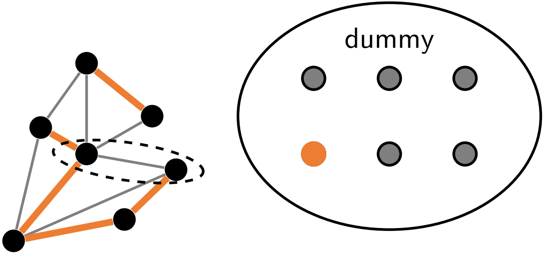
- $\#\{|, \bullet\} = |E|$. $\#\{|, \bullet\} = k$.
  $\bar{\mu}^{\uparrow}(S \cup Y) \propto \frac{1}{\binom{k}{|S|}} 1[E \setminus S = \text{forest}]$

Down-up walk on $\bar{\mu}^{\uparrow}$



Remove back edge that creates an orange cycle C

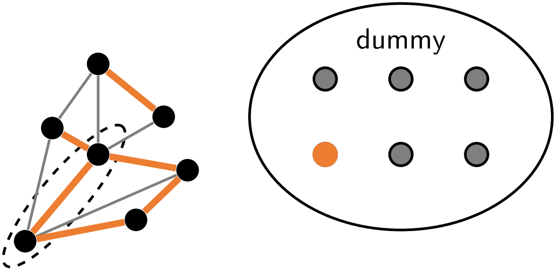
Down-up walk on $\bar{\mu}^{\uparrow}$



,

Add orange edge that creates an orange cycle C

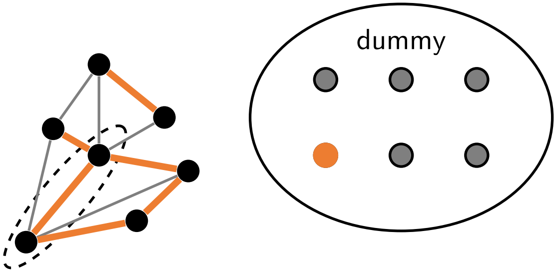
Down-up walk on $\bar{\mu}^{\uparrow}$



,

Add random black edge from C

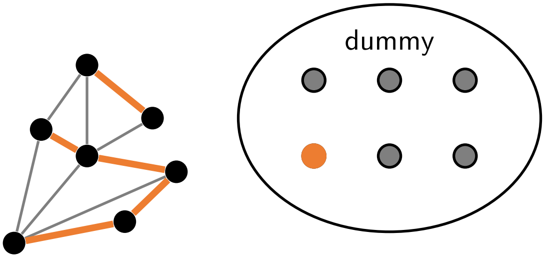
Down-up walk on $\bar{\mu}^{\uparrow}$



,

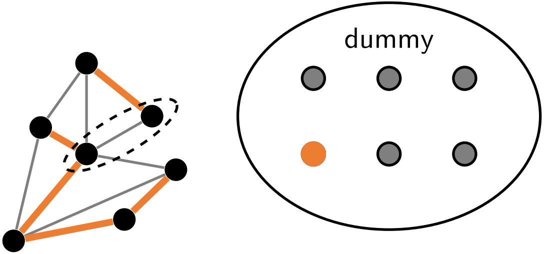
Remove random orange edge from C

Down-up walk on $\bar{\mu}^{\uparrow}$



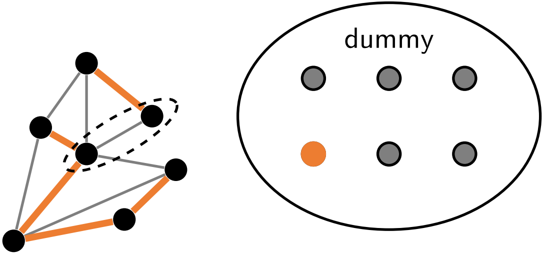
,

Down-up walk on $\bar{\mu}^{\uparrow}$



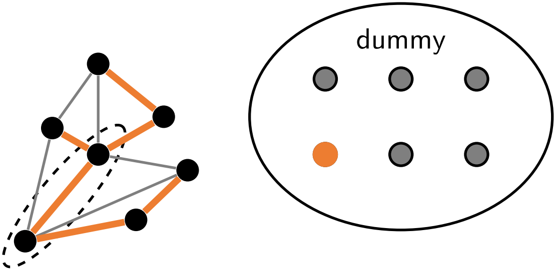
Remove back edge that does not create cycles

Down-up walk on $\bar{\mu}^{\uparrow}$



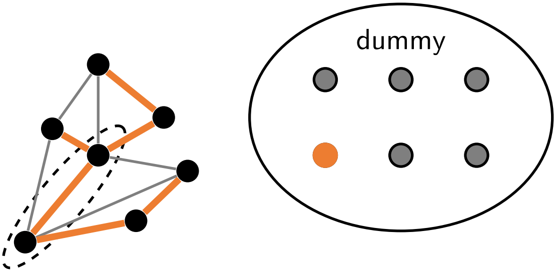
Add orange edge that does not create cycles

Down-up walk on $\bar{\mu}^{\uparrow}$



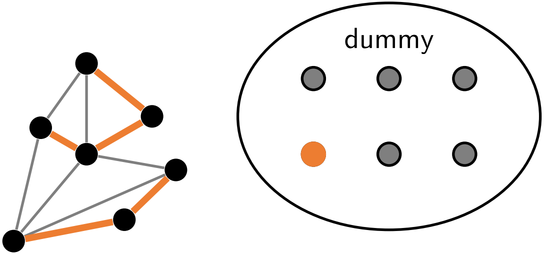
Add random black edge from forest

Down-up walk on $\bar{\mu}^{\uparrow}$

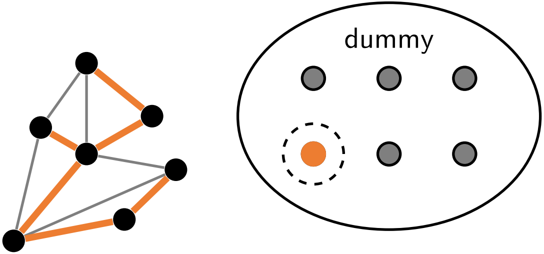


Remove random orange edge from forest

Down-up walk on $\bar{\mu}^{\uparrow}$

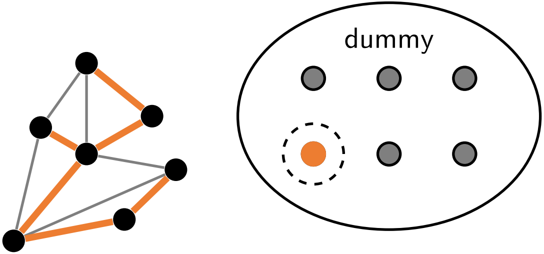


Down-up walk on $\bar{\mu}^{\uparrow}$



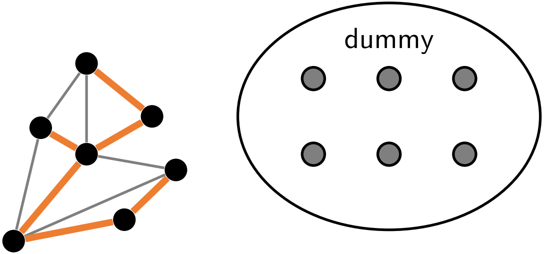
Add random black dummy node

Down-up walk on $\bar{\mu}^{\uparrow}$

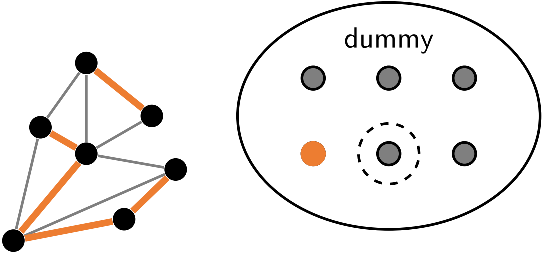


Remove random orange dummy node

Down-up walk on $\bar{\mu}^{\uparrow}$

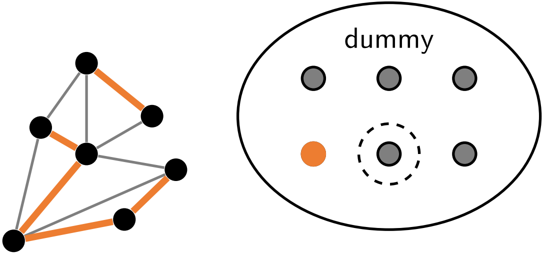


Down-up walk on $\bar{\mu}^{\uparrow}$



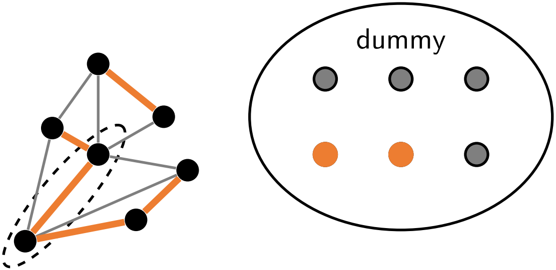
Remove random black dummy node

Down-up walk on $\bar{\mu}^{\uparrow}$



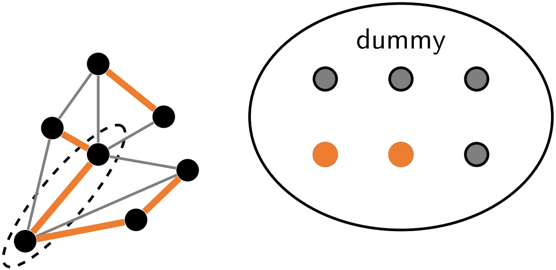
Add random orange dummy node

Down-up walk on $\bar{\mu}^{\uparrow}$



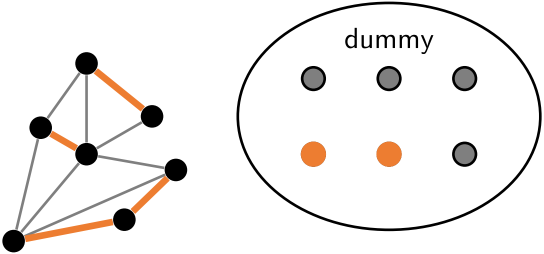
Add random black edge from forest

Down-up walk on $\bar{\mu}^{\uparrow}$



Remove random orange edge from forest

Down-up walk on $\bar{\mu}^{\uparrow}$



Sampling forests

Start with a forest $F_0 \in supp(\mu)$, run MCMC for $O(n \log n)$ steps:

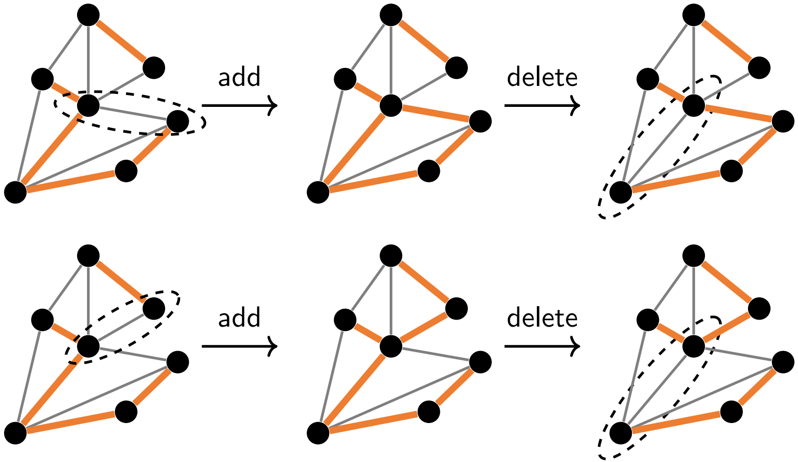
- With probability $1 - \frac{F_t}{n}$, sample $e \notin F_t$ uniformly at random and add e to F

Sampling forests

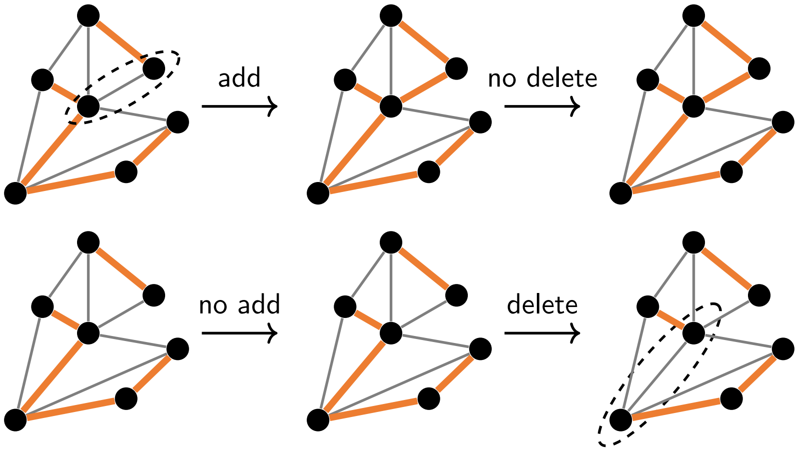
Start with a forest $F_0 \in supp(\mu)$, run MCMC for $O(n \log n)$ steps:

- With probability $1 - \frac{F_t}{n}$, sample $e \notin F_t$ uniformly at random and add e to F
- If there is a cycle C formed in F_t, sample $f \sim C$, update $F \leftarrow F \setminus f$
  Else, with probability $\frac{q}{1+q}$, sample $f \sim F_t$, update $F \leftarrow F \setminus f$

Figure



Figure



Runtime analysis

- Each step takes amortized $O(\log n)$ using link-cut tree [RTF'18]
- Mixing time is $O(n \log n)$ by [CGM19;ALOV**V**21]