Streaming Bayes GFlowNets

Tiago da Silva, Daniel Augusto de Souza, Diego Mesquita

Keywords — GFlowNets, Variational Bayesian inference

I. Background: GFlowNets

GFlowNets are amortized algorithms for sampling from distributions over compositional objects, i.e., over objects that can be sequentially constructed from an initial state through the application of simple actions (e.g., graphs via edge-addition).

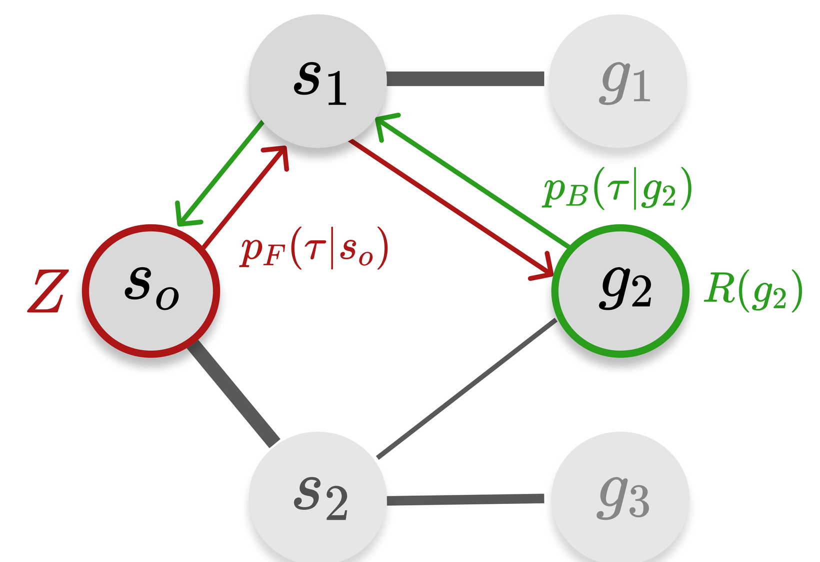


Figure 1: An illustration of the state graph as a DAG on $\mathcal{S}$.

To accomplish this, we learn a forward $p_F(\tau)$ and backward $p_B(\tau|x)$ policies on a state graph (illustrated above) such that

$$Zp_F(\tau) = p_B(\tau|x)R(x), \tag{1}$$

in which R is the unnormalized distribution of interest and Z is its partition function. If this condition is satisfied for each τ,

$$p_\top(x) = \sum_{\tau \rightsquigarrow x} p_F(\tau) \propto R(x), \tag{2}$$

ensuring that the correctness of the generative process.

II. Streaming Bayes GFlowNets

Let $\{R_t\}_{t \geq 1}$ be a sequence of unnormalized distributions. For each T, we train a GFlowNet G_T sampling in proportion to

$$\prod_{1 \leq t \leq T} R_t(x). \tag{3}$$

These GFlowNets approximately satisfy

$$Z_T p_F^{(T)}(\tau) = p_B^{(T)}(\tau|x) \prod_{1 \leq t \leq T} R_t(x), \tag{4}$$

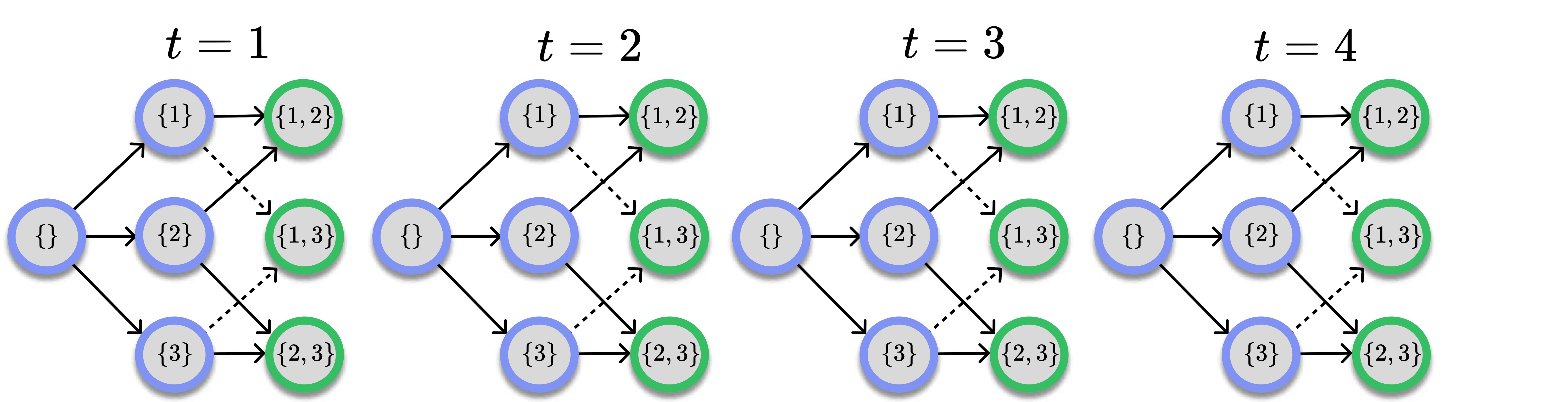
By noticing that

$$Z_T p_F^{(T)}(\tau) = p_B^{(T)}(\tau|x) R_T(x) \underbrace{\prod_{1 \leq t \leq T-1} R_t(x)}_{\approx p_F^{(T-1)}(\tau)/p_B^{(T-1)}(\tau|x)} \tag{5}$$

we train the Tth GFlowNet by minimizing

$$\mathbb{E}_\tau\left[\left(\log \frac{Z_T p_F^{(T)}(\tau)}{p_B^{(T)}(\tau|x) R_T(x)} - \log \frac{p_F^{(T-1)}(\tau)}{p_B^{(T-1)}(\tau|x)}\right)^2\right].$$

We introduce **Streaming Bayes GFlowNets** (SB-GFlowNets) as a general-purpose tool for streaming Bayesian inference over **discrete spaces.** Our model leverages a **GFlowNet** as a surrogate prior when updating the current posterior approximation based on new data, thereby **avoiding to repeatedly process old data** and **significantly accelerating** training convergence in a streaming setting.



$\log R_1^{(\alpha)}(S) = \frac{1}{\alpha} \sum_{d \in S} f_1(d)$ $\quad p_\top^{(1)}(S) \propto R_1^{(\alpha)}(S)$

$\log R_2^{(\alpha)}(S) = \frac{1}{\alpha} \sum_{d \in S} f_2(d)$ $\quad p_\top^{(2)}(S) \propto R_1^{(\alpha)}(S) R_2^{(\alpha)}(S)$

$\log R_3^{(\alpha)}(S) = \frac{1}{\alpha} \sum_{d \in S} f_3(d)$ $\quad p_\top^{(3)}(S) \propto R_1^{(\alpha)}(S) R_2^{(\alpha)}(S) R_3^{(\alpha)}(S)$

$\log R_4^{(\alpha)}(S) = \frac{1}{\alpha} \sum_{d \in S} f_4(d)$ $\quad p_\top^{(4)}(S) \propto R_1^{(\alpha)}(S) R_2^{(\alpha)}(S) R_3^{(\alpha)}(S) R_4^{(\alpha)}(S)$

Figure 2: Streaming amortized inference with SB-GFlowNets.

- **Under which circumstances are SB-GFlowNets useful?**
  1. when $\prod_{1 \leq t \leq T} R_t(x)$ is expensive to compute (e.g., in large-scale Bayesian inference — where each $R_t(x)$ is a likelihood function),
  2. and each GFlowNet is relatively cheap to evaluate (e.g., $p_F(\tau)$ is an MLP or a small GNN — which covers most applications).
- **What factors should we consider when training SB-GFlowNets?**

SB-GFlowNets are amenable to catastrophic error propagation; the accumulated errors should be carefully tracked.

Linear preference learning with integer-valued features.

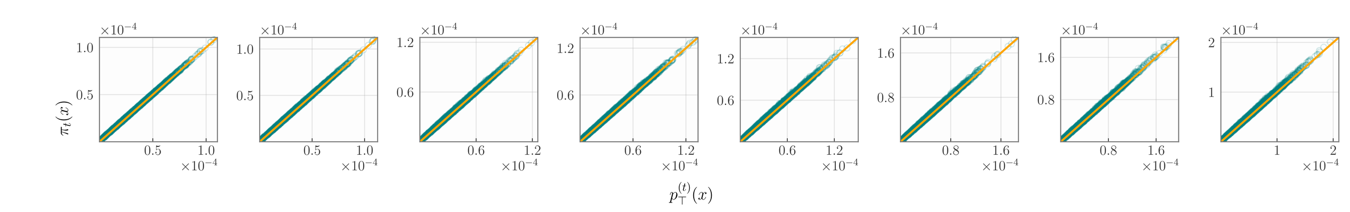


Figure 3: SB-GFlowNets accurately sample from the posterior distribution over the utility in integer-valued preference learning.

Online Bayesian phylogenetic inference.

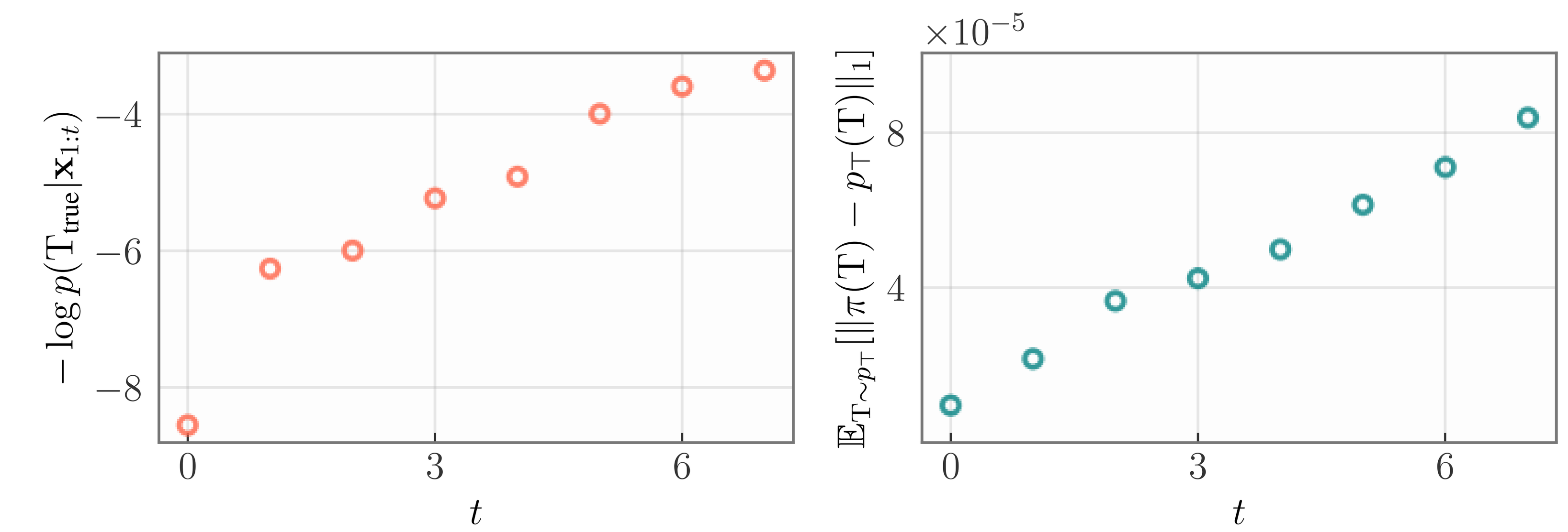


Figure 4: SB-GFlowNet's probability mass associated to the true phylogenetic tree increases as we observe more sequences.

		Model		
		GFlowNet	SB-GFlowNet	% ↑
# of leaves	7	2846.88 s	**1279.68** s	0%
	9	3779.11 s	**1714.49** s	−2%
	11	4821.74 s	**2303.99** s	0%

SB-GFlowNets **achieve faster training convergence** than conventional GFlowNets in a streaming context — while **maintaining a comparable performance** in terms of the TV distance (right column).
(Results averaged across 3 runs.)

{tiago.henrique, diego.mesquita}@fgv.br, daniel.souza.21@ucl.ac.uk