Bayesian Inference with GFlowNets

When do GFlowNets learn the right distribution?

Tiago da Silva[a]

December 5, 2024

Getulio Vargas Foundation (FGV)

[a]Final year Applied Mathematics Ph.D. candidate at FGV.

The sampling problem

Let

1. $\mathcal{X}$ be a set and
2. $R\colon \mathcal{X} \to \mathbb{R}_+$ be a positive function on $\mathcal{X}$.

The sampling problem

How to generate samples from

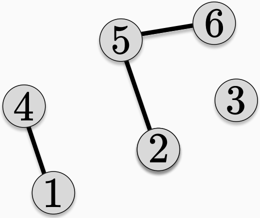
$$\pi(x) \propto R(x)?$$

For Bayesian inference: $R(x) = f(\mathcal{D}|x)p(x)$ (unnormalized posterior)

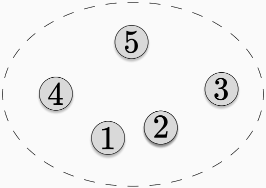
Usual solutions: MCMC, diffusion models, normalizing flows, etc.

Generative Flow Networks (GFlowNets)

GFlowNets solve the sampling problem for a **compositional space**.



Graphs



Sets

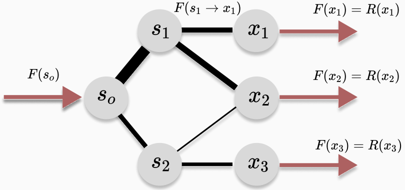
A natural language sentence has a compositional structure

Sentences

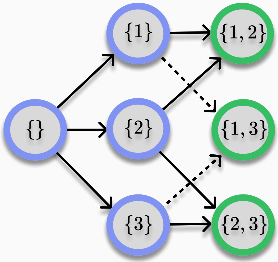
Compositional space

A space $\mathcal{X}$ is **compositional** if its elements can be sequentially constructed from a single **initial state**.

Generative Flow Networks (GFlowNets)



State graph (SG)

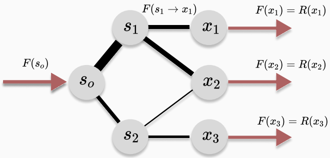


SG for generating 2-sized subsets of $\{1, 2, 3\}$

Central idea

Reframe the state graph as a **flow network** and the sampling problem as a **flow assignment problem**.

GFlowNets: Learning a flow



State graph (SG)

1. Incoming flow = outgoing flow.

$$\sum_{s' \in \mathrm{Ch}(s)} F(s \to s') = \sum_{s' \in \mathrm{Pa}(s)} F(s' \to s).$$

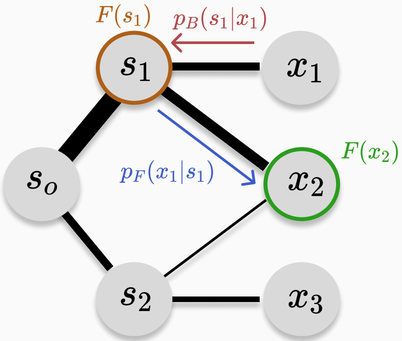
2. Terminal flow = $R(x)$.

$$\sum_{s' \in \mathrm{Pa}(x)} F(s' \to x) = R(x).$$

An ideal setup for Machine Learning

We have a clear goal (finding F) with a simple set of constraints.

GFlowNets: Learning a flow



1. We reparameterize

$$F(s \to s') = F(s) \quad \underbrace{p_F(s'|s)}_{\text{Markovian kernel}}$$

$$F(s' \to s) = F(s') \quad \underbrace{p_B(s|s')}_{\text{Markovian kernel}}$$

2. The constraints become

$$F(s)p_F(s'|s) = F(s')p_B(s|s')$$

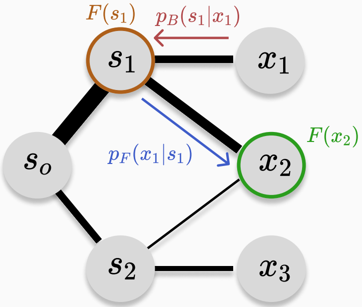
and $F(x) = R(x)$.

Transform the hard constraints into loss functions

$$\mathcal{L}_{DB}(s, s') = \left(\log \frac{F(s)p_F(s'|s)}{F(s')p_B(s|s')}\right)^2$$

This is the **detailed balance** loss. We hardcode $F := R$ for $x \in \mathcal{X}$.

GFlowNets: Learning a flow with neural networks



1. We define

$$F(s) = \underbrace{\mathrm{NN_F}}_{\text{Neural network}}(s)$$

2. Also,

$$p_F(s'|s) = \mathrm{Softmax}(\mathrm{NN_{PF}}(s))[s'],$$

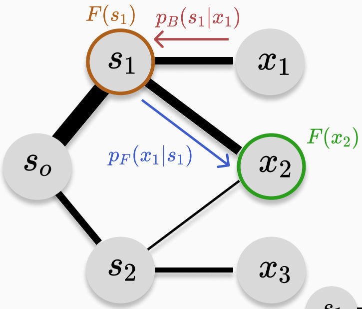
$$p_B(s|s') = \mathrm{Softmax}(\mathrm{NN_{PB}}(s'))[s].$$

A stochastic objective for learning the neural networks

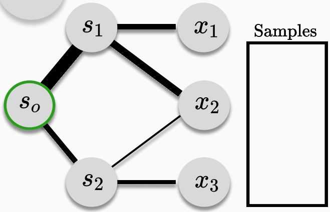
$$\mathcal{L}(F, p_F, p_B) = \mathbb{E}_{\tau\sim p_E}\left[\frac{1}{\#\tau}\sum_{(s,s')\in\tau}\mathcal{L}_{DB}(s,s')\right]$$

p_E is a policy. τ is a trajectory starting at s_o and finishing on $\mathcal{X}$ (e.g., x_2), and $\#\tau$ represents the number of transitions in τ.

GFlowNets: What should you keep in mind?

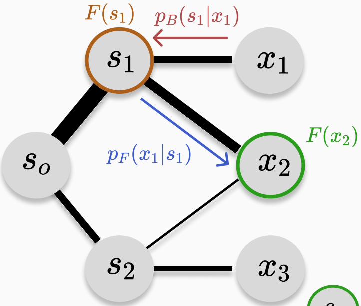


$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$

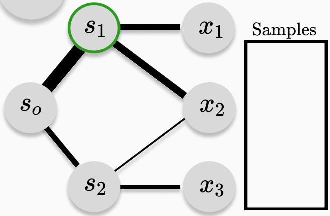


Sampling from ${p_F}^\star$

GFlowNets: What should you keep in mind?

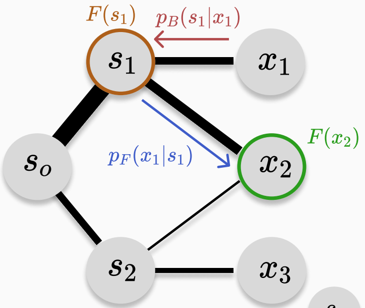


$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$

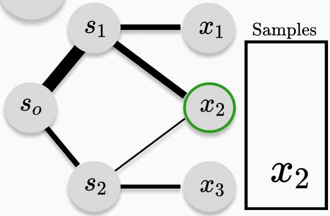


Sampling from ${p_F}^\star$

GFlowNets: What should you keep in mind?

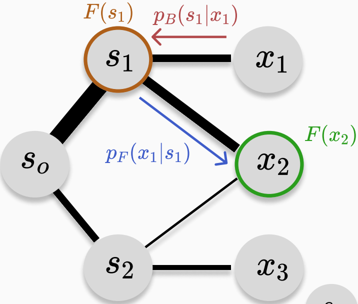


$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$

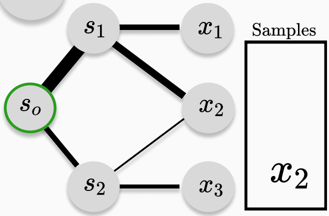


Sampling from ${p_F}^\star$

GFlowNets: What should you keep in mind?

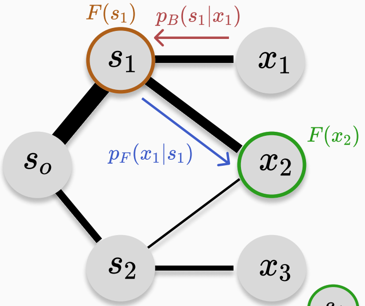


$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$

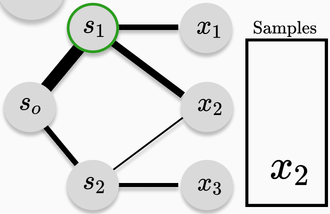


Sampling from ${p_F}^\star$

GFlowNets: What should you keep in mind?

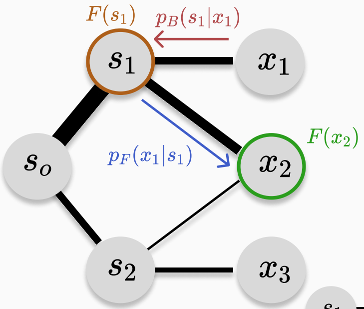


$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$

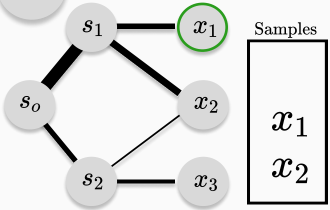


Sampling from ${p_F}^\star$

GFlowNets: What should you keep in mind?



$$F^\star, {p_F}^\star, {p_B}^\star = \underbrace{\arg\min_{F, p_F, p_B} \mathcal{L}_{\mathrm{DB}}(F, p_F, p_B)}_{\text{Stochastic gradient descent}}$$



Independent samples with **non-asymptotic** guarantees ($\neq$ MCMC)!

What do GFlowNets promise?

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2023, 2, 557

GFlowNets for AI-driven scientific discovery

Moksh Jain,[*ab] Tristan Deleu,[ab] Jason Hartford,[ab] Cheng-Hao Liu,[cb] Alex Hernandez-Garcia[ab] and Yoshua Bengio[abd]

Review

Scientific discovery in the age of artificial intelligence

https://doi.org/10.1038/s41586-023-06221-2

Received: 30 March 2022

Accepted: 16 May 2023

Published online: 2 August 2023

Check for updates

Hanchen Wang[1,2,37,38,39], Tianfan Fu[3,39], Yuanqi Du[4,39], Wenhao Gao[5], Kexin Huang[6], Ziming Liu[7], Payal Chandak[8], Shengchao Liu[9,10], Peter Van Katwyk[11,12], Andreea Deac[9,10], Anima Anandkumar[2,13], Karianne Bergen[11,12], Carla P. Gomes[4], Shirley Ho[14,15,16,17], Pushmeet Kohli[18], Joan Lasenby[1], Jure Leskovec[6], Tie-Yan Liu[19], Arjun Manrai[20], Debora Marks[21,22], Bharath Ramsundar[23], Le Song[24,25], Jimeng Sun[26], Jian Tang[9,27,28], Petar Veličković[18,29], Max Welling[30,31], Linfeng Zhang[32,33], Connor W. Coley[5,34], Yoshua Bengio[9,10] & Marinka Zitnik[20,22,35,36] ✉

Bayesian Structure Learning with Generative Flow Networks

Tristan Deleu[1] António Góis[1] Chris Emezue[2,*] Mansi Rankawat[1]
Simon Lacoste-Julien[1,4] Stefan Bauer[3,5] Yoshua Bengio[1,4,6]

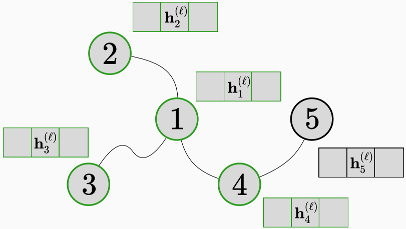
When do GFlowNets learn the right distribution?

What are the limits of GFlowNets?

How to assess a trained GFlowNet?

What are the limits of GFlowNets? A review of GNNs

GFlowNets use Graph Neural Networks (GNNs) to parameterize the flow functions for graph-structured objects. We review what a GNN is below.



1. Aggregate the the features of each node's neighbors.

$$\mathbf{h}_n^{(\ell),\text{agg}} = \sum_{m \in \underbrace{\mathcal{N}(n)}_{\text{Neighborhood of } n} \cup \{n\}} \mathbf{h}_m^{(\ell)}.$$

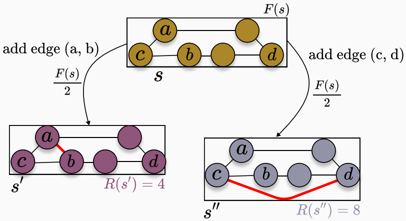
2. Update the node's feature.

$$\mathbf{h}_n^{(\ell+1)} = \mathrm{NN}_{\mathrm{G}}(\mathbf{h}_n^{(\ell),\text{agg}}).$$

Repeat this a few times.

Result: $\mathbf{H} = \left[\mathbf{h}_n^{(L)}\right]_n \in \mathbb{R}^{\#\text{nodes} \times d}$ used as input for downstream tasks.

What are the limits of GNN-based GFlowNets?

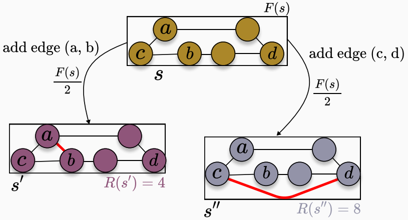


1. GNNs cannot distinguish edges (a, b) and (c, d) in s.
2. A GFlowNet always assigns the same flow to s's children — regardless of R.

GNN-based GFlowNets are fundamentally limited

The limited capacity of a GNN **constrains the expressivity** of the GNN-parameterized GFlowNet.

What are the limits of GNN-based GFlowNets?



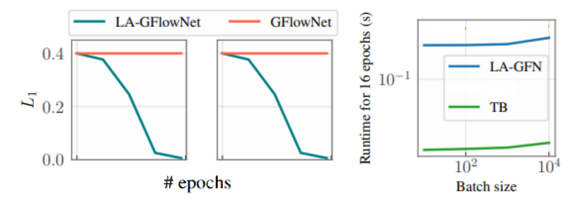
Solution: Look-Ahead (LA) GFlowNets.

$$p_F(s'|s) \propto \exp\left(\mathbf{w}^T \left[\phi_{(a,b)}^{(s')} || \phi_{(a,b)}^{(s)}\right]\right),$$

$$\phi^{(s)} = \text{GNN}(s),\ \phi^{(s')} = \text{GNN}(s').$$

LA-GFlowNets boost the expressiveness GNN-based GFlowNets

LA-GFlowNets incorporate **future embeddings** into the **current policy** to compute the transition probabilities.

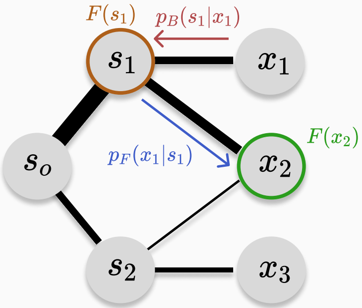


When do GFlowNets learn the right distribution?

What are the limits of GFlowNets?

How to assess a trained GFlowNet?

How to assess a trained GFlowNet?



$$\underbrace{p_\top(x)}_{\text{Marginal over } \mathcal{X}} := \sum_{\tau \rightsquigarrow x} p_F(\tau)$$

$$= \underbrace{\mathbb{E}_{\tau \sim p_B(\tau|x)}\left[\frac{p_F(\tau)}{p_B(\tau|x)}\right]}_{\text{Importance sampling}}$$

$$= \frac{1}{T} \sum_{1 \le i \le T} \frac{p_F(\tau_i)}{p_B(p_B|x)},$$

$$\{\tau_1, \dots, \tau_T\} \sim p_B(\cdot|x).$$

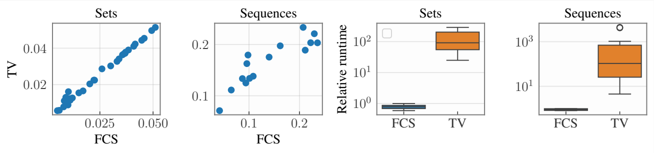
Flow Consistency in Subgraphs (FCS)

FCS consists of the expected total variation in random $\mathcal{X}$-subsets.

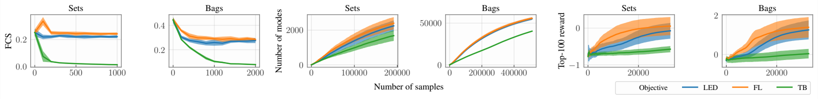
$$\mathbb{E}_{S(\subseteq \mathcal{X}) \sim P}\left[\mathrm{TV}(p_\top^{(S)}, R^{(S)}\right] = \frac{1}{2}\mathbb{E}_{S(\subseteq \mathcal{X}) \sim P}\left[\sum_{x \in S}\left|p_\top^{(S)}(x) - R^{(S)}(x)\right|\right],$$

with $R^{(S)}(x) = {}^{R(x)}/_{\sum_{y \in S} R(y)}$ and $p_\top^{(S)}(x) = {}^{p_\top(x)}/_{\sum_{y \in S} p_\top(y)}$.

How to assess a trained GFlowNet?



FCS is highly correlated and drastically faster-to-compute than TV.



Mode-coverage is not a reliable indicator of correctness.

> **FCS is a realiable measurement of GFlowNet's accuracy**
>
> FCS provides a relatively **fast-to-compute** and **accurate** surrogate for the distributional correctness of a GFlowNet.

Conclusions

Take-home message

GFlowNets are excellent samplers for compositional spaces.

GFlowNets cast the sampling problem as a flow assignment problem in a flow network. This network describes the sequential construction of complex objects from a single initial state.

GNN-based GFlowNets have fundamentally limited expressiveness.

GNN-based flow functions explicitly constraint the range of distributions realizable by the corresponding GFlowNet. Designing efficient solutions to this problem remains an open issue.

FCS has the potential to standardize the evaluation of GFlowNets.

As such, FCS stands as the best available diagnostic for this family of models. It is a computationally amenable and theoretically sound proxy for the distributional accuracy of GFlowNets.

Questions?

Thank you!