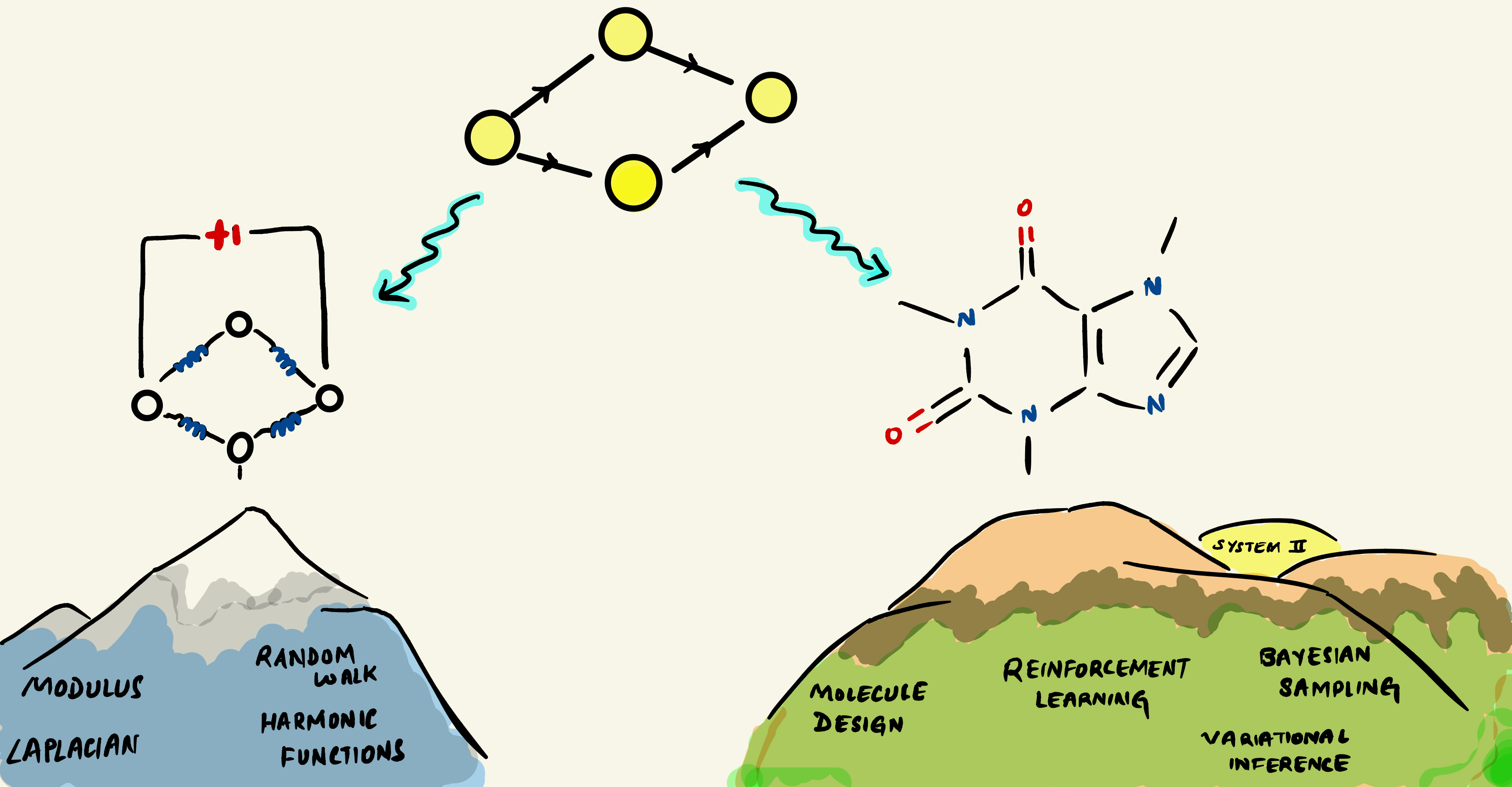
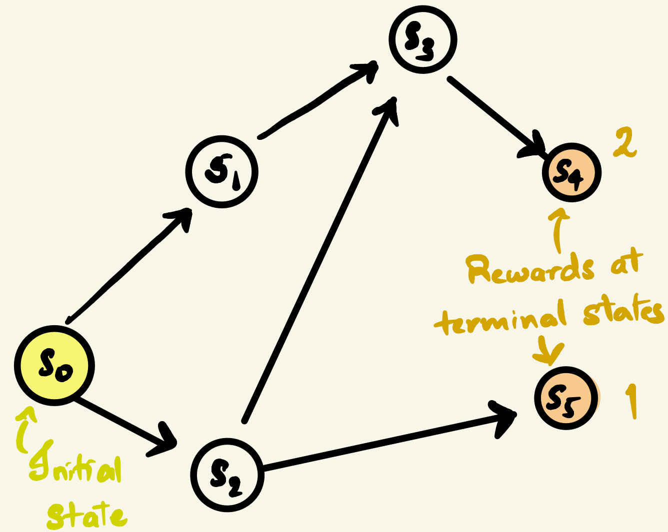
(G)ENERATIVE (F)LOW (NET)WORKS



SETUP

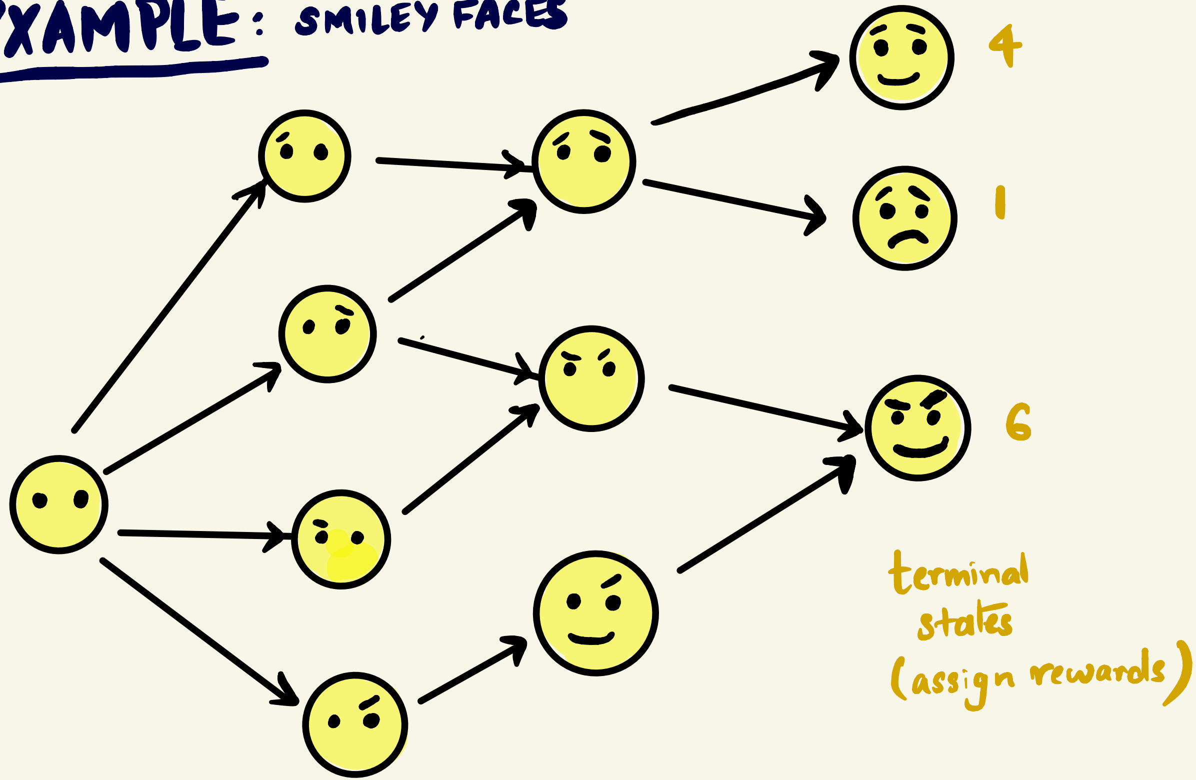
Directed acylic graph $G = (S, \mathbb{A})$, where S is a finite set of states, and $\mathbb{A}$ is a subset of $S \times S$ representing directed edges called actions/transitions.



<u>Goal:</u> Learn a stochastic policy for generating objects from a sequence of actions such that the probability of generating an object is proportional to the given reward at that object.

$$\boxed{P(x) \propto R(x)}$$

EXAMPLE: SMILEY FACES

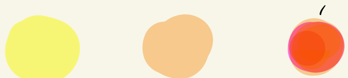


EXAMPLE: SUBGREEDY WALK

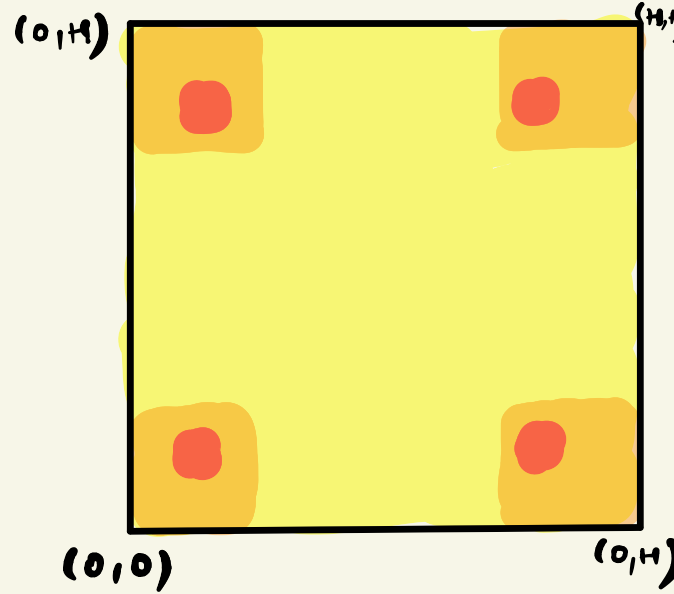
- You start at (0,0).
- Only allowed to increase coordinate. (This is the action a_i).
- Can terminate at any state s and get the reward $r(s)$.
- Reward

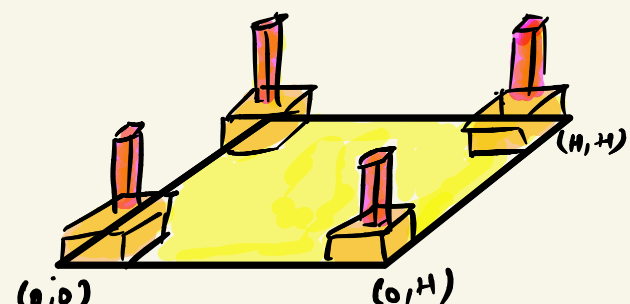
$$R(\vec{x}) = R_0 + R_1 \prod_{i=1}^{2} \mathbb{1}\left\{0.2 < \left|\frac{x_i}{H} - 0.5\right|\right\}$$

$$+ R_2 \prod_{i=1}^{2} \mathbb{1}\left\{0.3 < \left|\frac{x_i}{H} - 0.5\right| < 0.4\right\}$$

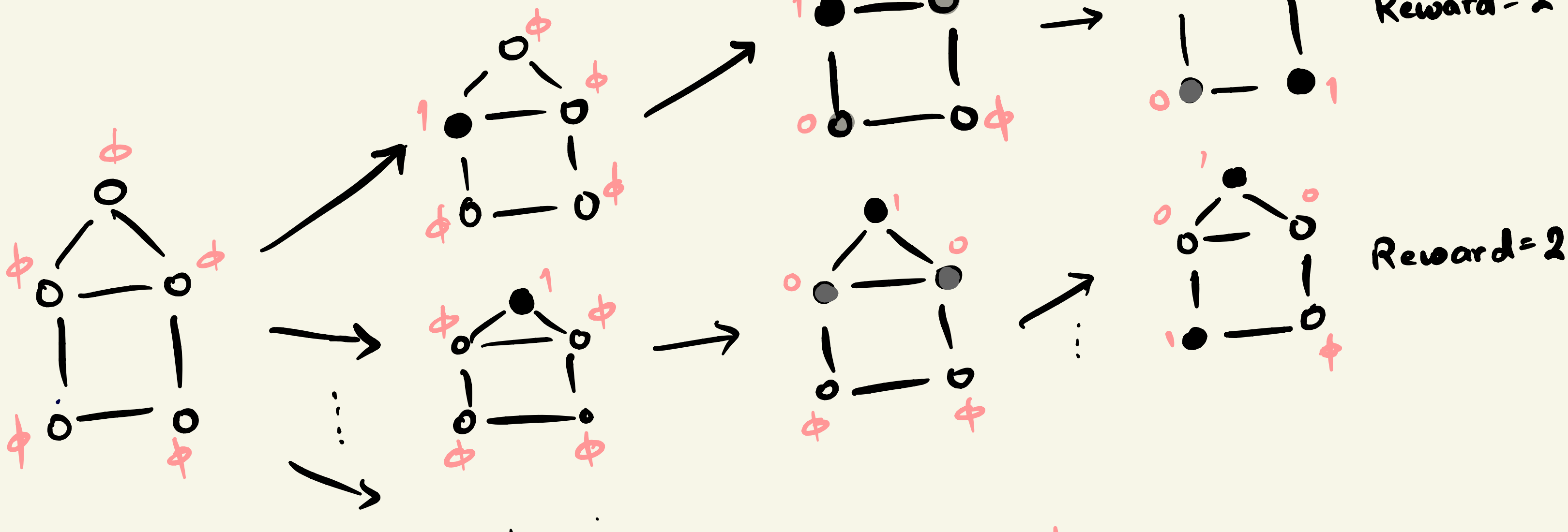


$(0 < R_0 < R_1 < R_2)$

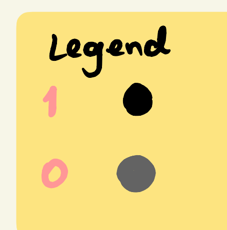


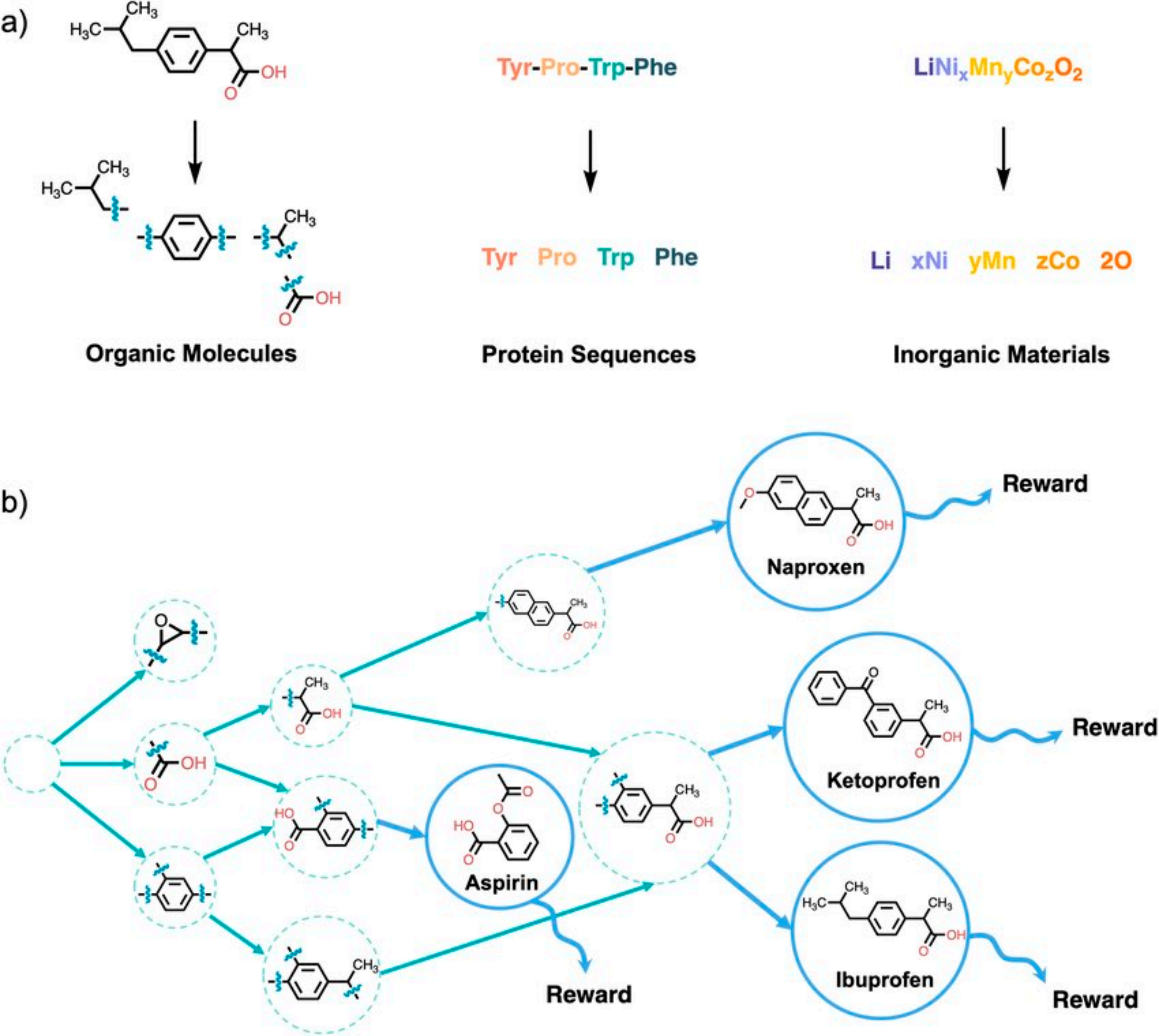


EXAMPLE: MAXIMUM INDEPENDENT SET



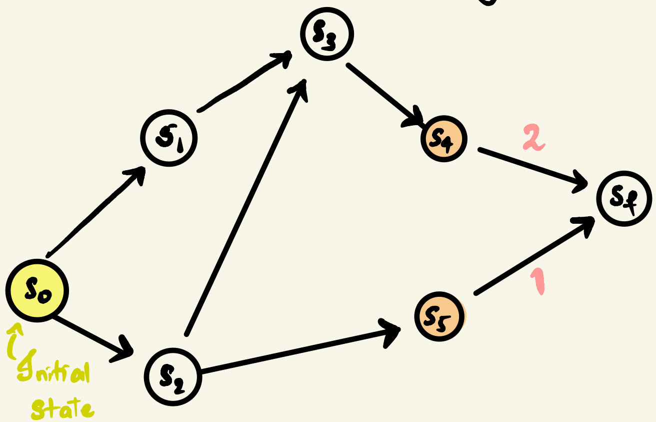
- Start with a graph G. Label all the vertices ϕ.
- At each stage update the node label with 1 or 0; 1 represents include in the independent set; 0 represents exclude.
- Terminate when all nodes are assigned 0 or 1.




a)
Tyr-Pro-Trp-Phe
LiNixMnyCozO2
Tyr Pro Trp Phe
Li xNi yMn zCo 2O
Organic Molecules
Protein Sequences
Inorganic Materials
b)
Naproxen
Reward
Ketoprofen
Reward
Aspirin
Reward
Ibuprofen
Reward

GFLOWNETS

- First we introduce an auxillary final state s_f to encode the rewards.



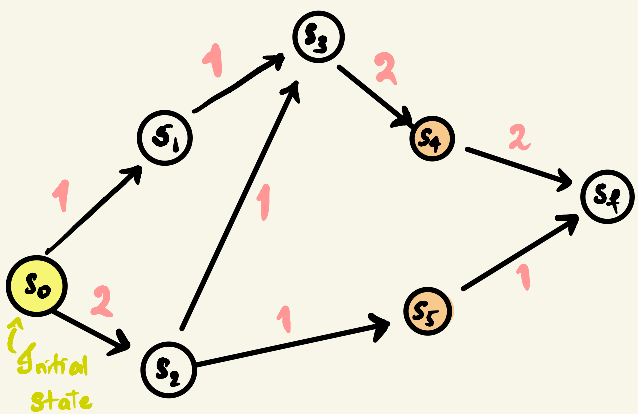
- Then assign a forward probability

$$P_F(s \to s') = \frac{F(s \to s')}{F(s)}$$

- Then s_4 will be sampled with probability $2/3$ and s_5 with probability $1/3$.

Idea:

- Learn a flow.



Theorem. Define a policy π that generates trajectories starting at s_0 by sampling actions $a \in \mathcal{A}(s)$ according to

$$\pi(a|s) = \frac{F(s,a)}{F(s)}$$

($F(s,a)$ is flow through (s,a)
$F(s)$ is flow through s)

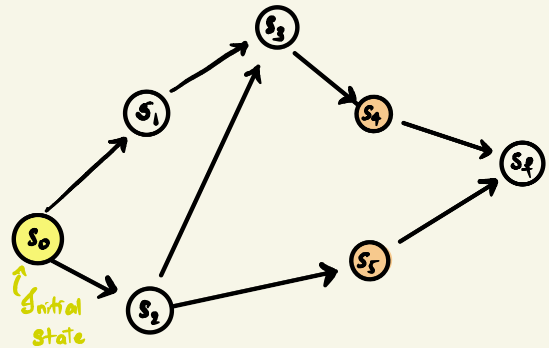
Let $\pi(s)$ denote the probability of visiting state s when starting at s_0 and following $\pi(\cdot|\cdot)$. Then

a) $\pi(s) = \dfrac{F(s)}{F(s_0)}$

b) $F(s_0) = \sum_{x \in \mathcal{X}} R(x)$ ($\mathcal{X}$ denotes terminal states)

c) $\pi(x) = \dfrac{R(x)}{\sum_{x' \in \mathcal{X}} R(x')}$

This was the goal.



Proof a) Note $\pi(s_0)=1$, because we always start from s_0.

For the children of s_0, say s, we have $\pi(s) = \dfrac{F(s_0 \to s)}{F(s_0)} = \dfrac{F(s)}{F(s_0)}$

Induction: Assume the statement is true for all the parents of a state s'.

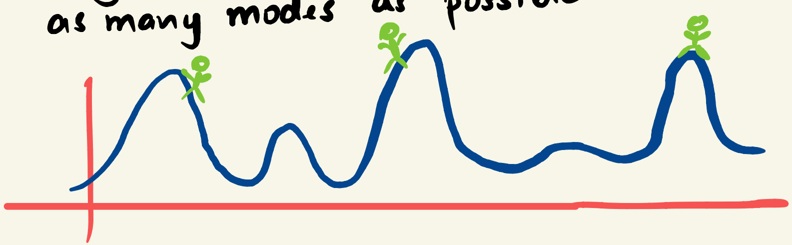
$$\pi(s') = \sum_{\substack{\text{parents} \\ s \text{ of } s'}} \frac{F(s)}{F(s_0)} \cdot \frac{F(s \to s')}{F(s)} = \frac{\sum_{\substack{\text{parents} \\ s \text{ of } s'}} F(s \to s')}{F(s_0)} = \frac{F(s')}{F(s_0)}$$

For terminal states $x \in \mathcal{X}$, $\pi(x) = \dfrac{R(x)}{F(s_0)}$

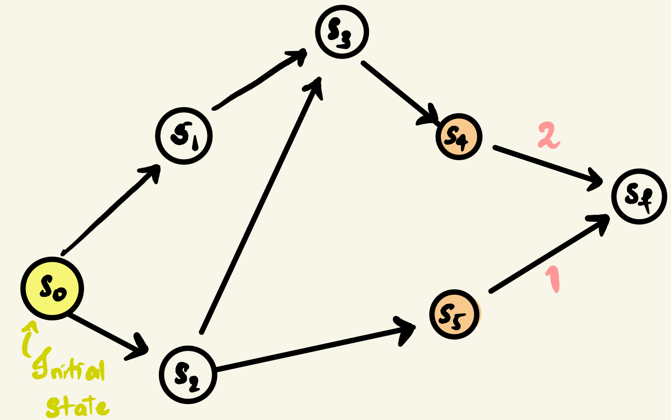
Since outflow at s_0 = Inflow at s_f, $\pi(x) = \dfrac{R(x)}{\sum_{x' \in \mathcal{X}} R(x')}$

HOW TO TRAIN GFLOWNETS?

- For small number of states, no problem learning a flow.
- For large state spaces,
  - * we don't have immediate access to rewards
  - * reaching all the terminal states is not feasible
- Maybe a more accesible goal is to learn as many modes as possible

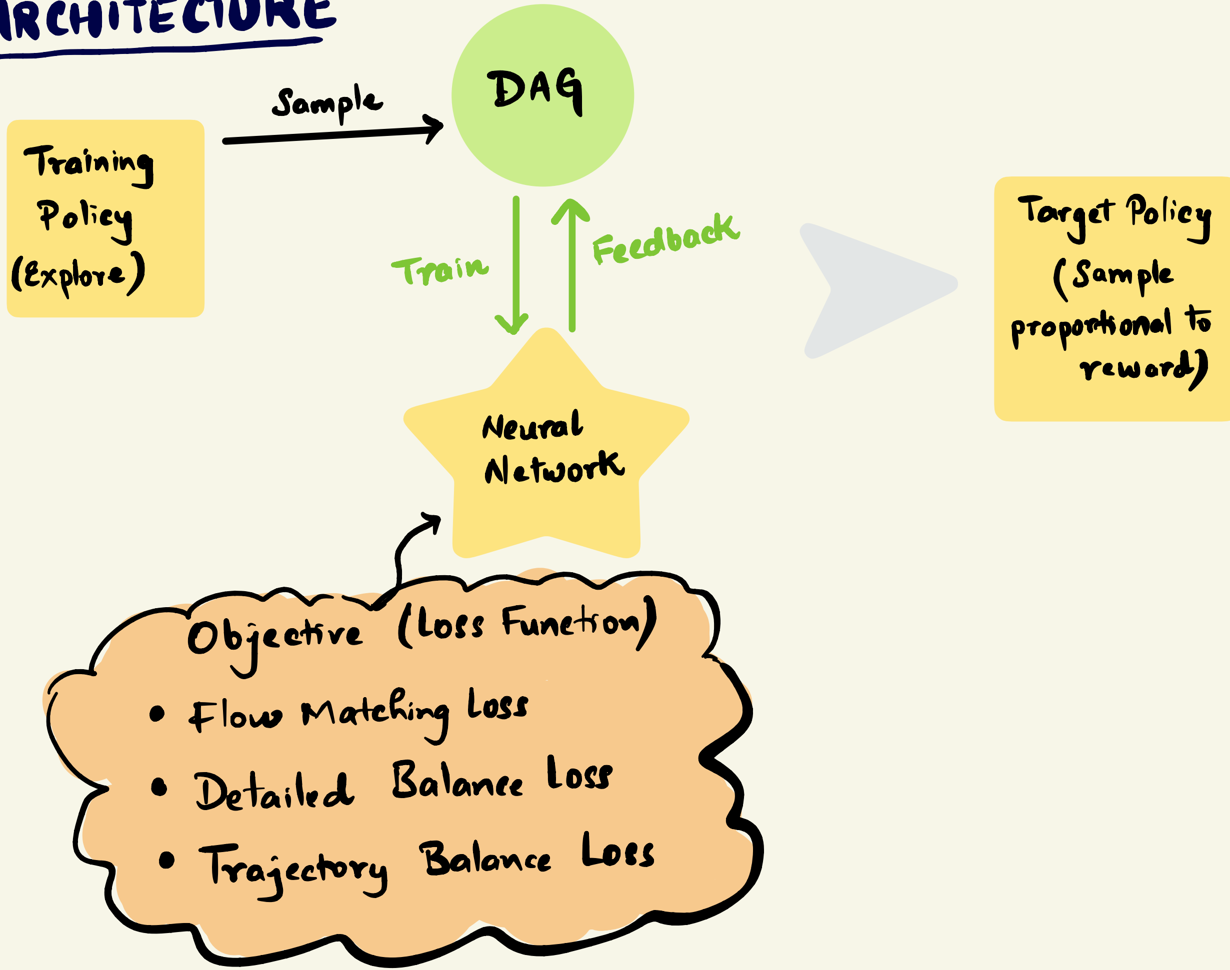


REWARD DISTRIBUTION



PROBLEMS IN
- MOLECULE DISCOVERY
- COMBINATORIAL OPTIMIZATION
- GENERATIVE MODELS

ARCHITECTURE

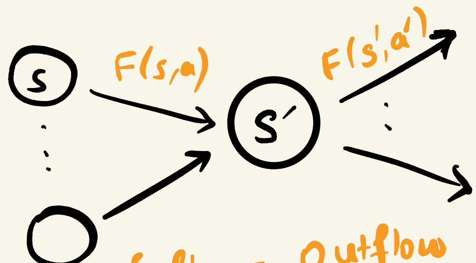


FLOW MATCHING LOSS

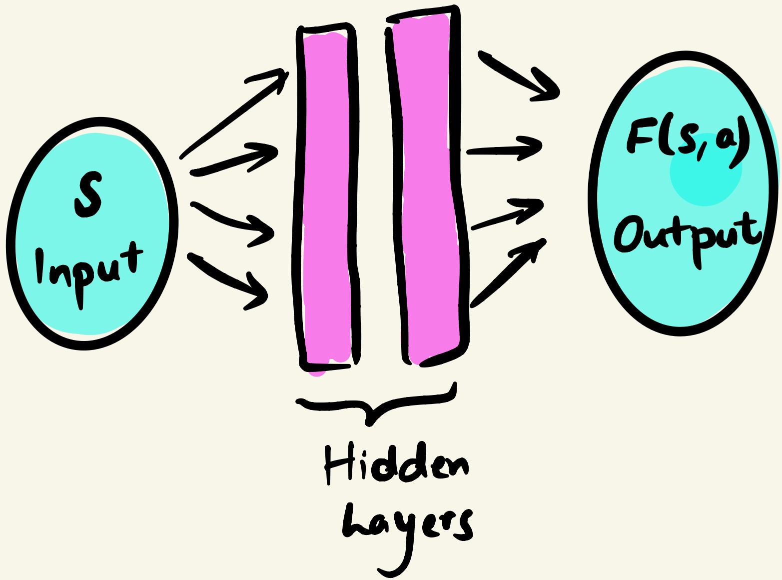
$$\tilde{\mathcal{L}}_{\theta}(\tau) = \sum_{\substack{s' \in \tau \neq s_0 \\ \neq s_f}} \left(\sum_{\substack{s:\, s \text{ is parent of } s' \\ \text{and } T(s,a) = s'}} F_{\theta}(s,a) - \sum_{\substack{a':\, a' \text{ is an action} \\ \text{from } s'}} F_{\theta}(s',a') \right)^2$$

τ is a trajectory, $s_0 \rightarrow \ldots \rightarrow s_f$

θ are the parameters of NN.



Inflow = Outflow



LIMITATION:

- Since the flows for nodes near the root are supposed to be high, compared to the nodes near the leaves, the loss will be high compared to the nodes near the leaves. So gradient updates will be imbalanced.

IMPROVEMENT:

log scale

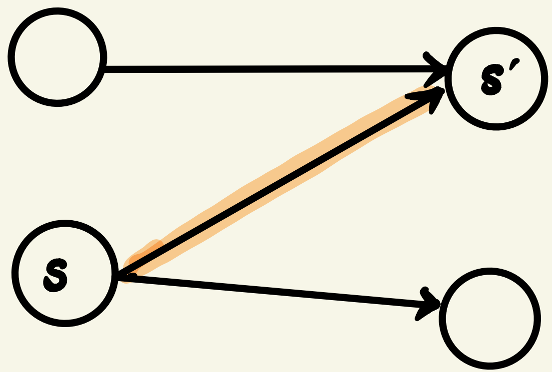
$$\mathcal{L}_{\theta,\epsilon}(\tau) = \sum_{s' \in \tau \neq s_0, s_f} \left(\log\left[\epsilon + \sum_{s,a : T(s,a)=s'} \exp F_\theta^{\log}(s,a)\right] - \log\left[\epsilon + \sum_{a' \in A(s')} \exp F_\theta^{\log}(s',a')\right]\right)^2$$

$\log(\epsilon + \text{inflow}) \quad - \log(\epsilon + \text{outflow})$

- gradient updates are similar for large and small flows.
- ϵ is added to avoid log of small numbers and also to give more weight to errors on large flows.

DETAILED BALANCE LOSS

- Learn
  - * forward policy $P_F(\cdot|s)$ for each s
  - * state flow $F(s)$ for each s
  - * Backward policy $P_B(\cdot|s)$ for noninitial s



Define

$$P_F(v|s) = \frac{F(s\to v)}{\sum_{v':(s,v')\in A} F(s\to v')}$$

$$P_B(u|s) = \frac{F(u\to s)}{\sum_{u':(u',s)\in A} F(u'\to s)}$$

Note

For $\nearrow$, $F(s\to s') = F(s)\,P_F(s'|s) = F(s')\,P_B(s|s')$

Detailed Balance Objective:

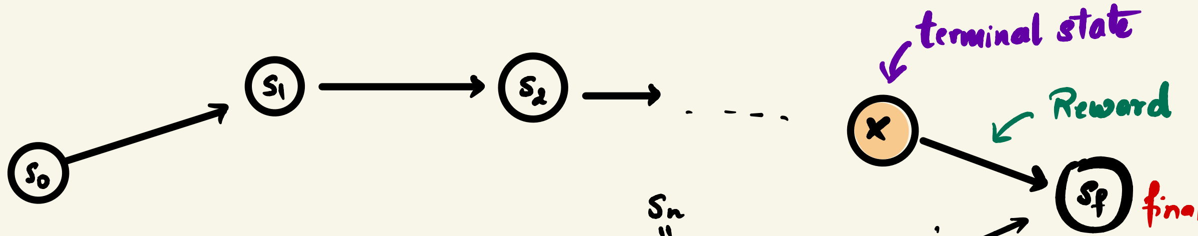
$$\mathcal{L}_{DB}(s \to s') = F_\theta(s) P_F^\theta(s'|s) - F_\theta(s') P_B^\theta(s|s')$$

IMPROVEMENT

$$\mathcal{L}_{DB}(s \to s') = \left[\log\left(\frac{F_\theta(s) P_F^\theta(s'|s)}{F_\theta(s') P_B^\theta(s|s')}\right)\right]^2$$

- solves the problem of summing over large namber of parents in the flow matching loss.
- Slow credit assigment because it is local.

TRAJECTORY BALANCE LOSS



For a trajectory $\tau = s_0 \to s_1 \to s_2 \to \cdots \to \underset{\overset{\shortparallel}{s_n}}{x} \to s_f$

the probability of sampling τ using $P_F(\cdot)$ is

$$P_F(s_1|s_0) \cdot P_F(s_2|s_1) \cdots P_F(s_{n-1}|x)$$

the probability of sampling τ using $P_B(\cdot)$ is

$$\frac{R(x)}{F(s_0)} \cdot P_B(s_{n-1}|x) \cdots P_B(s_0|s_1)$$

Thus,

$$F(s_0)\, P_F(s_1|s_0)\, P_F(s_2|s_1) \cdots P_F(s_{n-1}|x) = R(x)\, P_B(s_{n-1}|x) \cdots P_B(s_0|s_1)$$

TRAJECTORY BALANCE OBJECTIVE:

$$\mathcal{L}_{TB}(\tau) = \left(\log \frac{F_\theta(s_0) \prod_{i=1}^{n} P_F^\theta(s_i \mid s_{i-1})}{R(x) \prod_{i=1}^{n} P_B^\theta(s_{i-1} \mid s_i)} \right)^2 = \left(\log \frac{F_\theta(s_0)\, P_F^\theta(\tau)}{R(x)\, P_B^\theta(\tau \mid x)} \right)^2$$

Theorem. If $\mathcal{L}_{TB}(\tau) = 0$ for all τ, then P_F samples proportional to reward.

Advantages:

- global objective
- faster credit assignment

REFERENCES

- E. Bengio et al. *Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation.* NEURIPS 2021.
- Y. Bengio et al. *GFlownet Foundations.* JOURNAL OF MACHINE LEARNING RESEARCH 24 (2023) 1-76.
- Zhang et al. *Let the Flows Tell: Solving Graph Optimization Problems with GFlowNets.* NEURIPS 2023.
- Mila GFlowNet Workshop 2023. gflownet.org