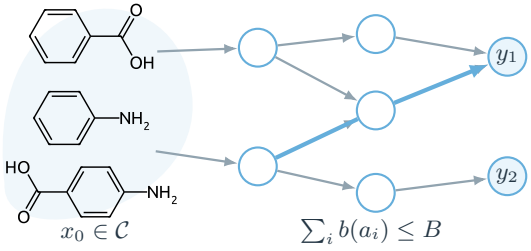
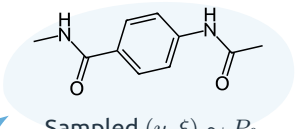
Parents, admissible actions, budget B, outcome weights r_z, route temperature τ



$$v(s) = \log \sum_{e:u \to s} e^{v(u) - c(e)/\tau}$$

$$q(e \mid s) \propto e^{v(u) - c(e)/\tau}$$

$$\log Z + \log P_\theta = \log r_z + \log Q_\phi$$



Sampled $(y, \xi) \sim P_\theta$

$x_0 \xrightarrow{a_1} x_1 \xrightarrow{a_2} y$

DooABLe

$$P^*(\xi) = \underbrace{\frac{r_z(y)}{\sum_{y' \in \mathcal{Y}_B} r_z(y')}}_{\text{outcome probability}} \underbrace{\frac{e^{-C(\xi)/\tau}}{\sum_{\xi' : y(\xi') = y} e^{-C(\xi')/\tau}}}_{\text{conditional route probability}}$$