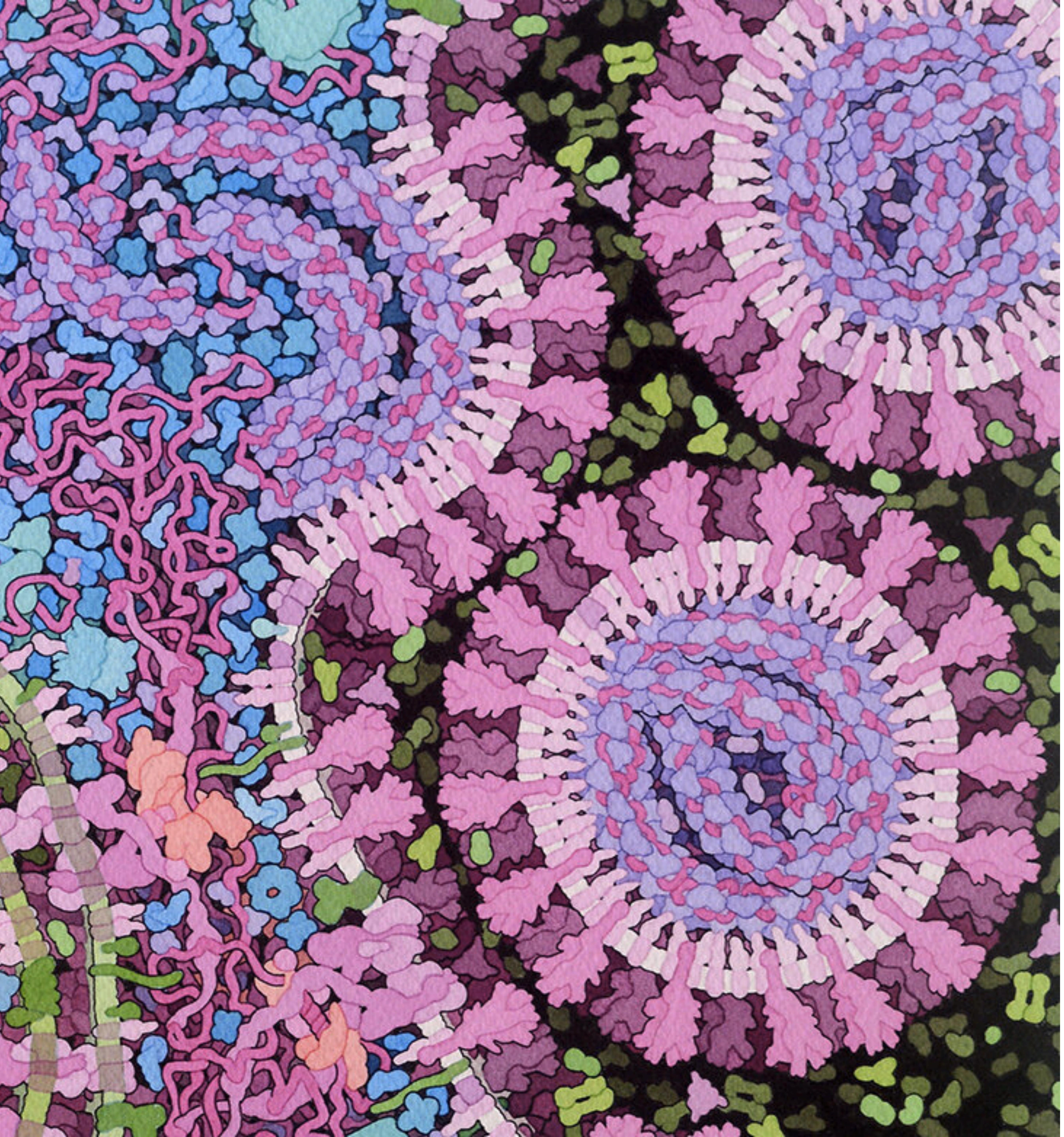


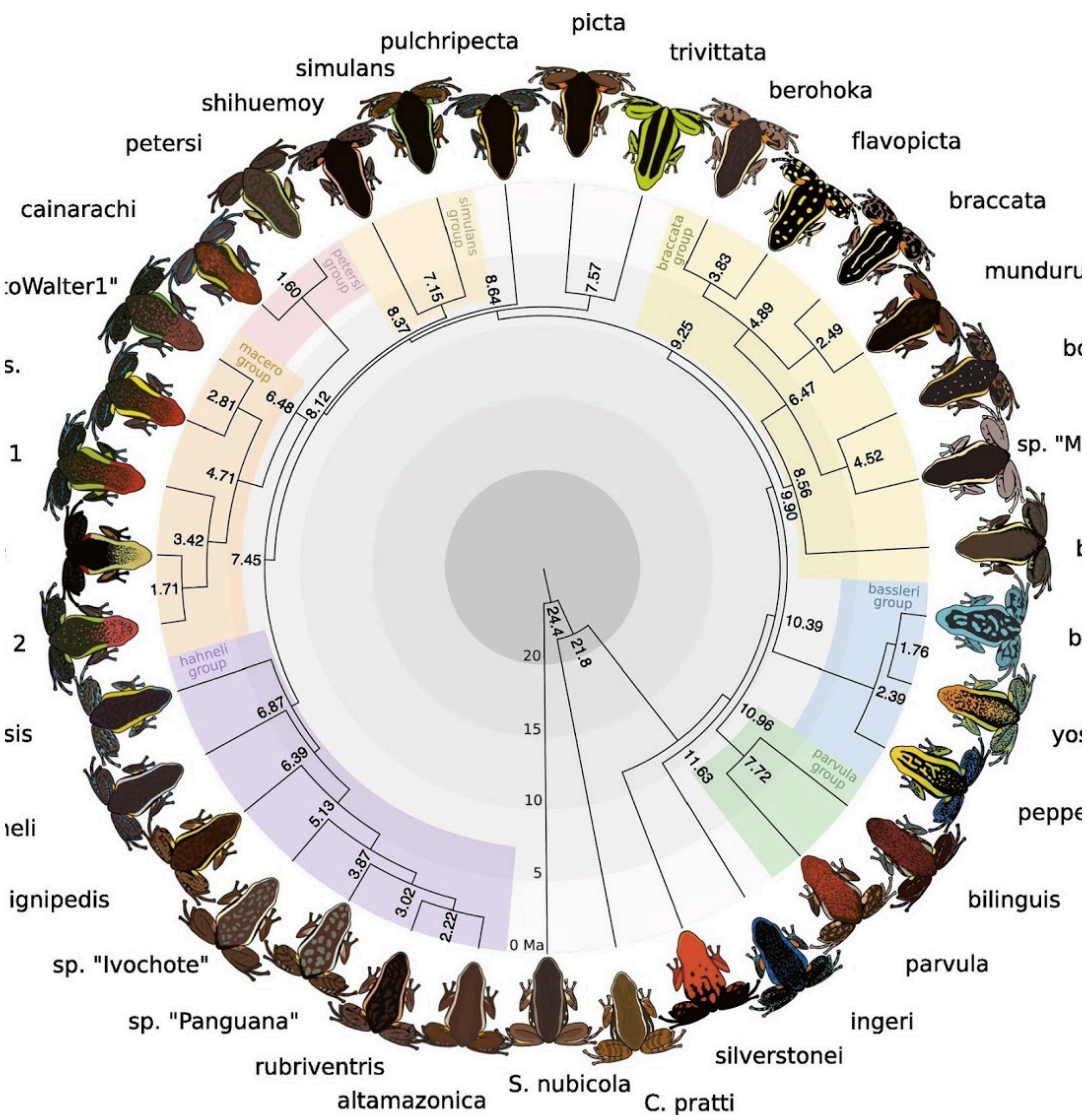
بسم الله الرحمن الرحیم

ژنومیک محاسباتی

جلسه ۵: بازسازی درخت تبارزایی (۲)

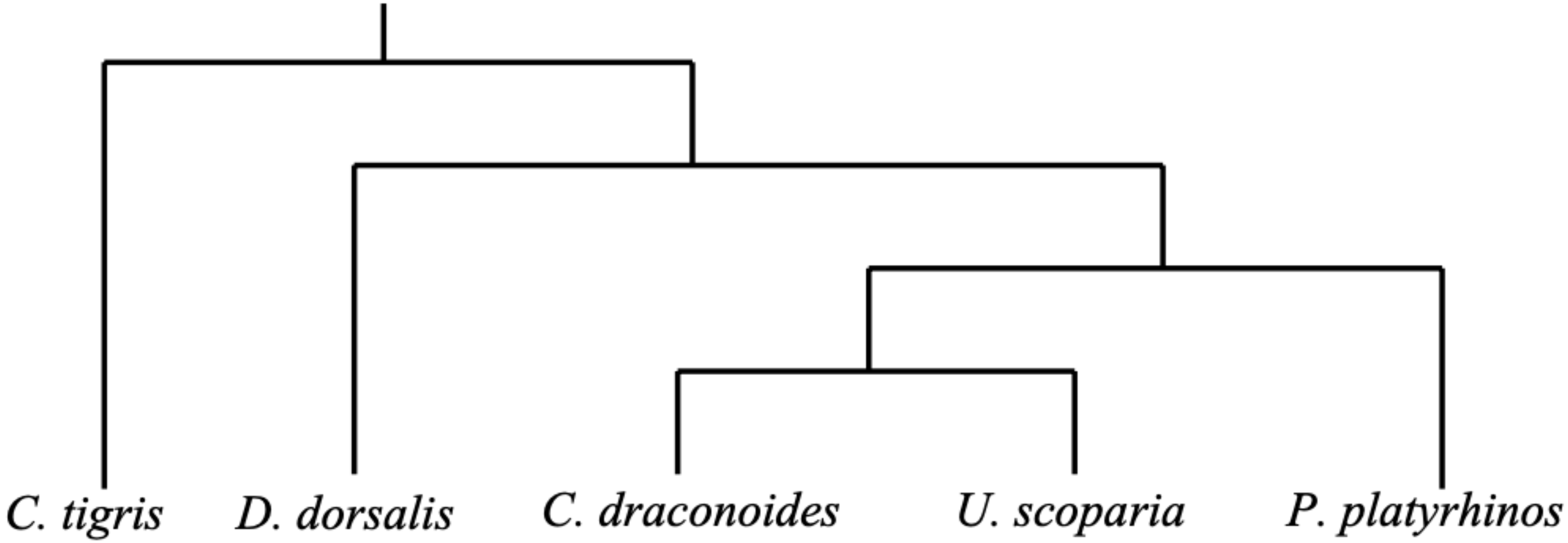
ترم پاییز ۱۴۰۰-۱۴۰۱

مرور



picta
trivittata
berohoka
flavopicta
braccata
pulchripecta
simulans
shihuemoy
petersi
cainarachi
simulans group
petersi group
braccata group
macero group
bassleri group
hahneli group
parvula group
bilinguis
parvula
ingeri
silverstonei
C. pratti
S. nubicola
altamazonica
rubriventris
sp. "Panguana"
sp. "Ivochote"
ignipedis
0 Ma
5
10
15
20

درخت تبارزایی، چیستی و چرایی؟



معیار بهترین درخت، ویژگی_ مبنا

- بیشینه صرفه‌جویی
- یک مدل احتمالاتی
- معیارهای دیگر؟

مسئله صرفه‌جویانه‌ترین درخت

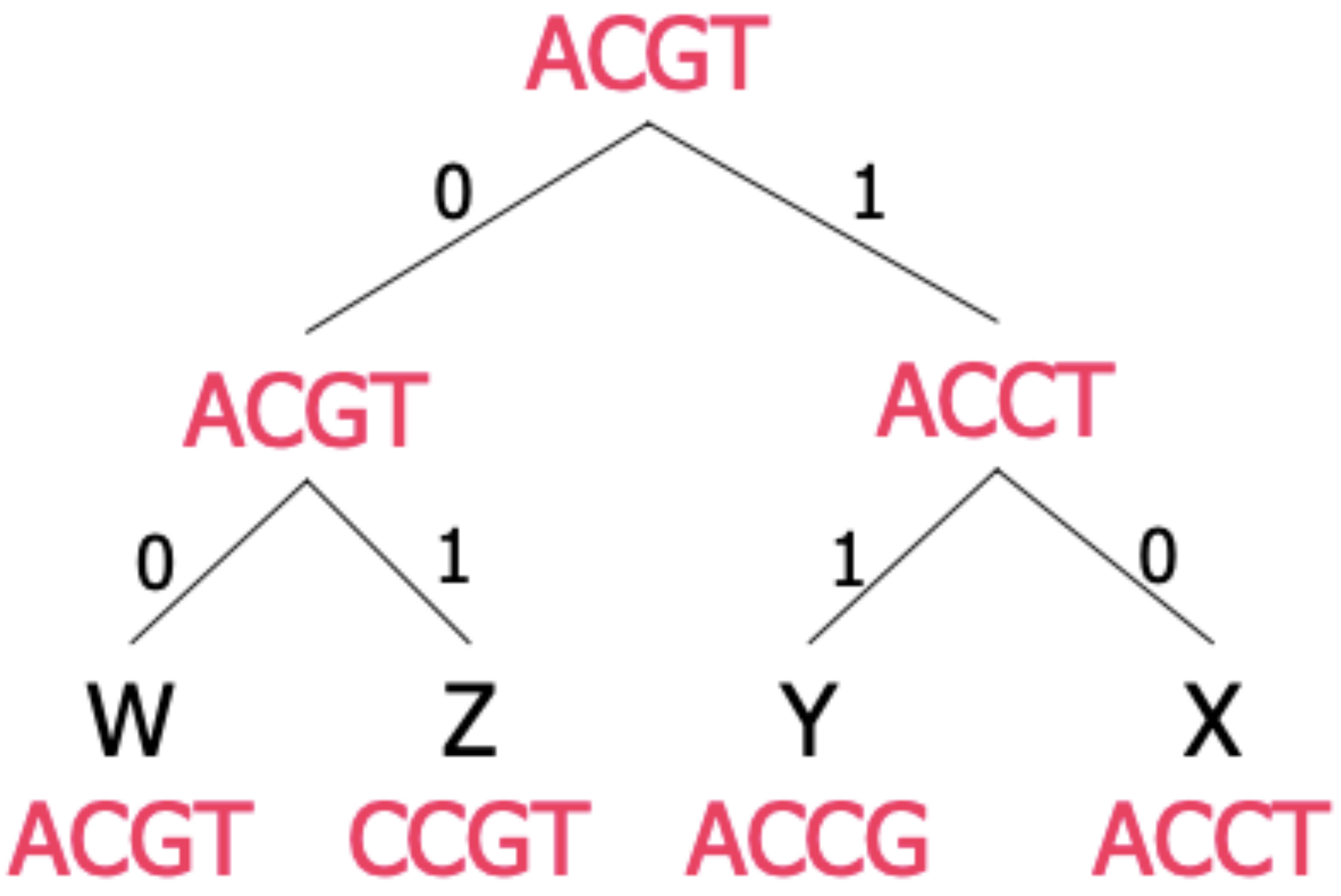
مسئله صرفه‌جویانه‌ترین درخت

	1	2	3	4
W	A	C	G	T
X	A	C	C	T
Y	A	C	C	G
Z	C	C	G	T

مسئله صرفه‌جویانه‌ترین درخت

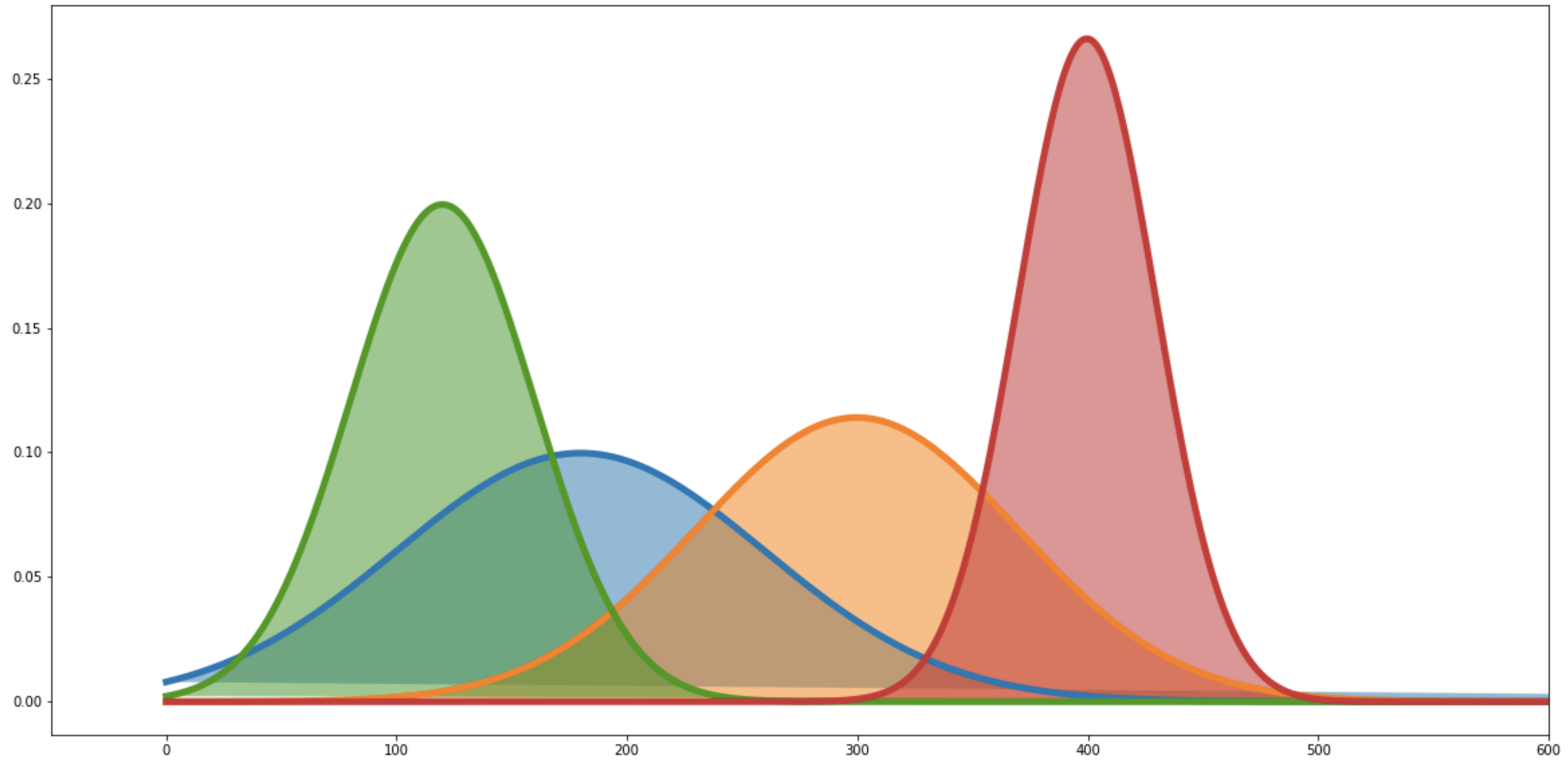
	1	2	3	4
W	A	C	G	T
X	A	C	C	T
Y	A	C	C	G
Z	C	C	G	T

مسئله صرفه‌جویانه‌ترین درخت



	1	2	3	4
W	A	C	G	T
X	A	C	C	T
Y	A	C	C	G
Z	C	C	G	T

برآورد درست‌نمایی بیشینه



درست‌نمایی

ورودی (ویژگی‌ها)

درخت و
پارامترهای درخت

$$L(\theta) := P(D \mid \theta)$$

$$\max P(D \mid \theta)$$

درست‌نمایی
بیشینه

محاسبه درست‌نمایی

- ورودی: ماتریس $M[v,i]$ (ویژگی i برای برگ v)
- تعریف: $L_i(v,s)$ =
  - بیشینه احتمال مشاهده ویژگی i روی برگ‌های زیردرخت v، به شرط اینکه ویژگی i روی راس v برابر با s باشد.
- جواب: $\prod_i \left(\frac{1}{2}L_i(r,0) + \frac{1}{2}L_i(r,1)\right)$
- حالت پایه: برای برگ v: $L_i(v,s) = 1_{s=M[v,i]}$
- رابطه بازگشتی: برای راس میانی u با فرزندان v و w:

$$L_i(u,s) = \left[\sum_{y\in\{0,1\}} L_i(v,y)Pr(v_i = y|u_i = s)\right]\left[\sum_{x\in\{0,1\}} L_i(w,x)Pr(w_i = x|u_i = s)\right]$$

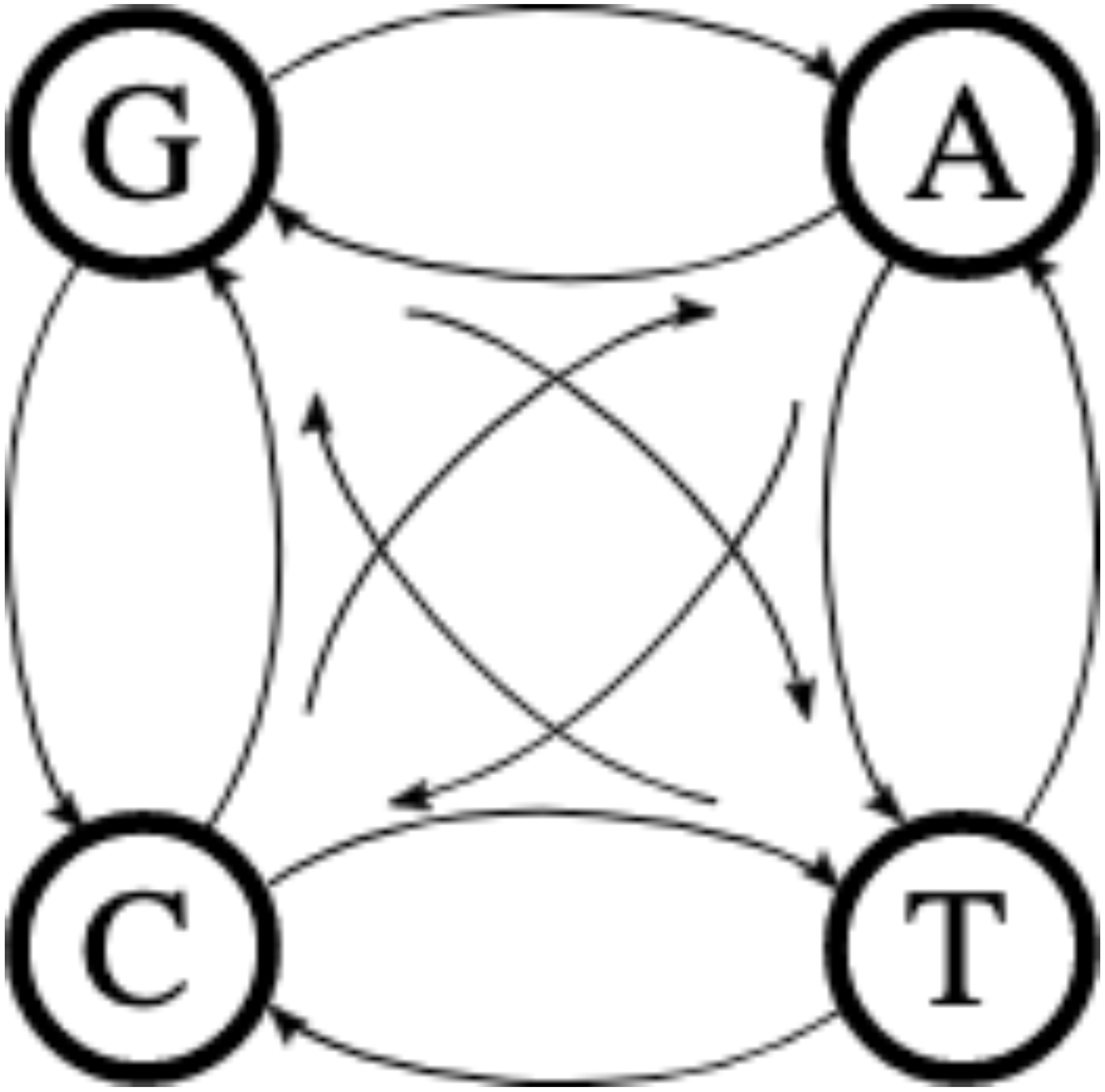
اگر رشته ویژگی‌ها روی هر راس را داشته باشیم:

مسئله بازسازی درخت با درست‌نمایی بیشینه:

$$\max_{p_e,T} P(i_u \forall u \in L(T) \mid \{T, p_e\})$$

- الگوریتم‌های ابتکاری
- هر دفعه
  - یک تغییر روی درخت و روی peها،
  - مشاهده L

مدل تکامل DNA



مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

$$p'(t) = p(t)Q$$

مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

$$p'(t) = p(t)Q$$

جمع سطر = ۰

مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

$$p'(t) = p(t)Q$$

جمع سطر = ۰

$$p(t) = p(0)e^{tQ}$$

مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

$$p'(t) = p(t)Q$$

جمع سطر = ۰

$$p(t) = p(0)e^{tQ}$$

توزیع پایای یکتا

$$\pi e^{tQ} = \pi$$

مدل تکامل مارکوف

احتمال مشاهده حالت‌ها در زمان t: $p(t)$

جمع سطر = ۰

$$p'(t) = p(t)Q$$

$$p(t) = p(0)e^{tQ}$$

توزیع پایای یکتا

$$\pi e^{tQ} = \pi$$

بازگشت‌پذیر

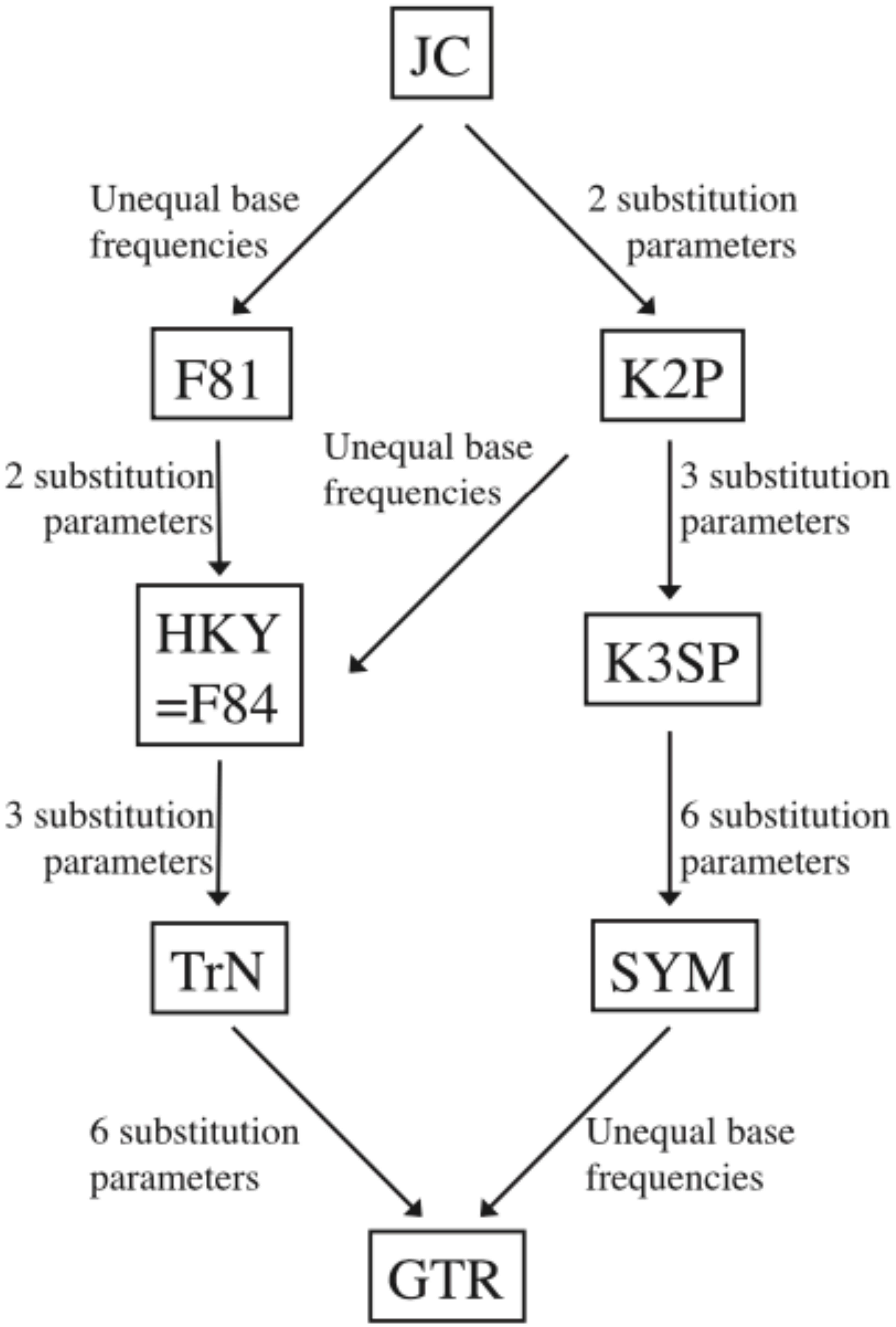
$$xe^{tQ} = y \Leftrightarrow ye^{tQ} = x$$

مدل JC69 (Jukes and Cantor)

- یک پارامتر μ: نرخ جهش در یک واحد زمانی

$$P = \begin{pmatrix} \frac{1}{4}+\frac{3}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} \\ \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}+\frac{3}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} \\ \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}+\frac{3}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} \\ \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}-\frac{1}{4}e^{-t\mu} & \frac{1}{4}+\frac{3}{4}e^{-t\mu} \end{pmatrix}$$

$$Q = \begin{pmatrix} * & \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & * & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & \frac{\mu}{4} & * & \frac{\mu}{4} \\ \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} & * \end{pmatrix}$$


JC
Unequal base frequencies
2 substitution parameters
F81
K2P
2 substitution parameters
Unequal base frequencies
3 substitution parameters
HKY =F84
K3SP
3 substitution parameters
6 substitution parameters
TrN
SYM
6 substitution parameters
Unequal base frequencies
GTR

JC

ماتریس احتمال => محاسبه احتمال (درخت + ویژگی راس‌های میانی)

JC

ماتریس احتمال => محاسبه احتمال (درخت + ویژگی راس‌های میانی)

محاسبه ویژگی راس‌های میانی
با بیش‌ترین احتمال (امتیاز)

=> محاسبه احتمال (درخت)

مسئله

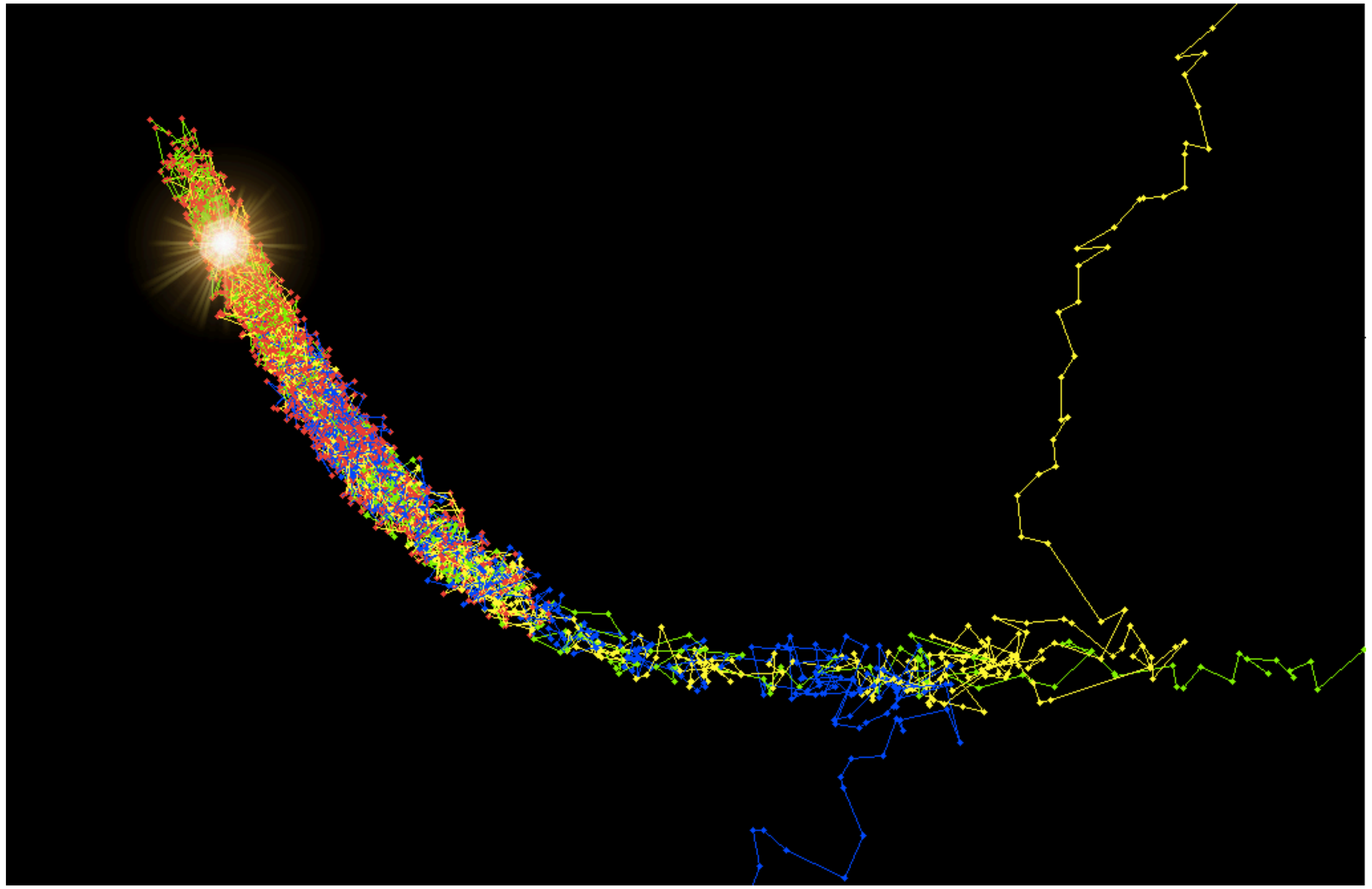
- ورودی:
  - ماتریس ویژگی‌ها

- خروجی:
  - درخت

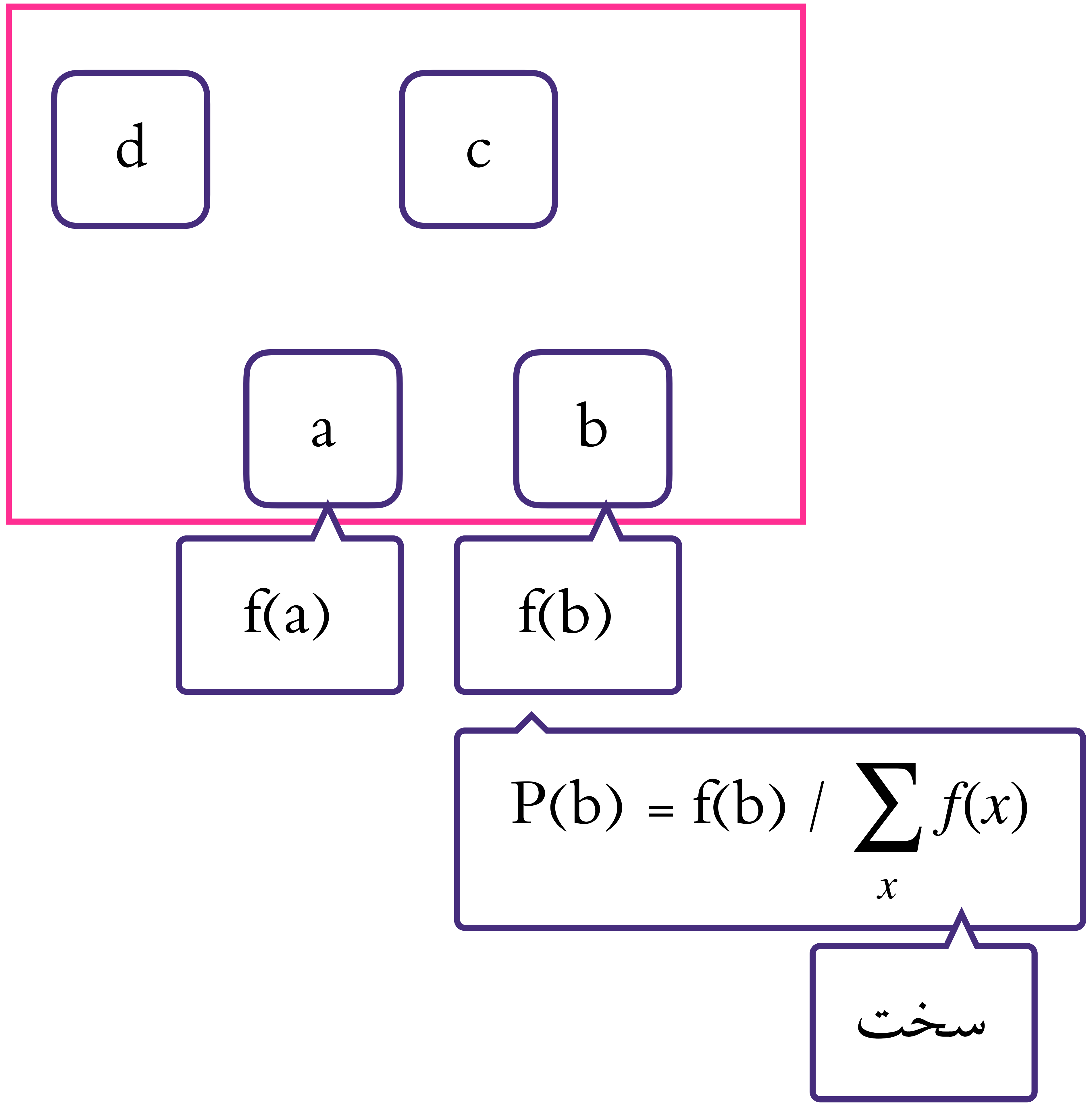
- الگوریتم

؟؟

MCMC



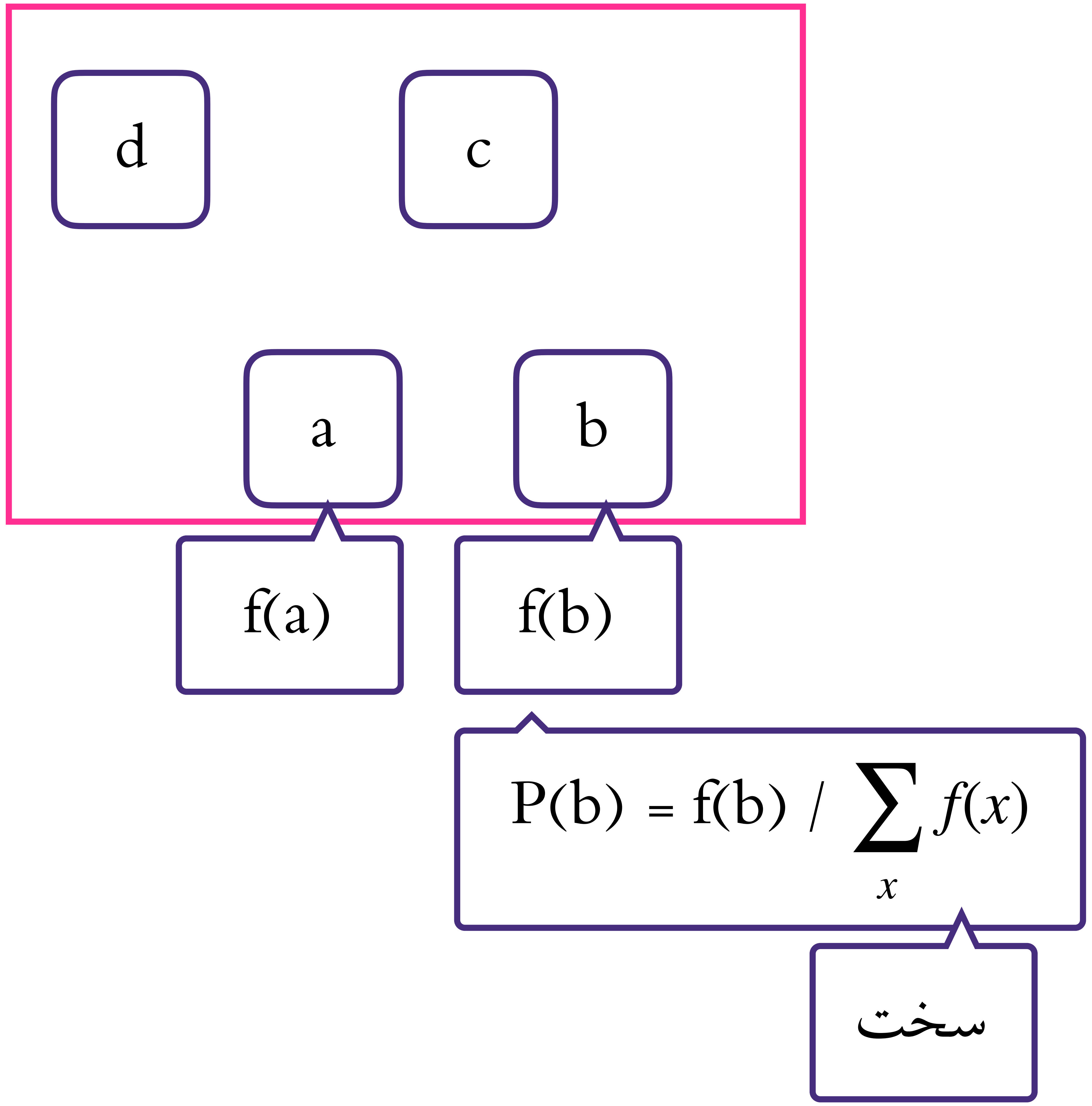
MCMC



MCMC

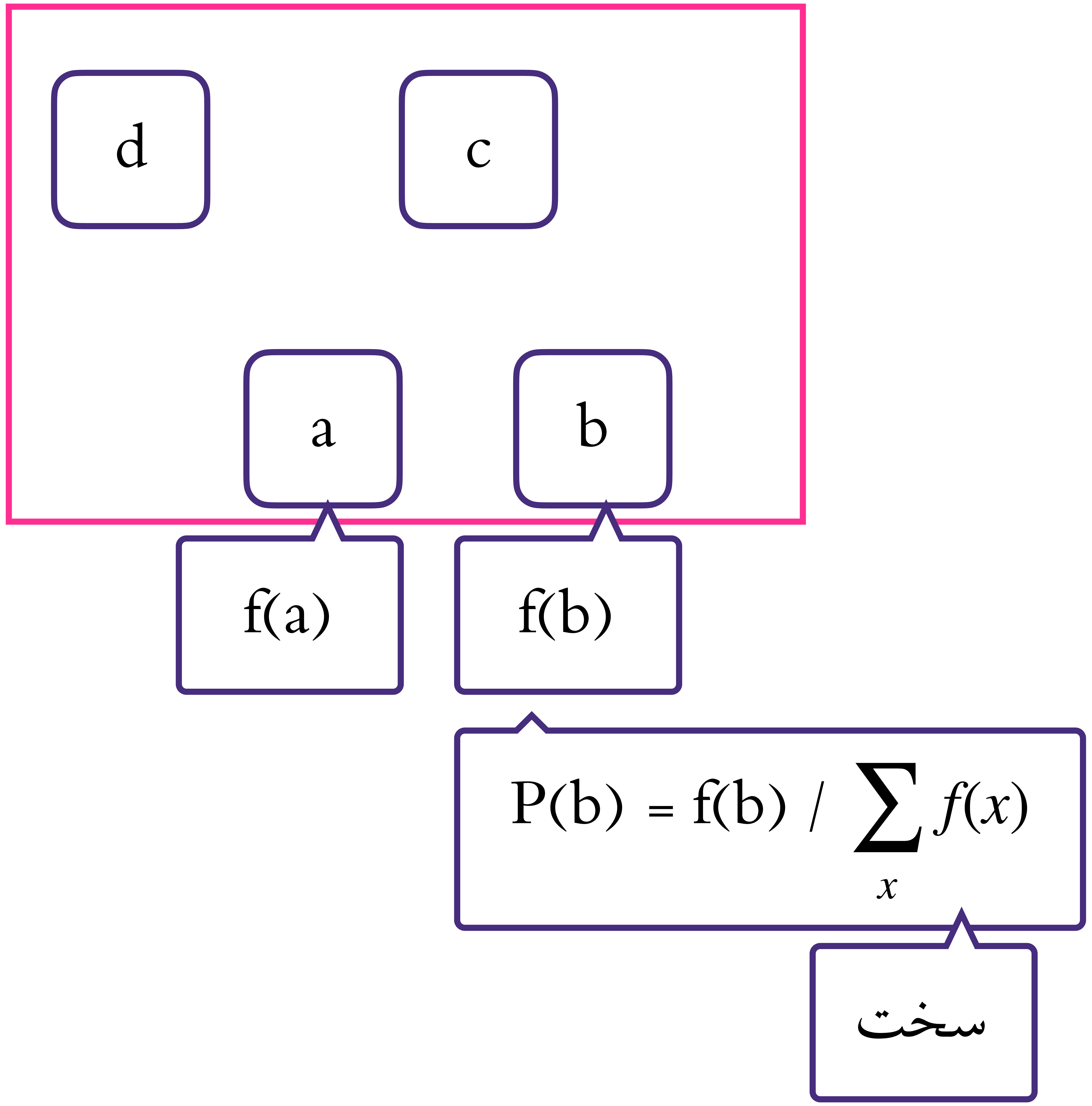
- یک نقطه ابتدایی x0
- هر دفعه
  - یک همسایه از توزیع $g(x' | x_t)$ انتخاب کن
  - با احتمال تغییر را بپذیر

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$



MCMC

- یک نقطه ابتدایی x0
- هر دفعه
  - یک همسایه از توزیع $g(x'|x_t)$ انتخاب کن
  - با احتمال تغییر را بپذیر



$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$

کافی است نسبت احتمال را محاسبه کنیم

MCMC

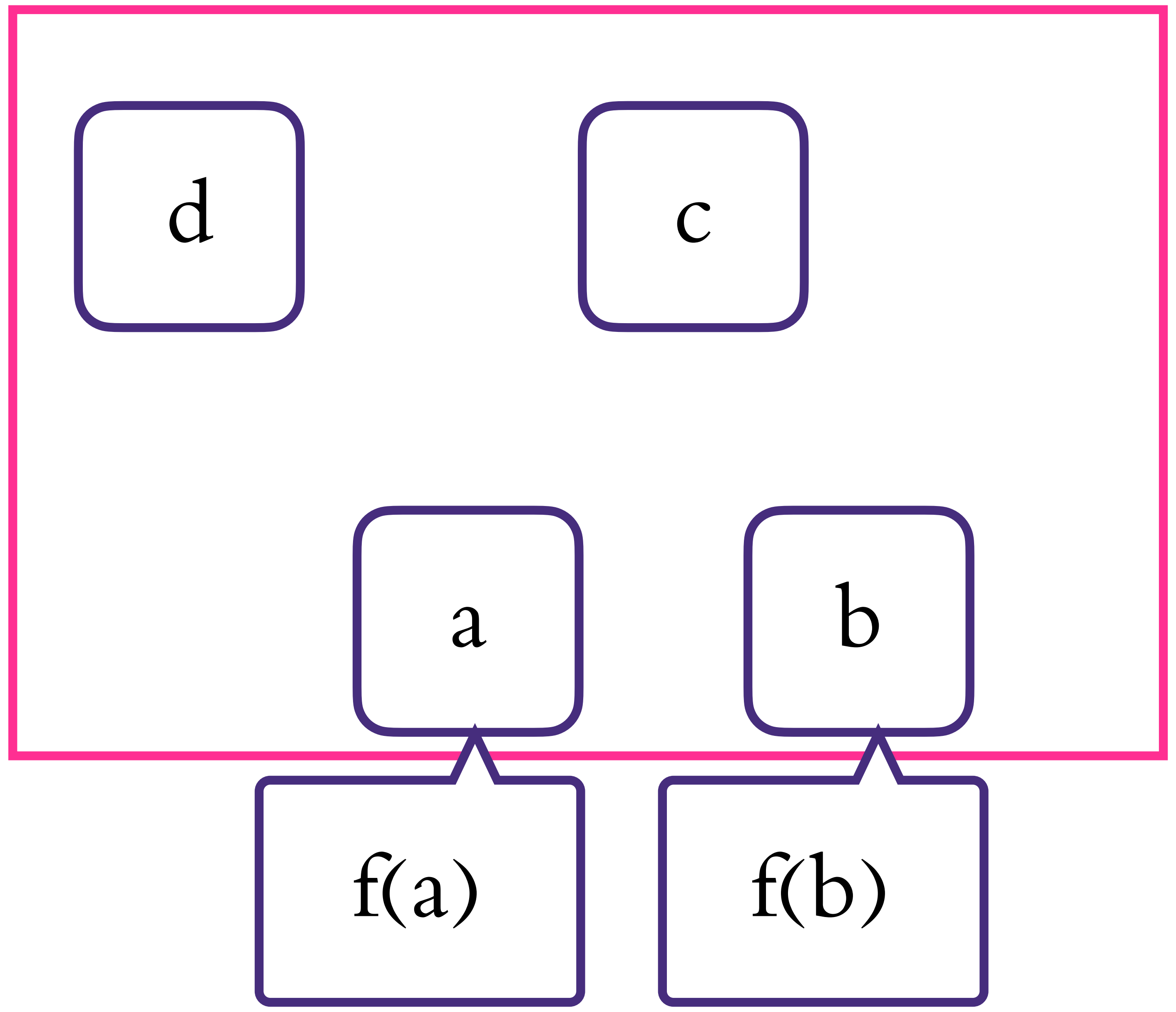
- یک نقطه ابتدایی x0
- هر دفعه
  - یک همسایه از توزیع $g(x' | x_t)$ انتخاب کن
  - با احتمال تغییر را بپذیر

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)}\frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$

کافی است نسبت احتمال را محاسبه کنیم

برای t بزرگ

احتمال x_t = $P(x_t)$



$P(b) = f(b) / \sum_x f(x)$

سخت

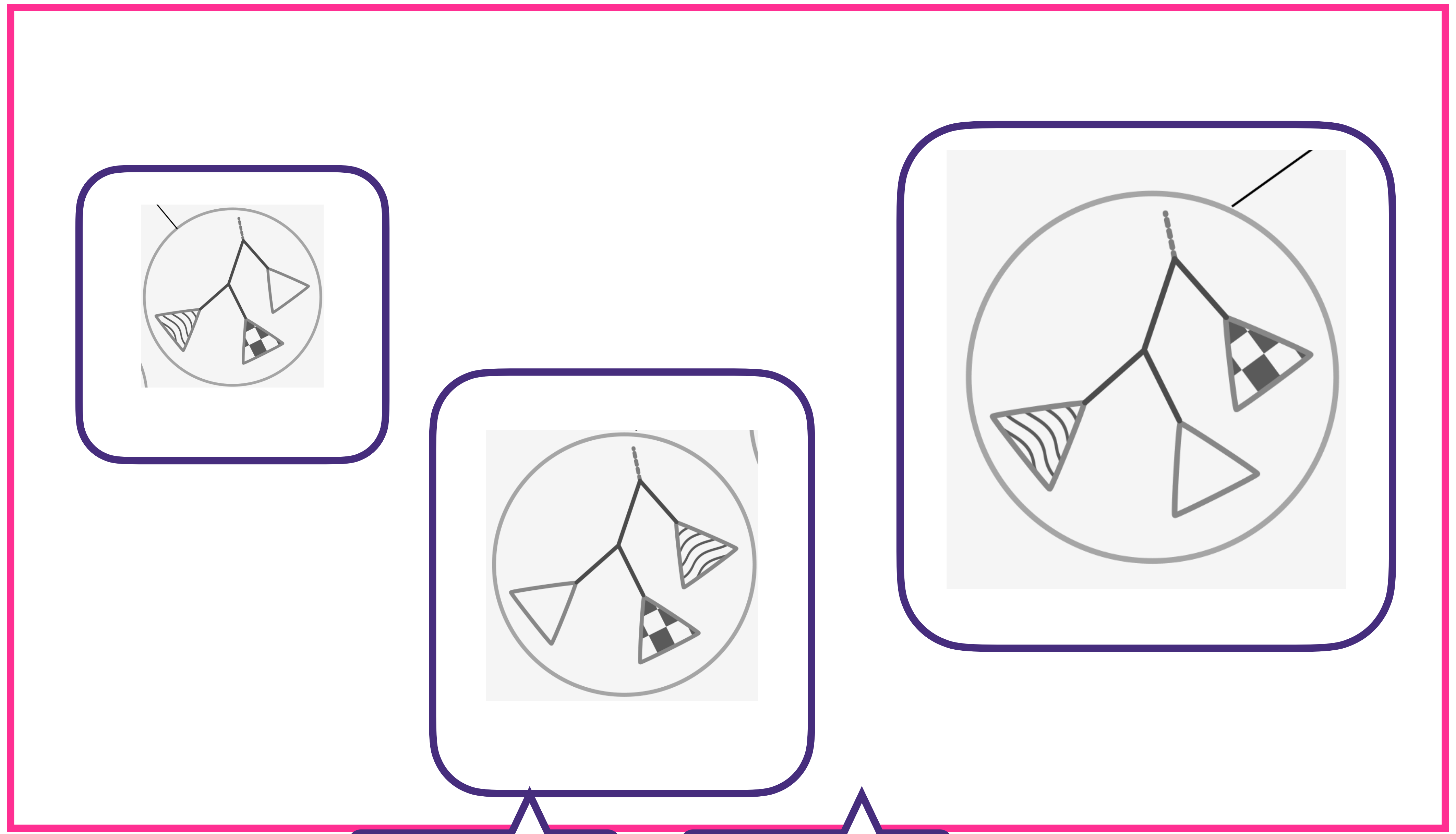
MCMC برای درخت

برای t بزرگ

احتمال x_t = $P(x_t)$

- یک نقطه ابتدایی x0
- هر دفعه
  - یک همسایه از توزیع g(x’|xt) انتخاب کن
  - با احتمال تغییر را بپذیر

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)}\frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$



L(a)

L(b)

$P(b) = f(b) / \sum_x f(x)$

MCMC برای درخت

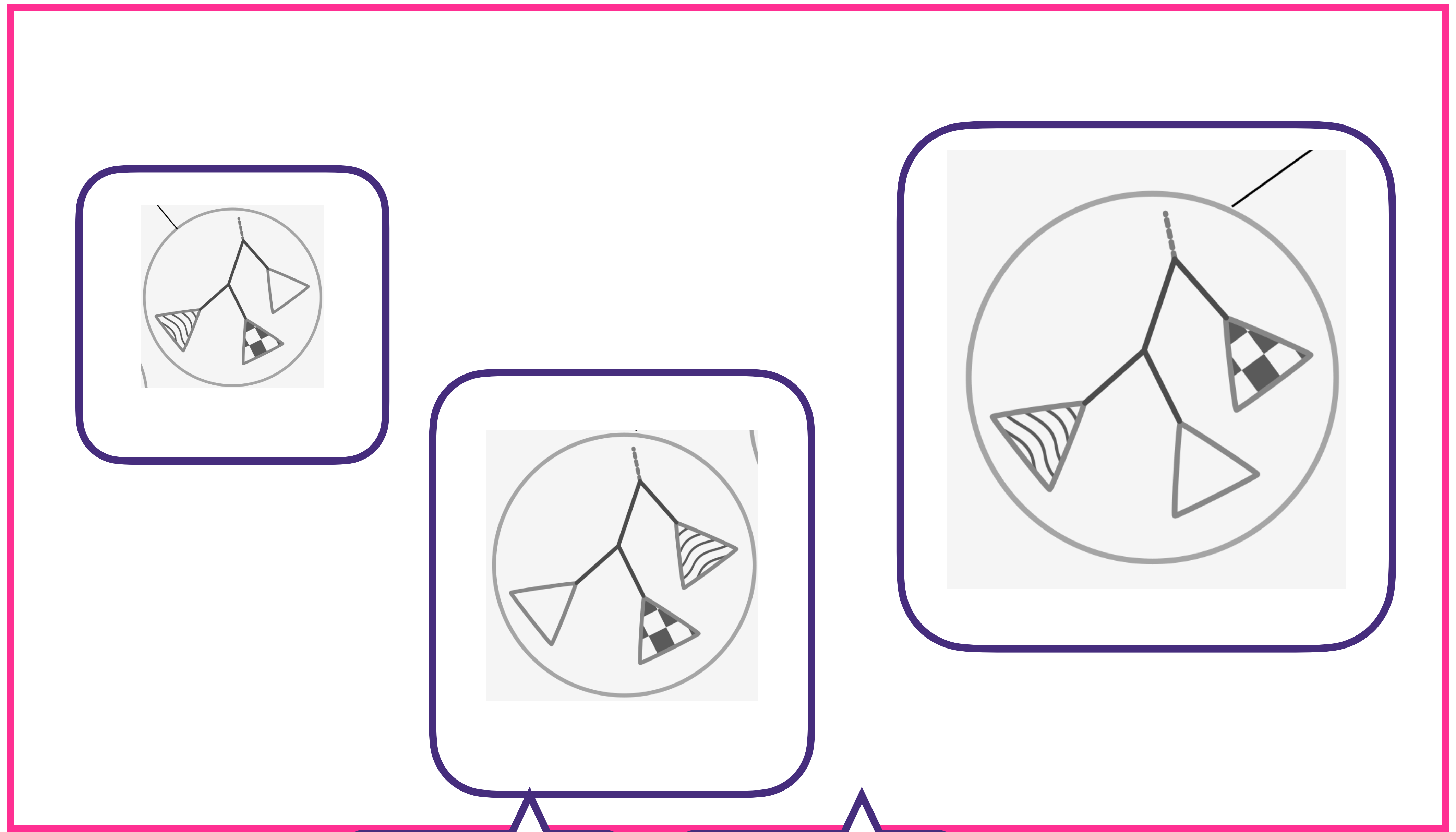
برای t بزرگ

احتمال x_t = $P(x_t)$

- یک نقطه ابتدایی x0
- هر دفعه
  - یک همسایه از توزیع g(x'|xt) انتخاب کن
  - با احتمال تغییر را بپذیر

یک زیردرخت را بکنیم جای دیگر بچسبانیم

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$



L(a)

L(b)

$P(b) = f(b) / \sum_x f(x)$

MCMC برای درخت

احتمال x_t =
$P(x_t)$

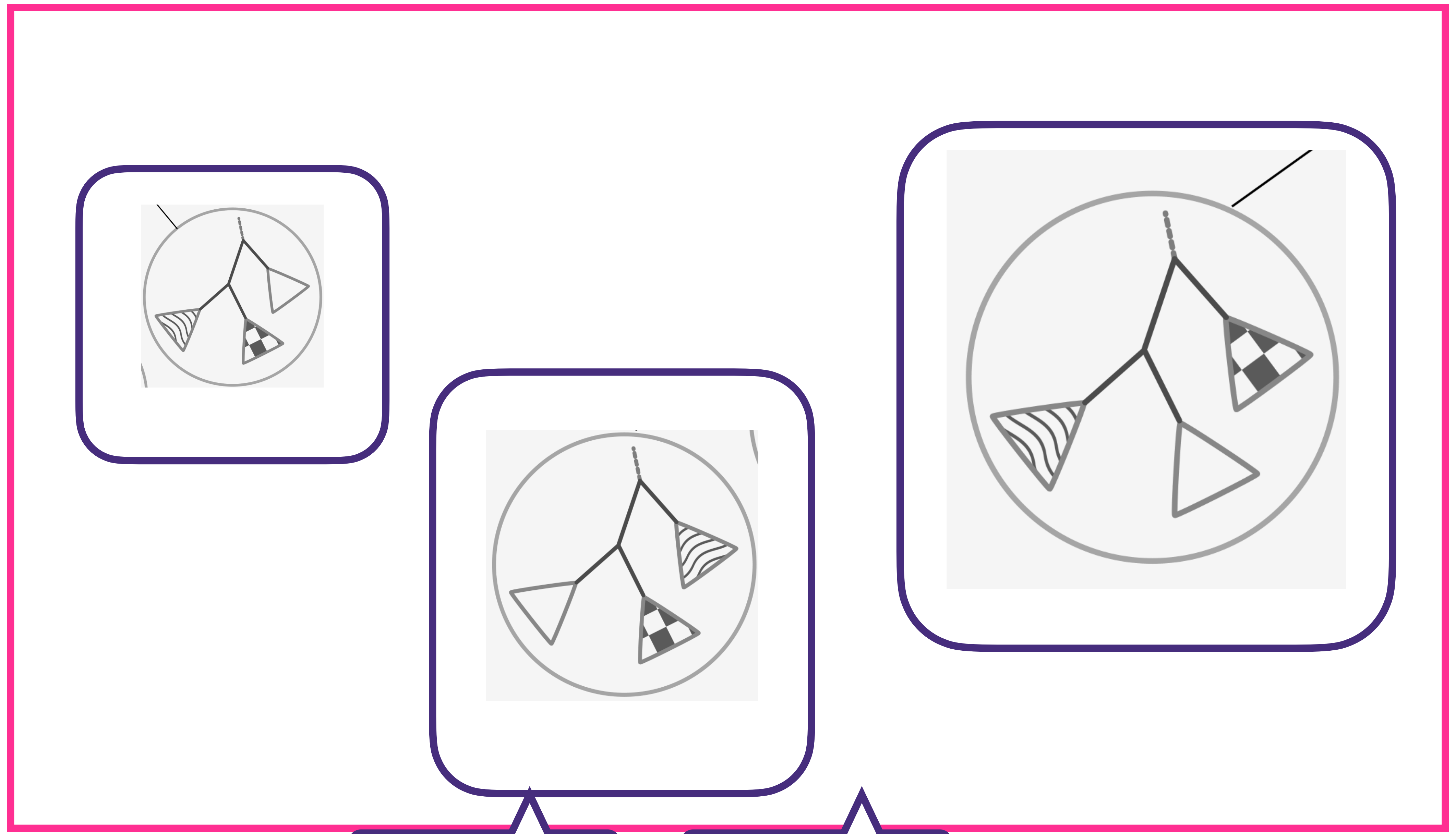
برای t بزرگ

- یک نقطه ابتدایی x0
- هر دفعه
- یک همسایه از توزیع g(x'|xt) انتخاب کن
- با احتمال تغییر را بپذیر

یک زیردرخت را بکنیم
جای دیگر بچسبانیم

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$

$L(x') \,/\, L(x_t)$



L(a)

L(b)

$P(b) = f(b) \,/\, \sum_x f(x)$

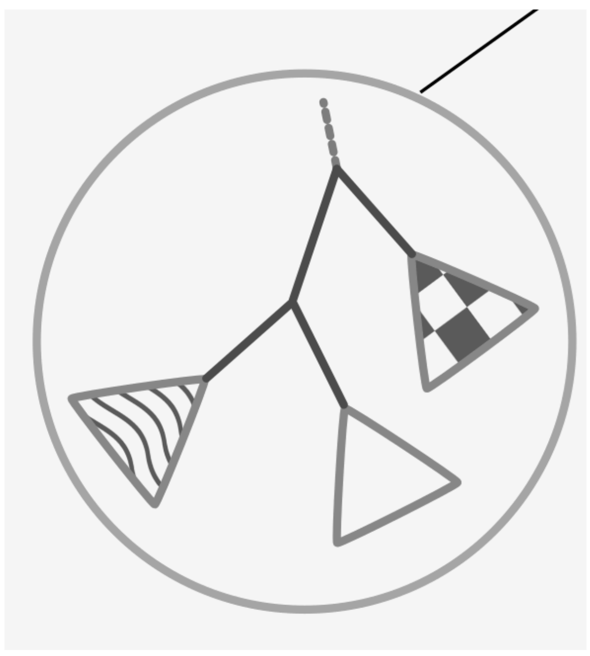
MCMC برای درخت

برای t بزرگ

احتمال x_t = $P(x_t)$

- یک نقطه ابتدایی x0
- هر دفعه
- یک همسایه از توزیع g(x'|xt) انتخاب کن
- با احتمال تغییر را بپذیر

یک زیردرخت را بکنیم جای دیگر بچسبانیم



L(a)

L(b)

$P(b) = f(b) \,/\, \sum_x f(x)$

$$A(x', x_t) = \min\left(1, \frac{P(x')}{P(x_t)} \frac{g(x_t \mid x')}{g(x' \mid x_t)}\right)$$

$L(x') \,/\, L(x_t)$

1