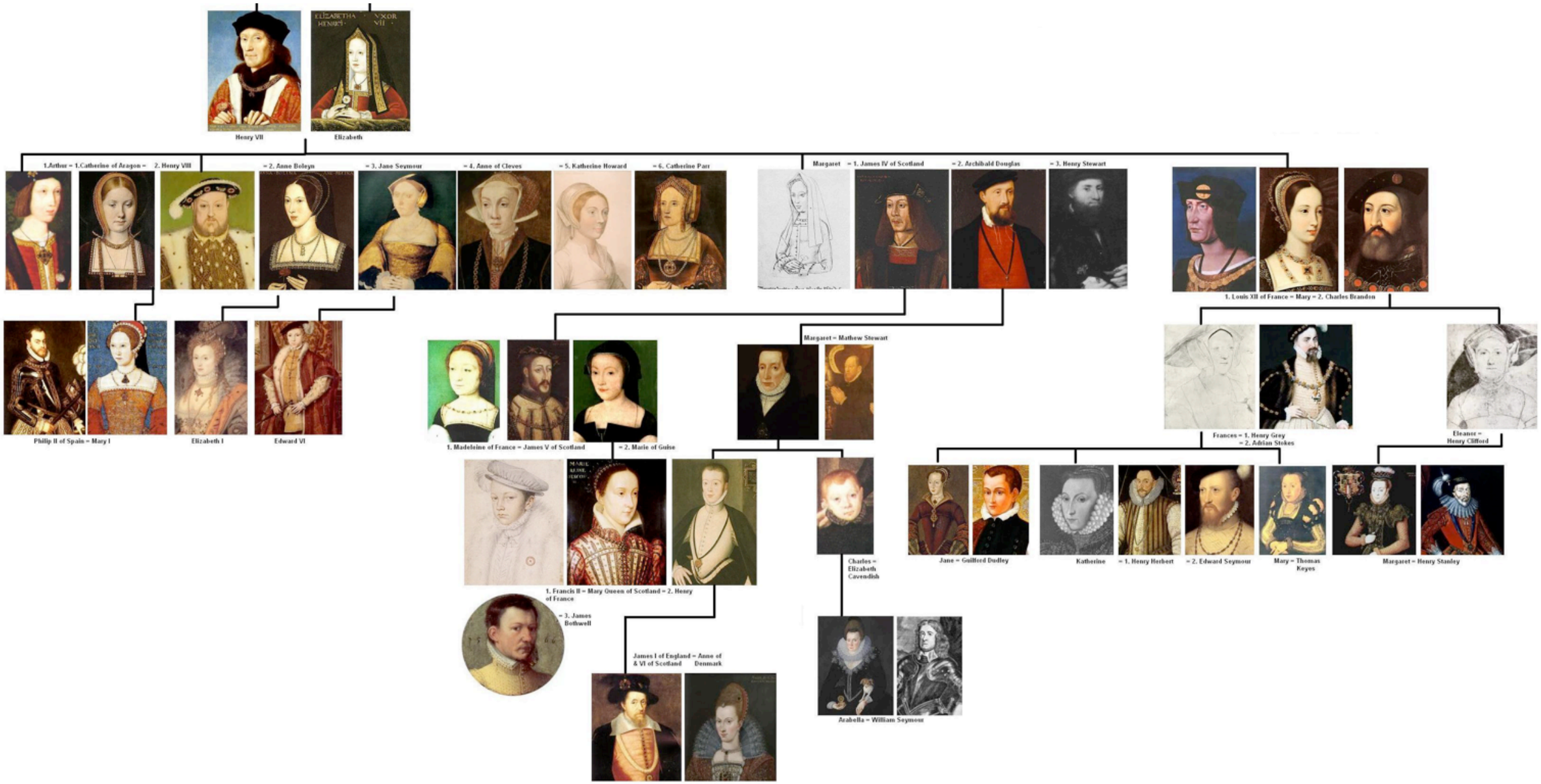
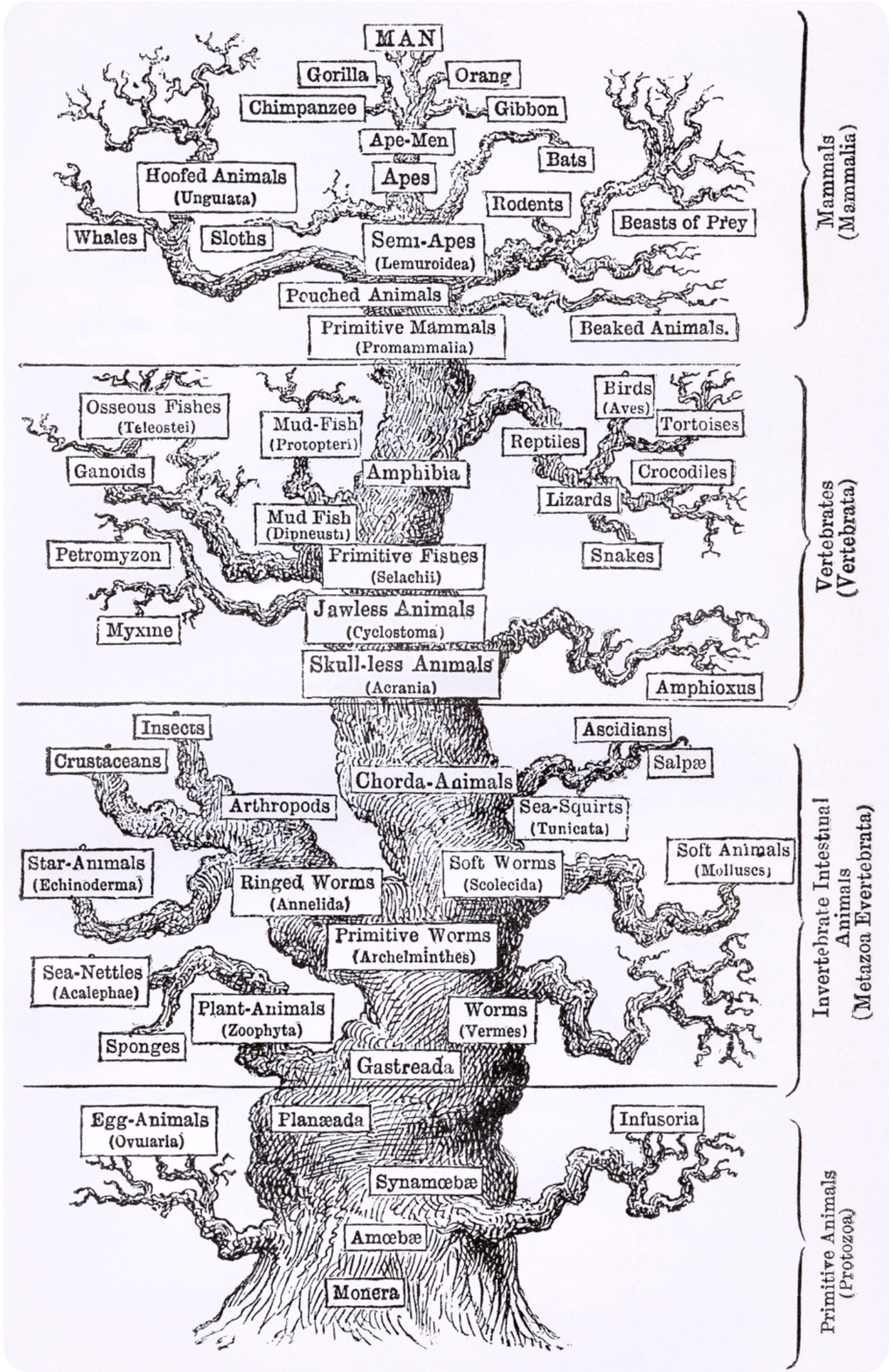
Phylogenies

Phylogenies describe history

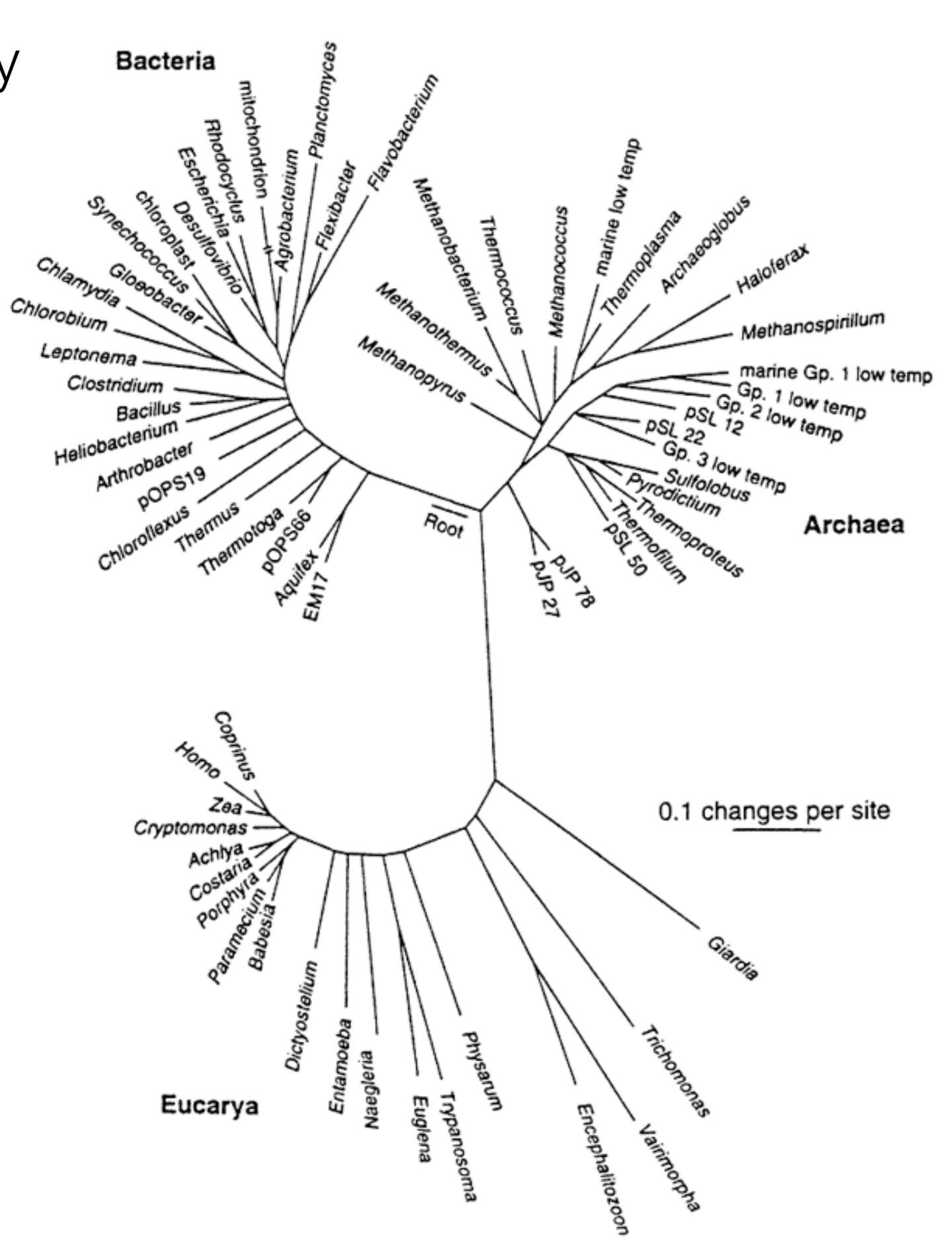


Phylogenies describe history



Haeckel. 1879.

Phylogenies describe history

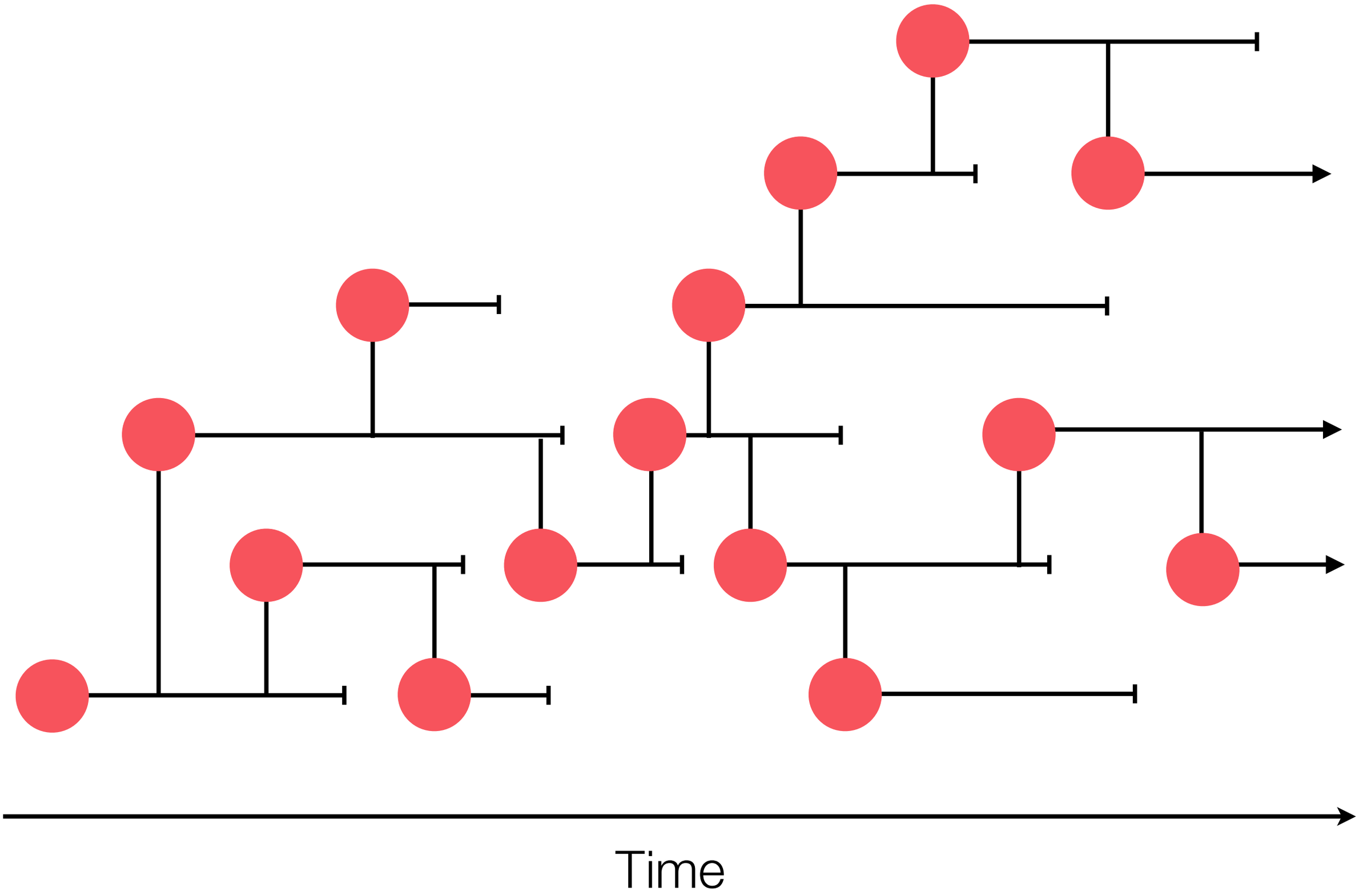


Pace. 1997. Science.

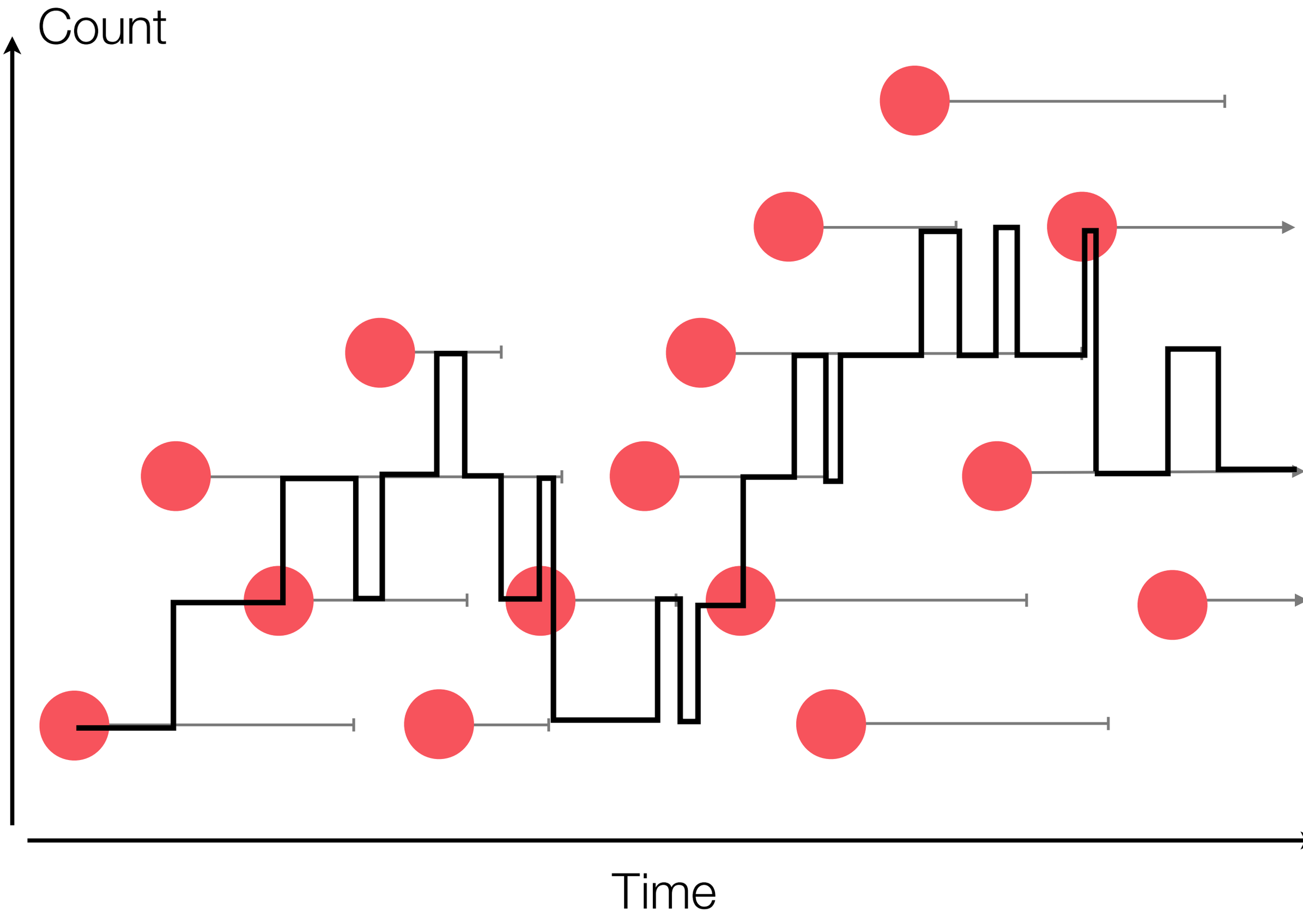
Phylogenies are the result of branching processes

Timeseries and phylogeny are dual outcomes of an infectious process

Epidemic process



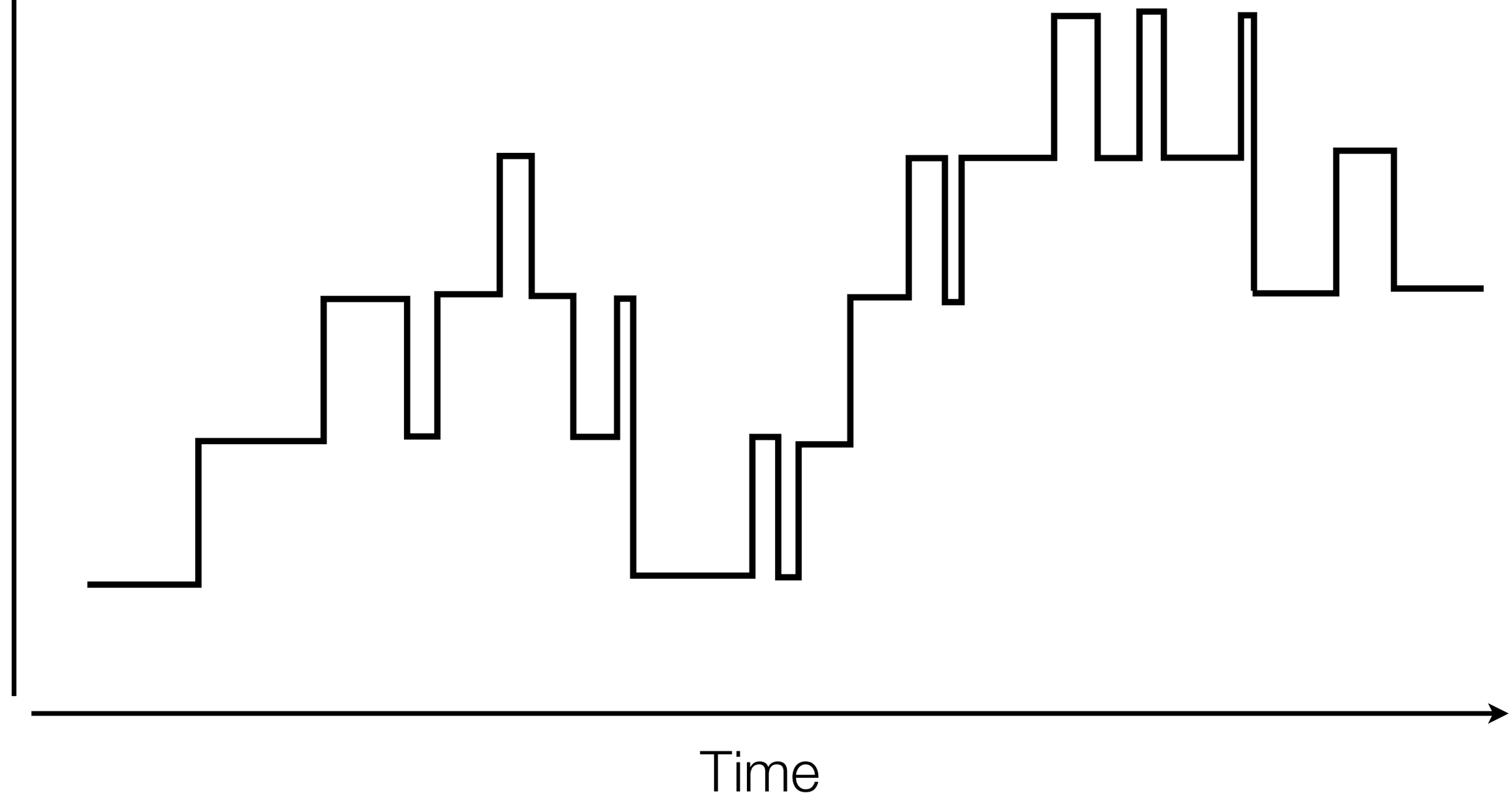
Epidemic process



Epidemic process

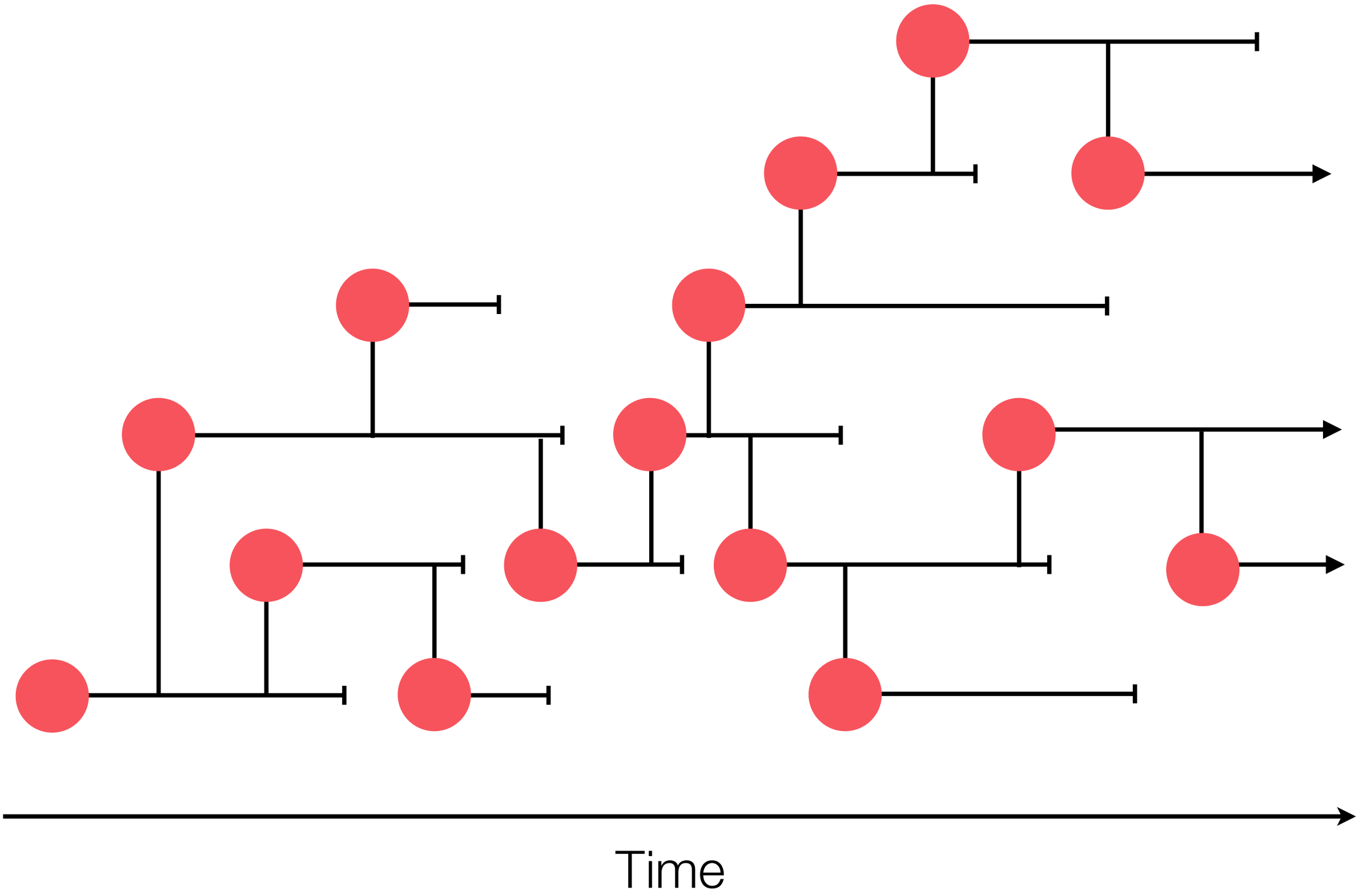
Count

Can ask for the probability of observing this timeseries given epidemiological parameters β and γ.



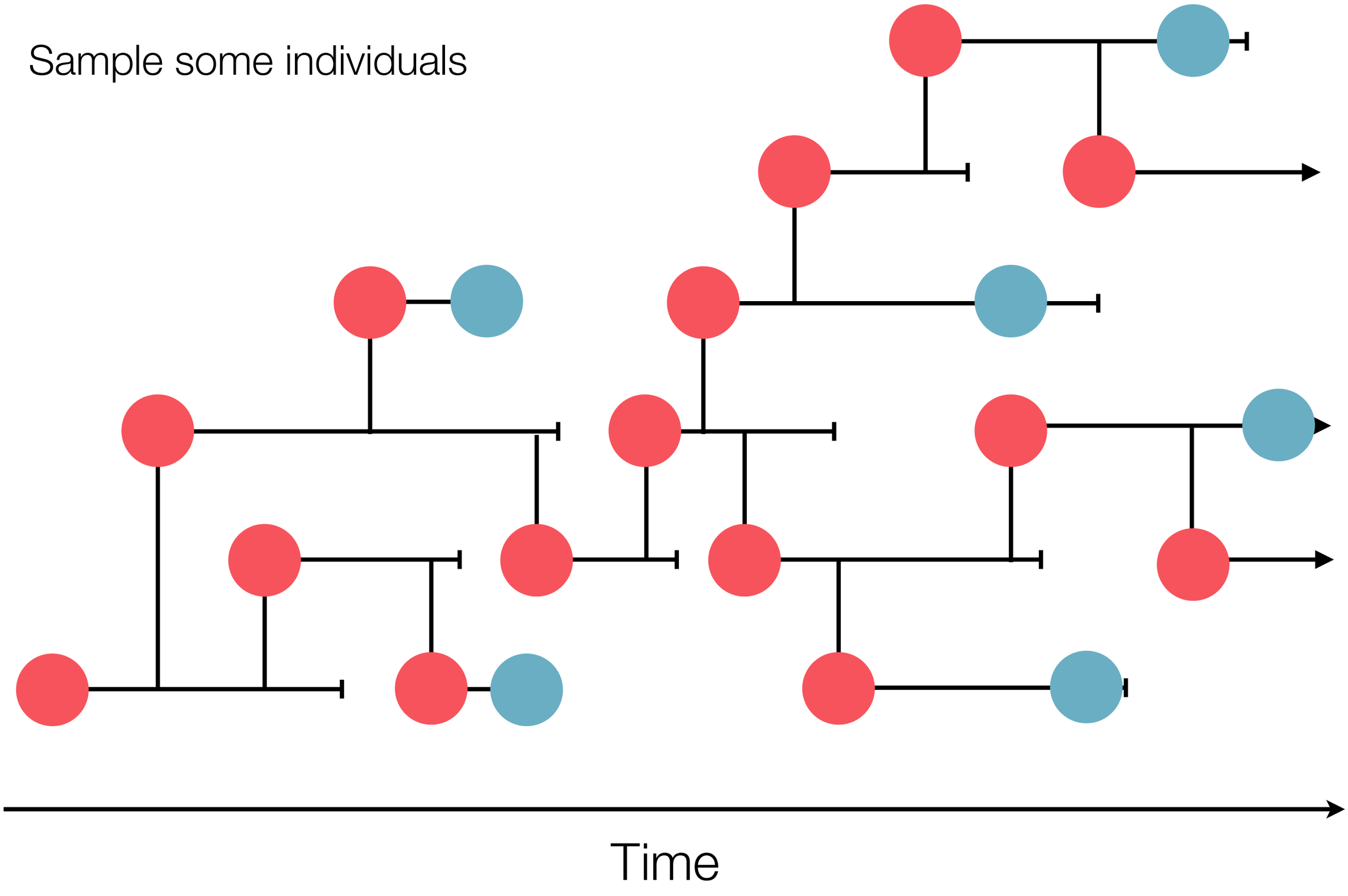
Time

Epidemic process

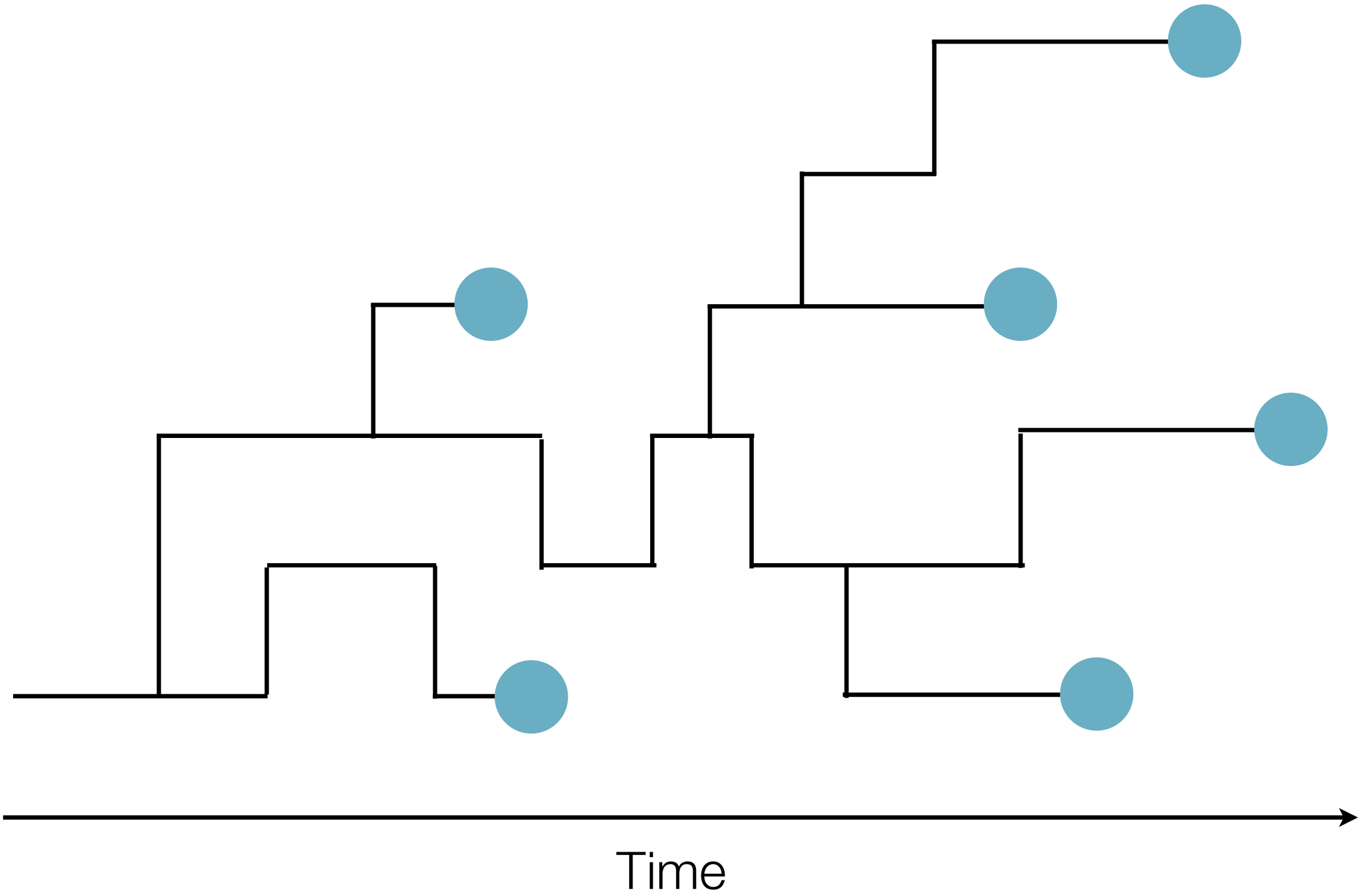


Epidemic process

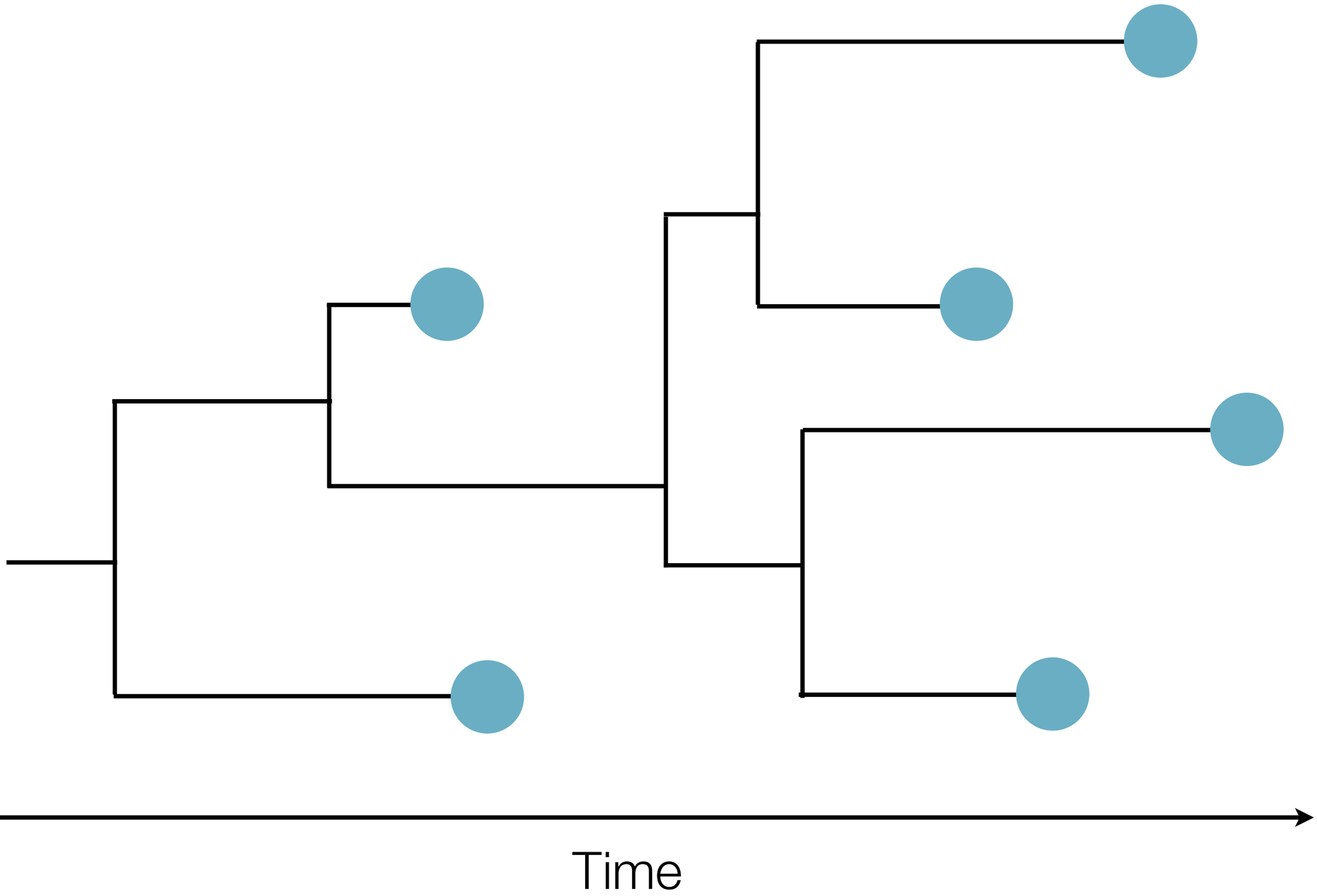
Sample some individuals



Epidemic branching process

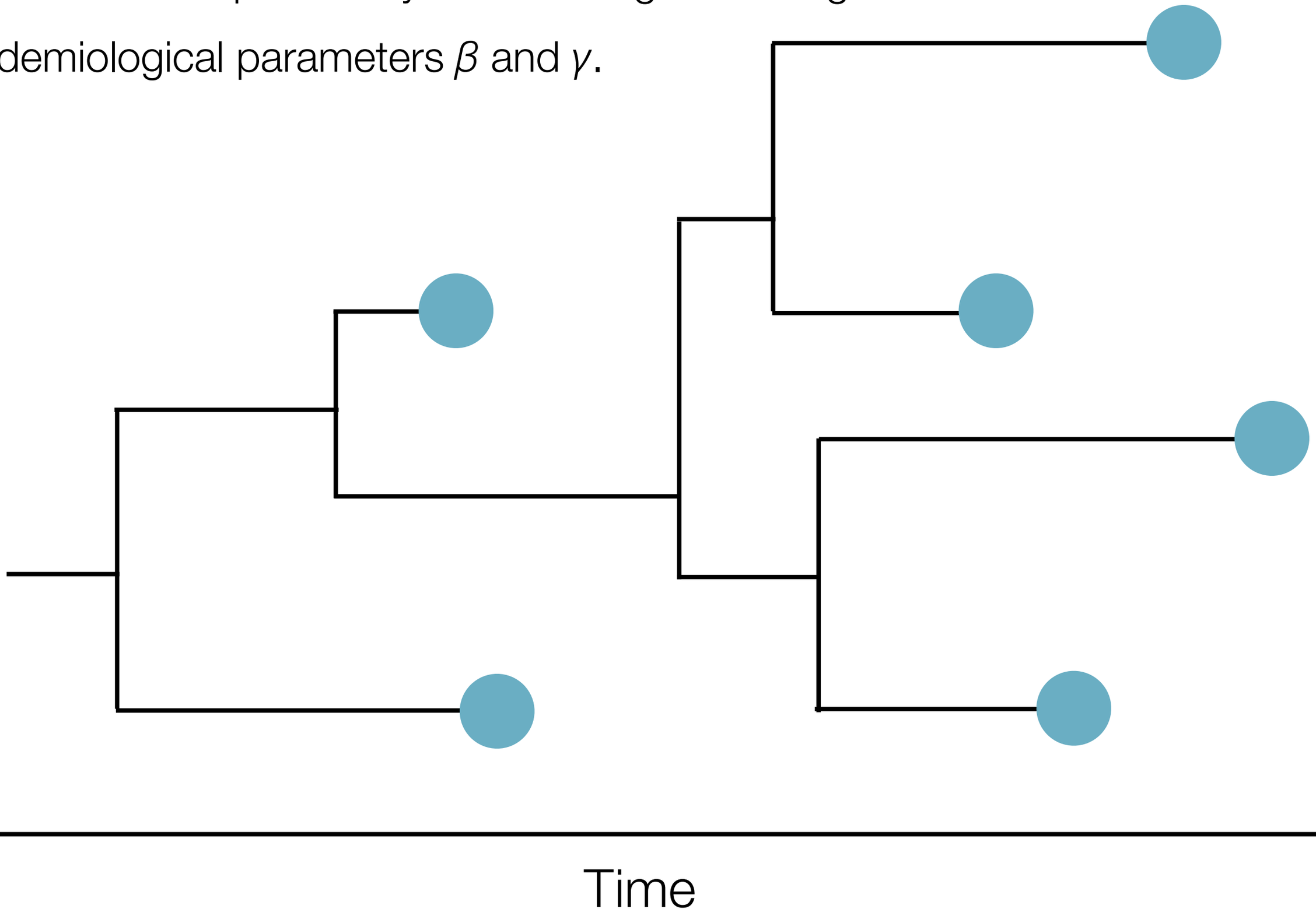


Epidemic branching process



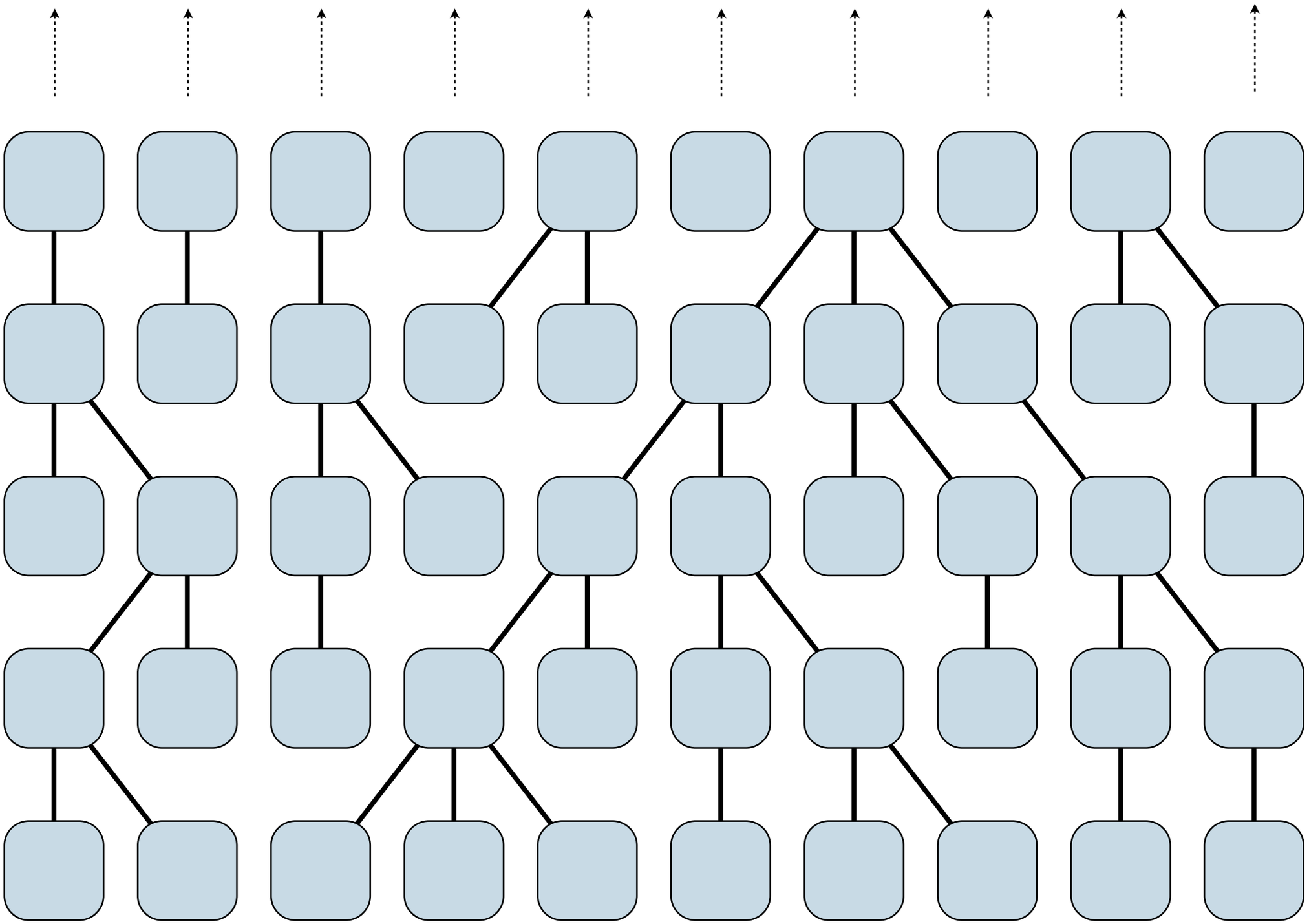
Epidemic branching process

Can ask for the probability of observing this tree given epidemiological parameters β and γ.



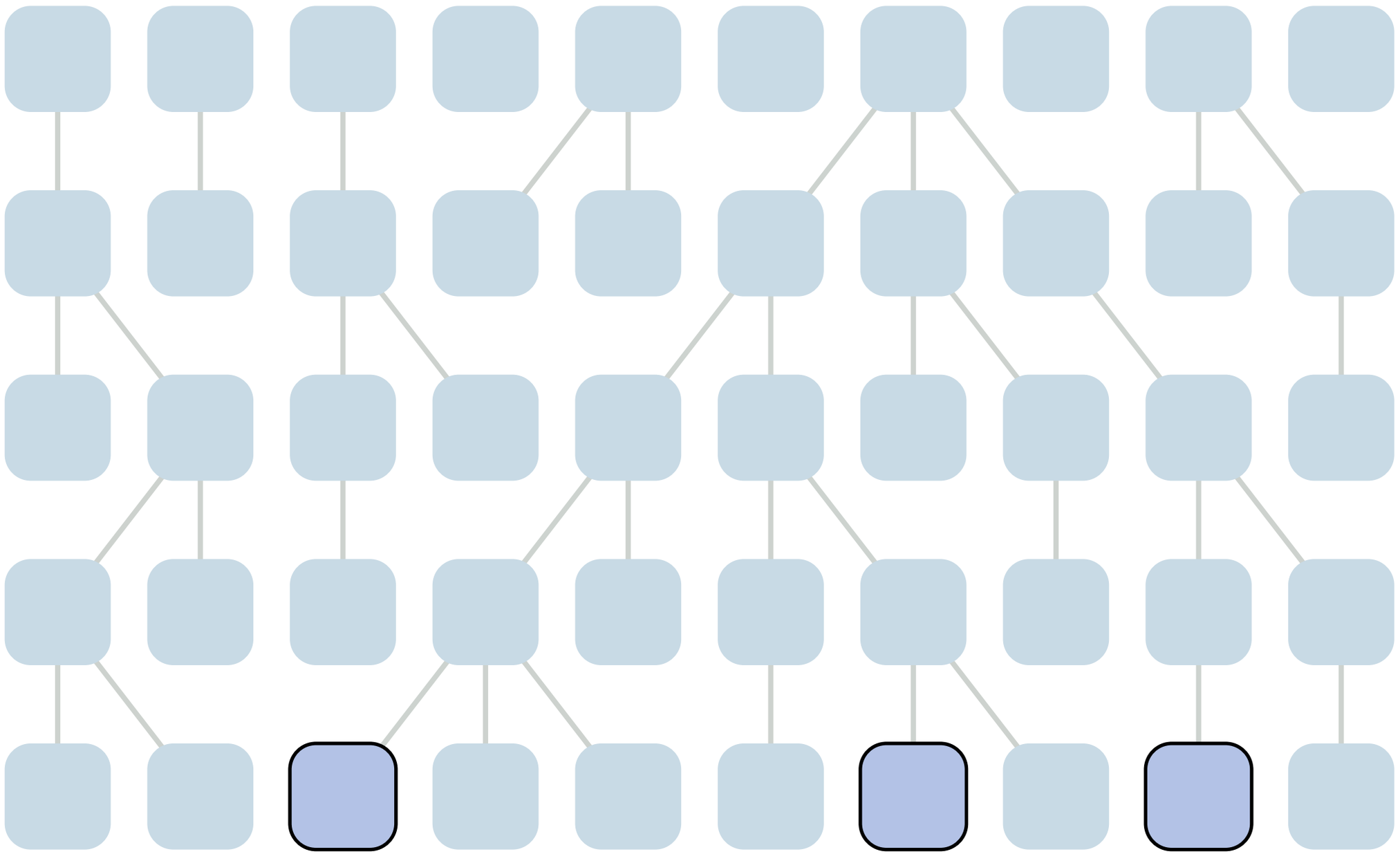
The coalescent

Assume equilibrium number of infecteds. Call this equilibrium *N*.



The coalescent

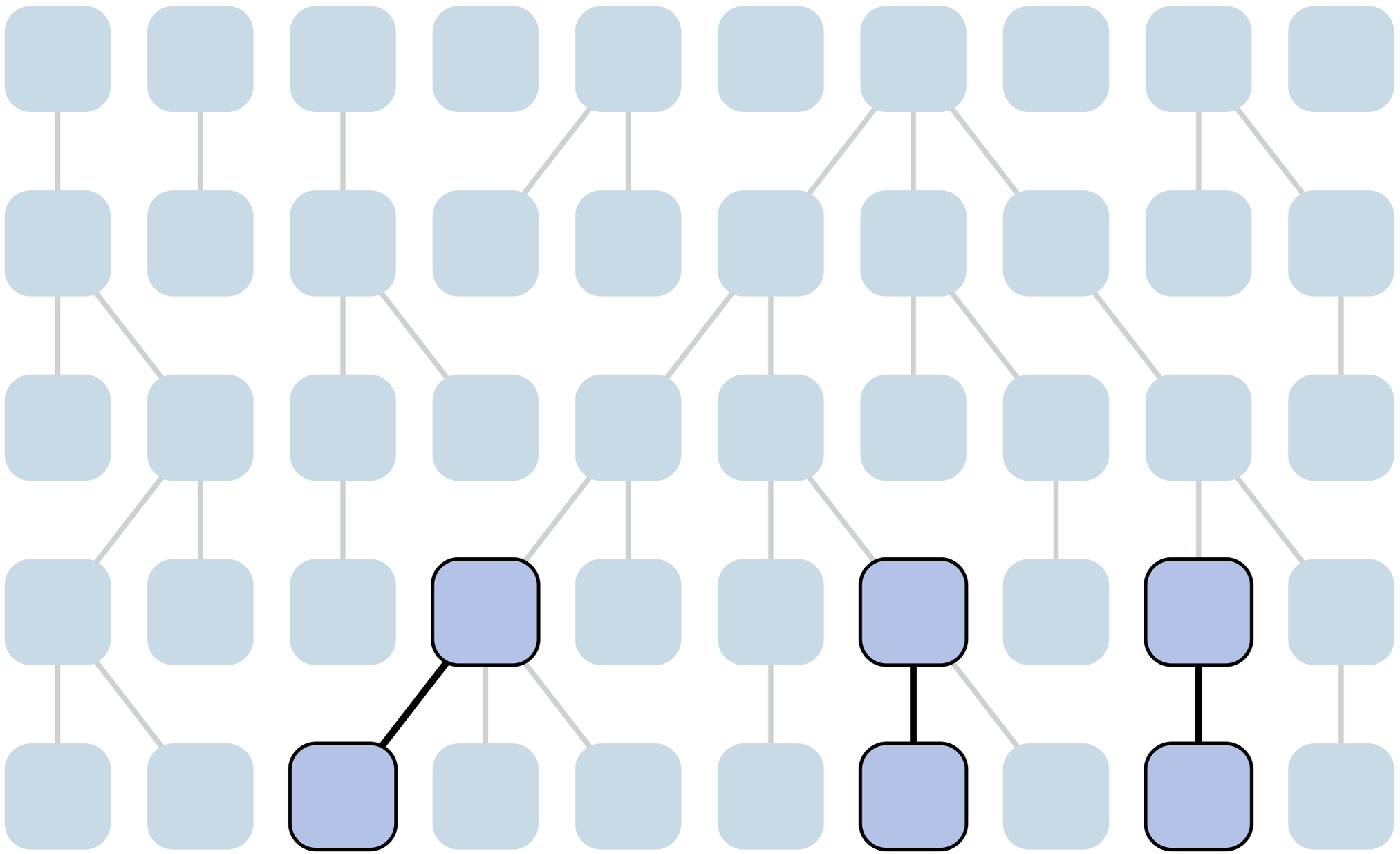
Sample some individuals



The coalescent

Each generation, there is a small chance for coalescence for each pair

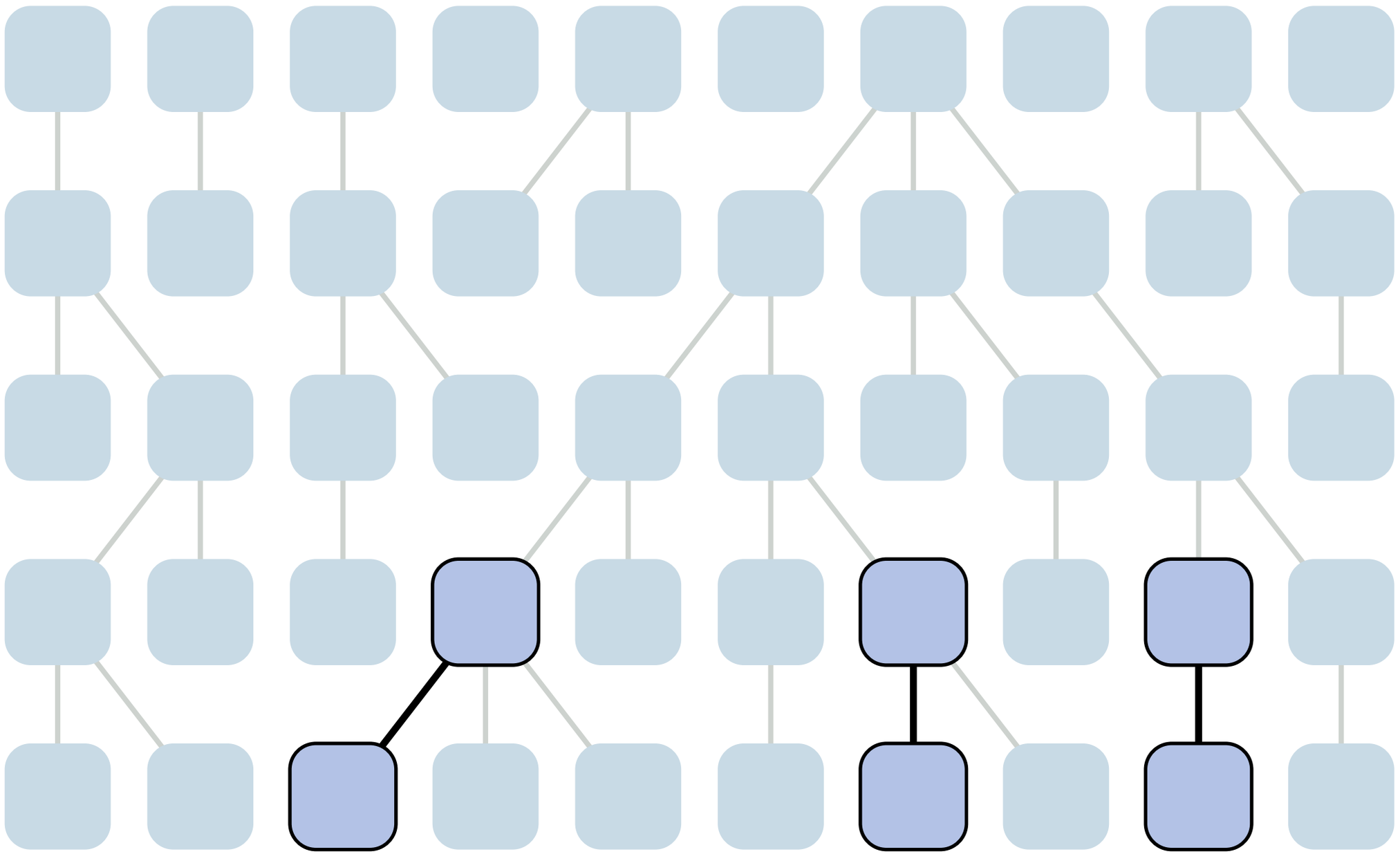
$$\Pr(\text{coal}|i=2) = \frac{1}{N}$$



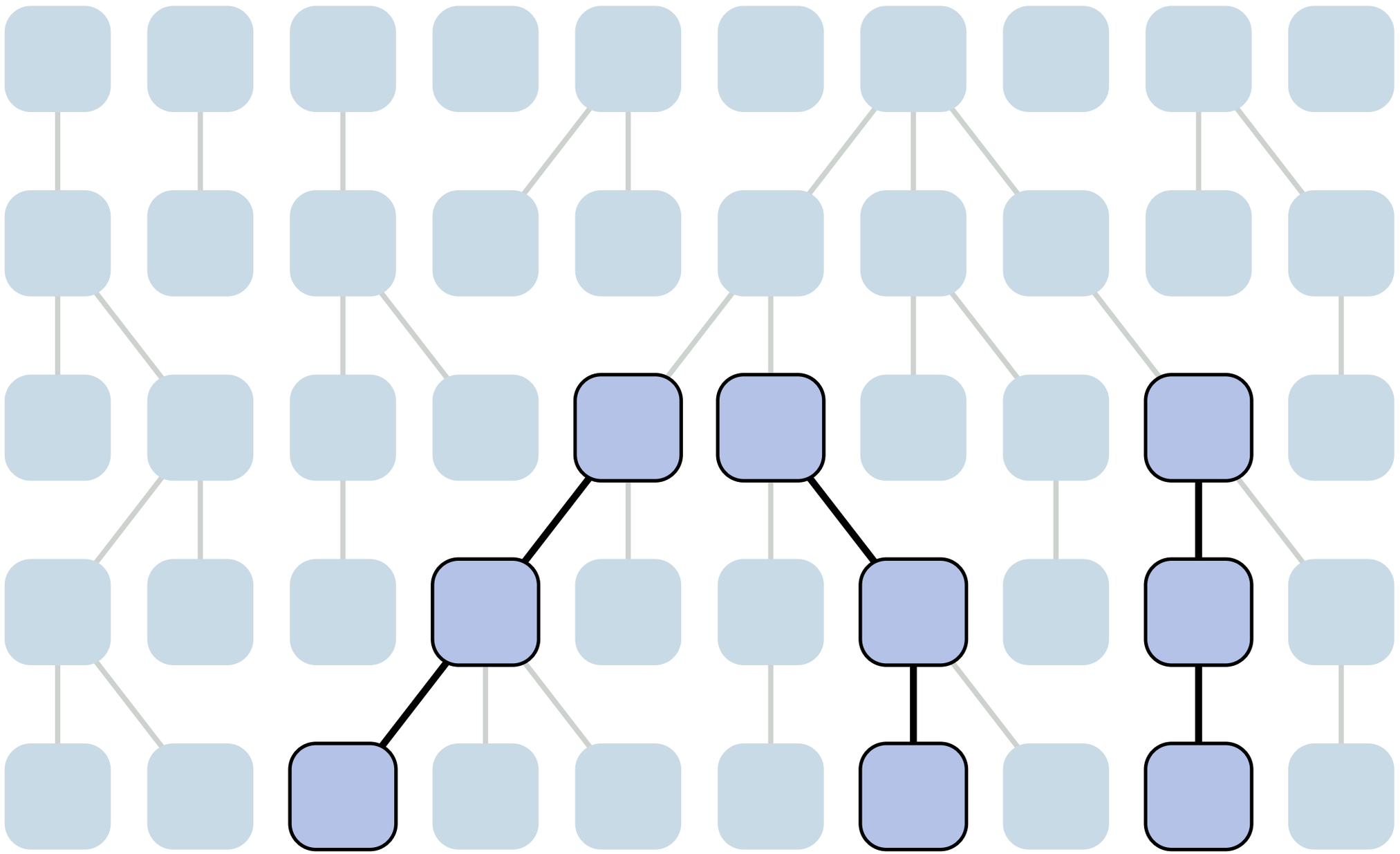
The coalescent

Probability of coalescence scales quadratically with lineage count

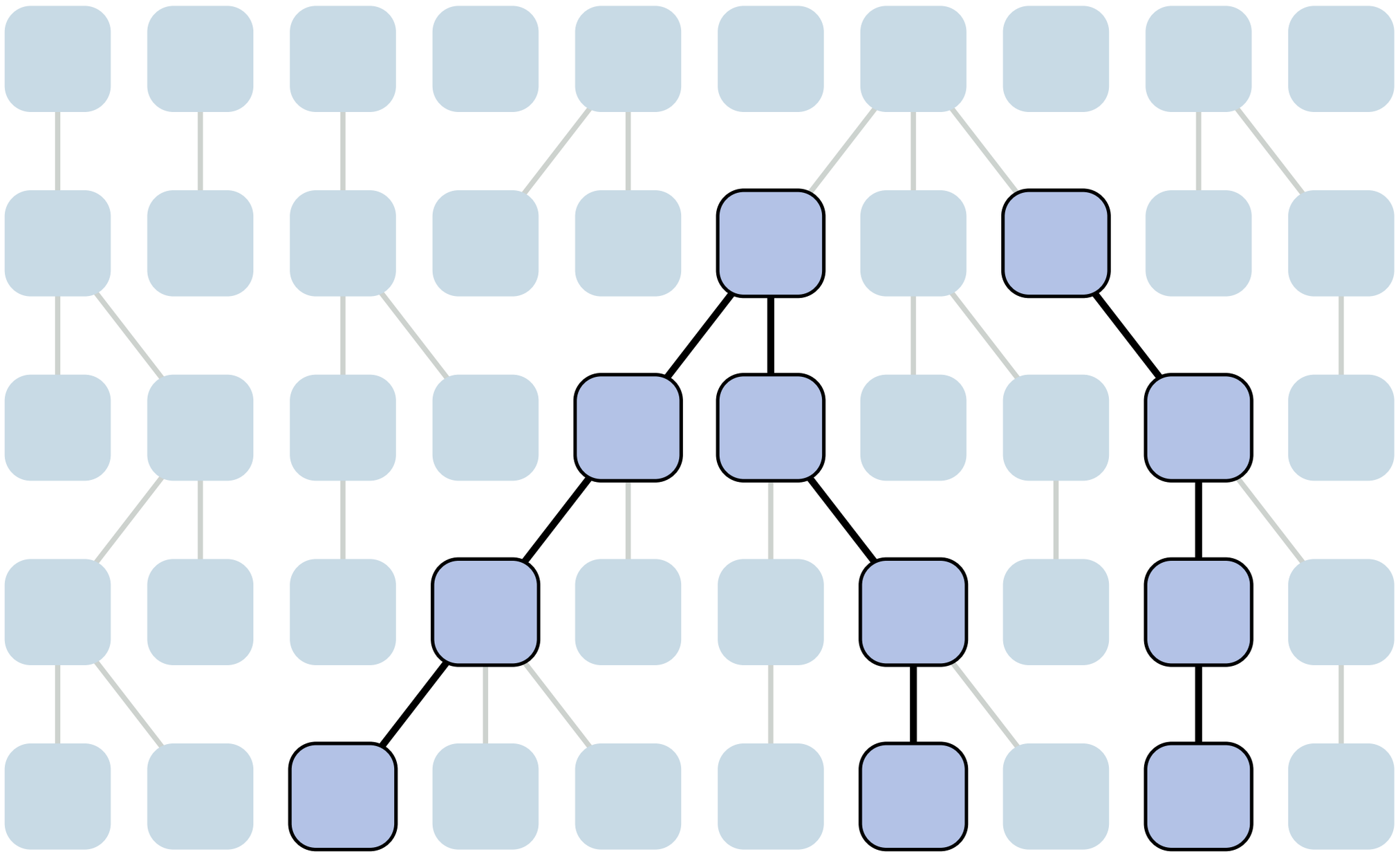
$$\Pr(\text{coal}) = \binom{i}{2}\frac{1}{N} = \frac{i(i-1)}{2N}$$



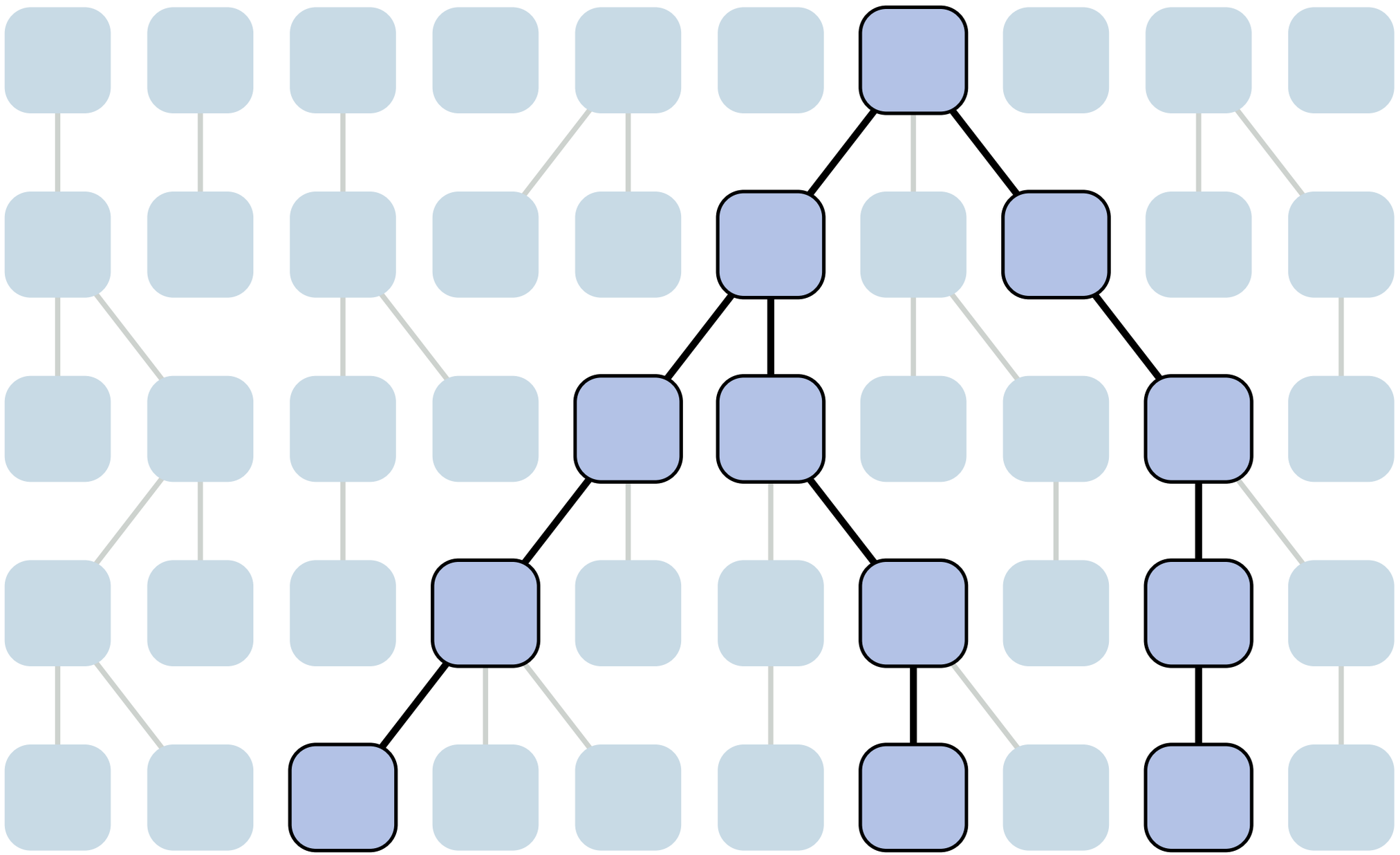
The coalescent



The coalescent

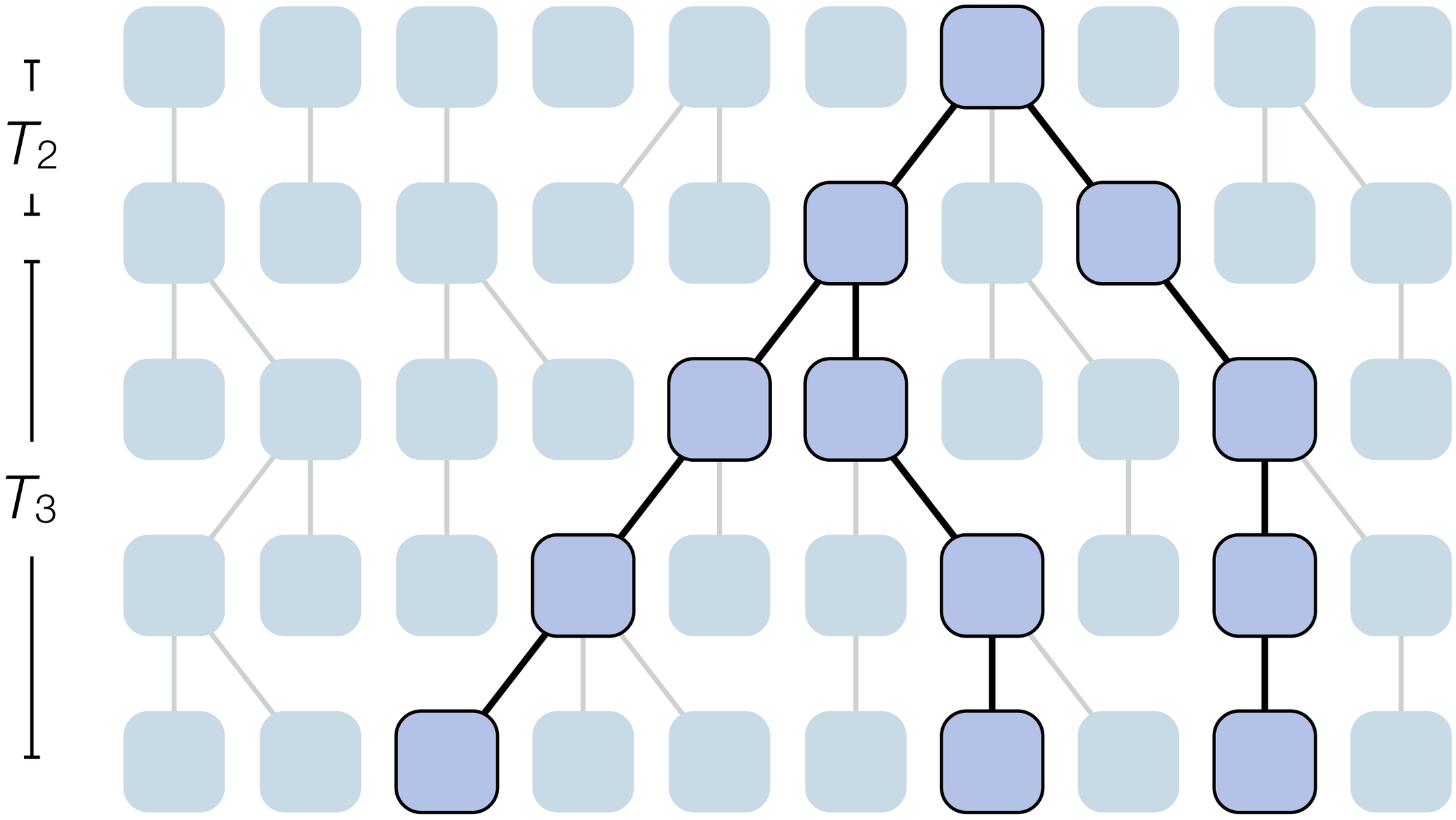


The coalescent



The coalescent

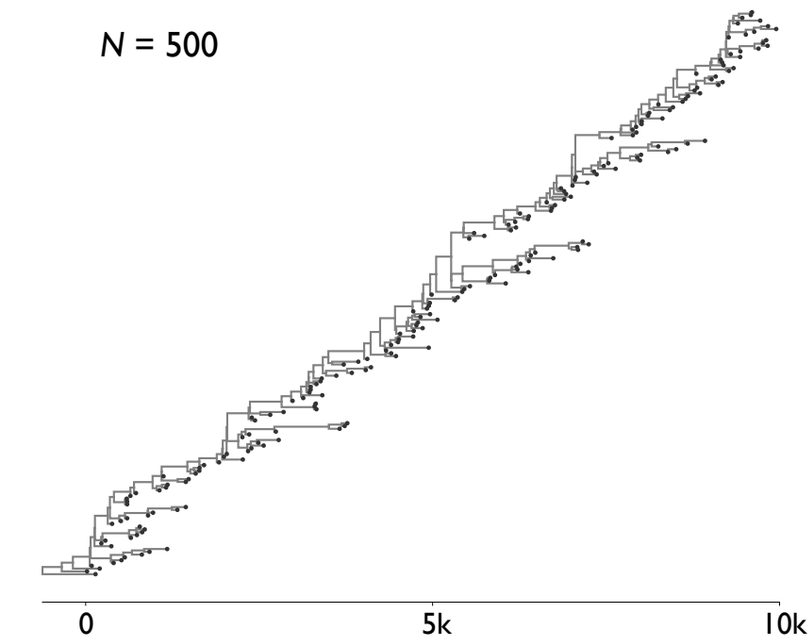
$$T_i \sim \text{Exponential}\left(\frac{2N}{i(i-1)}\right)$$

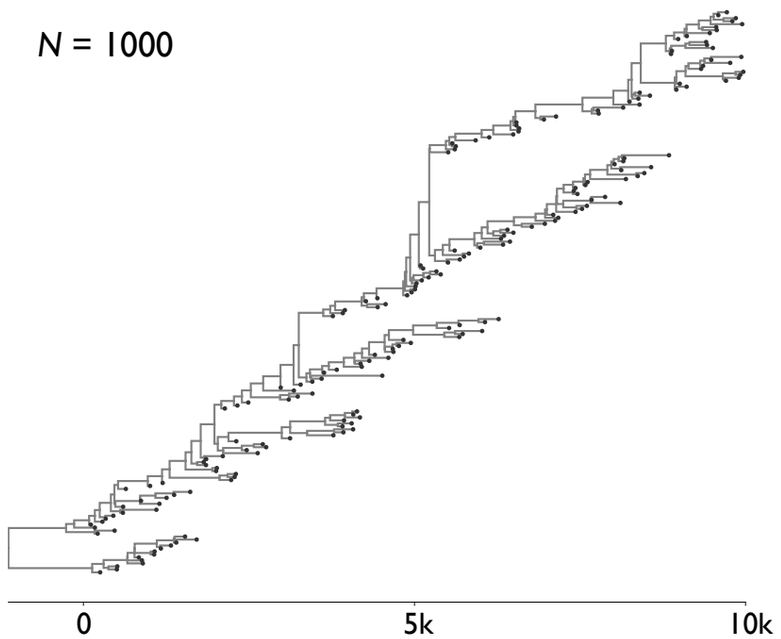


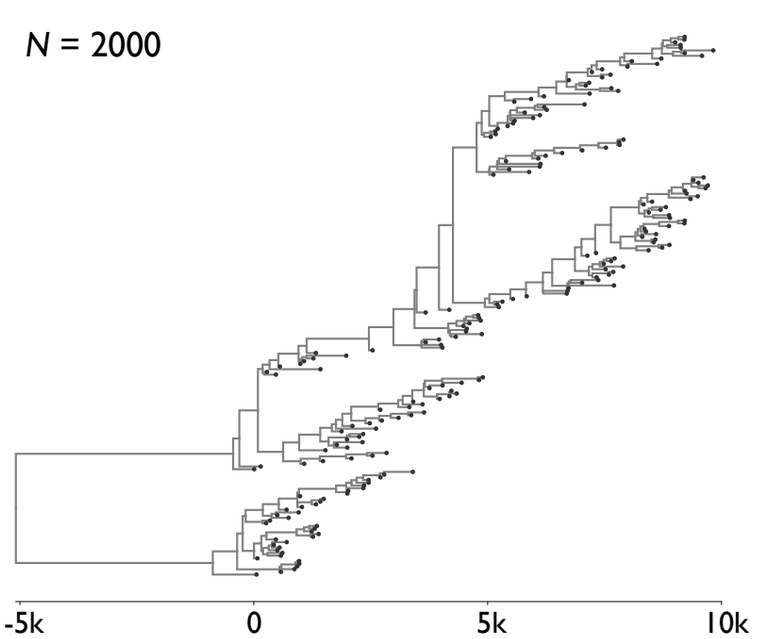
Demo

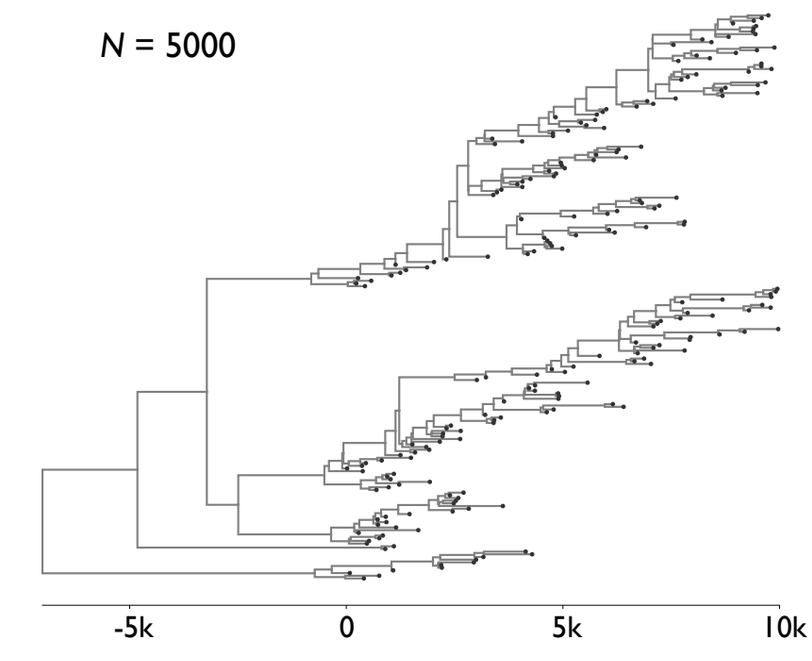
Population size affects tree shape

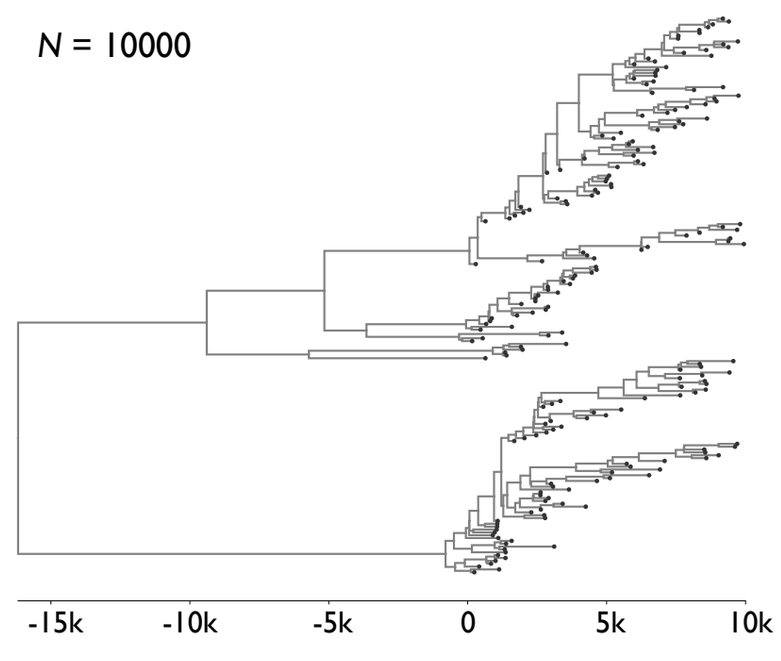
The rate of coalescence decreases linearly with the population size *N*.

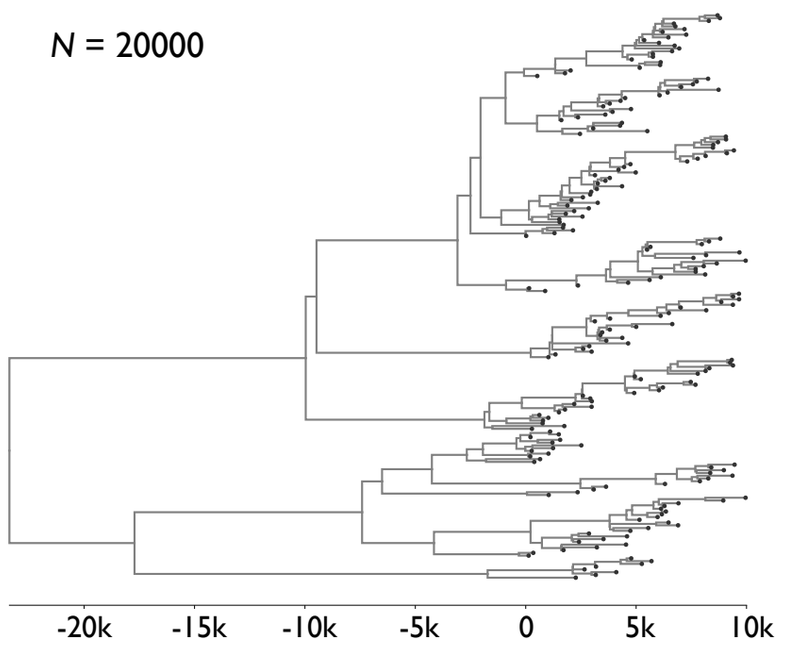




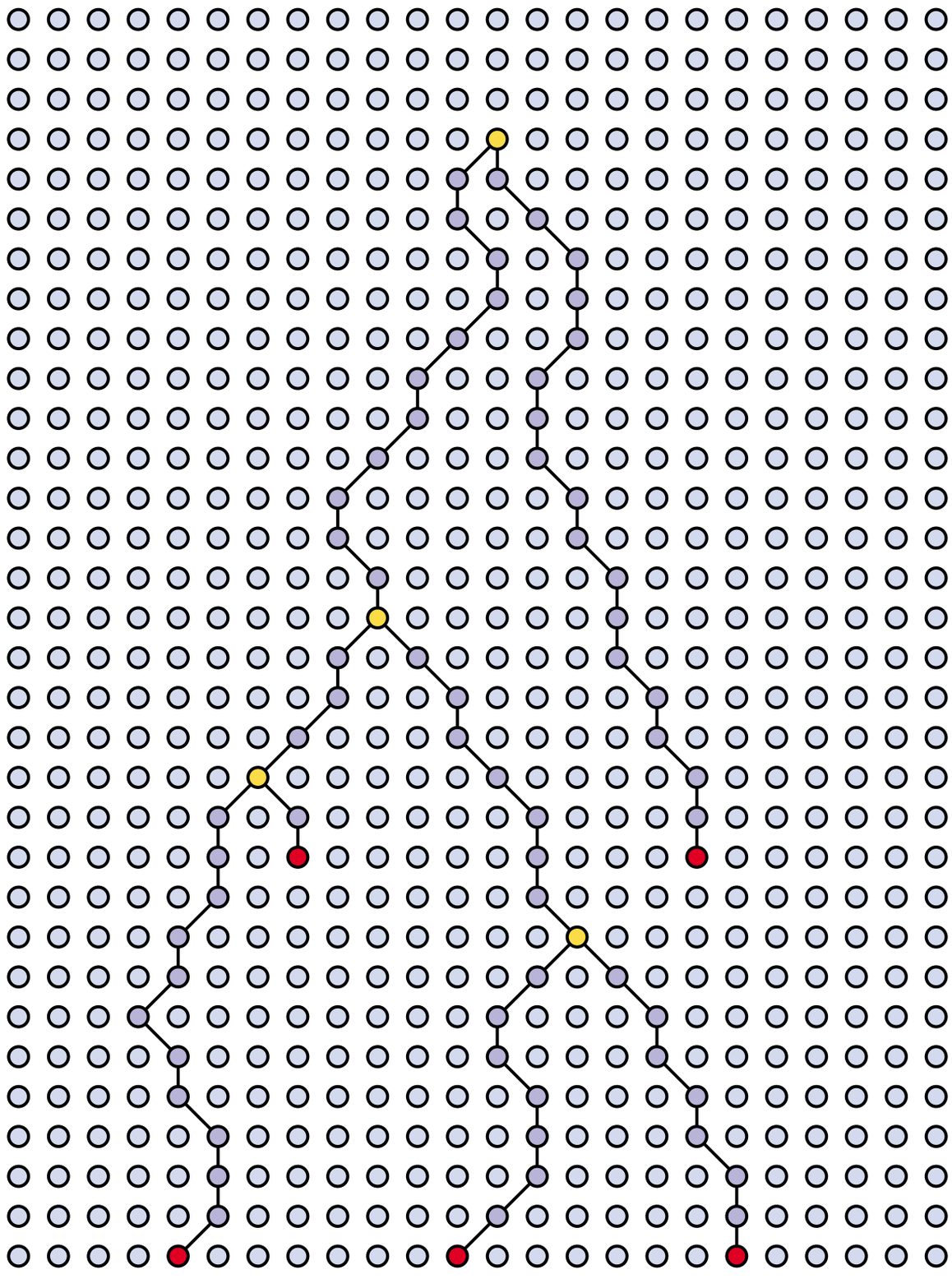




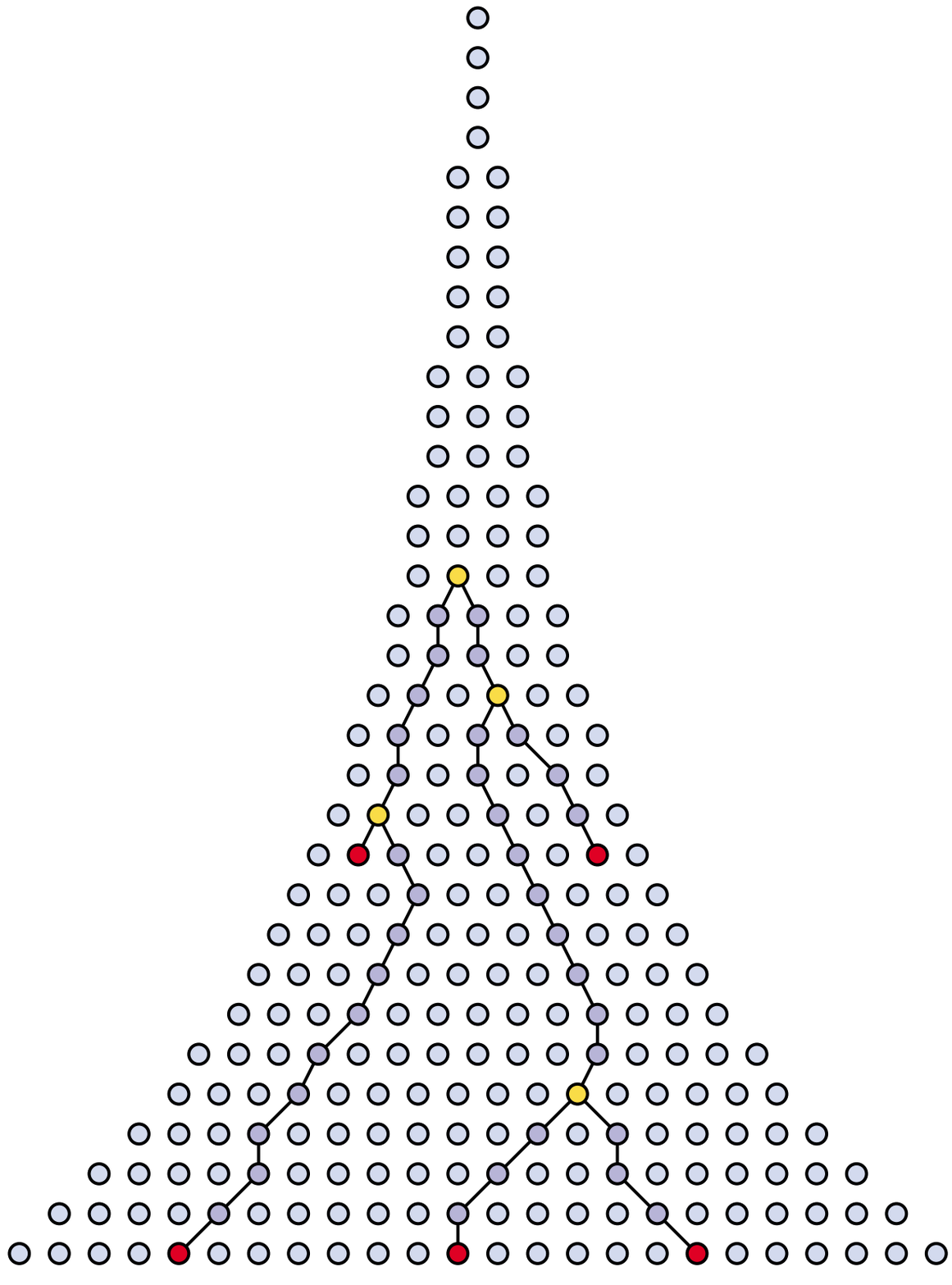




Changing population size

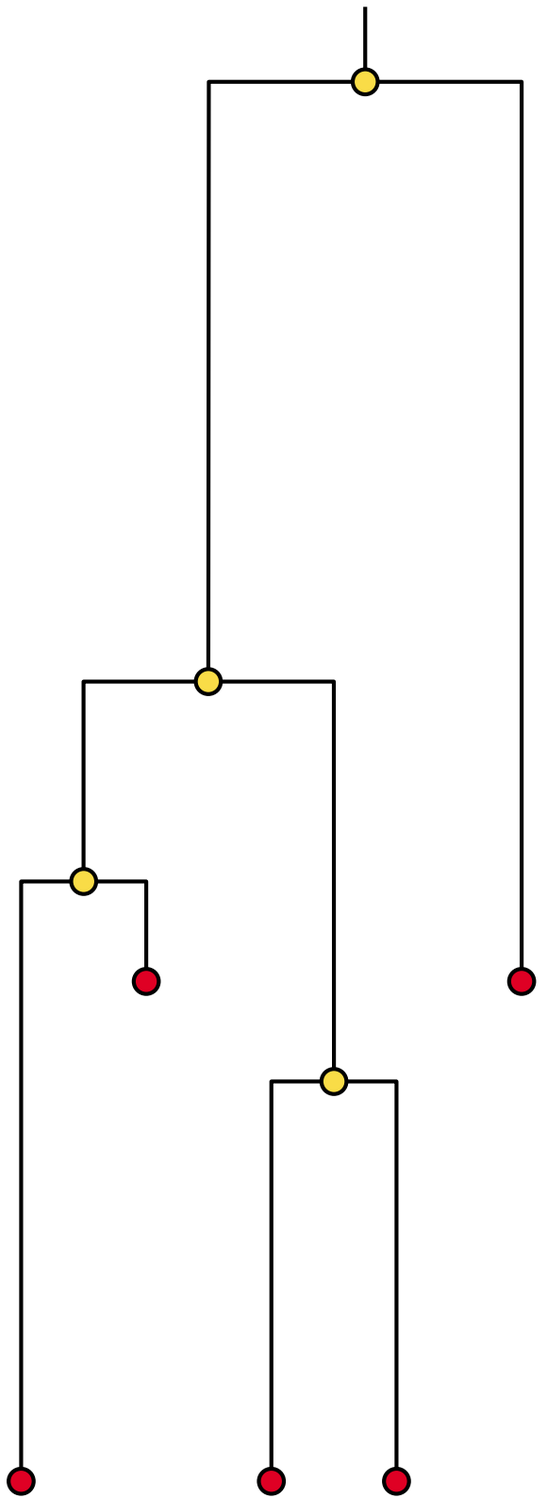


Constant size

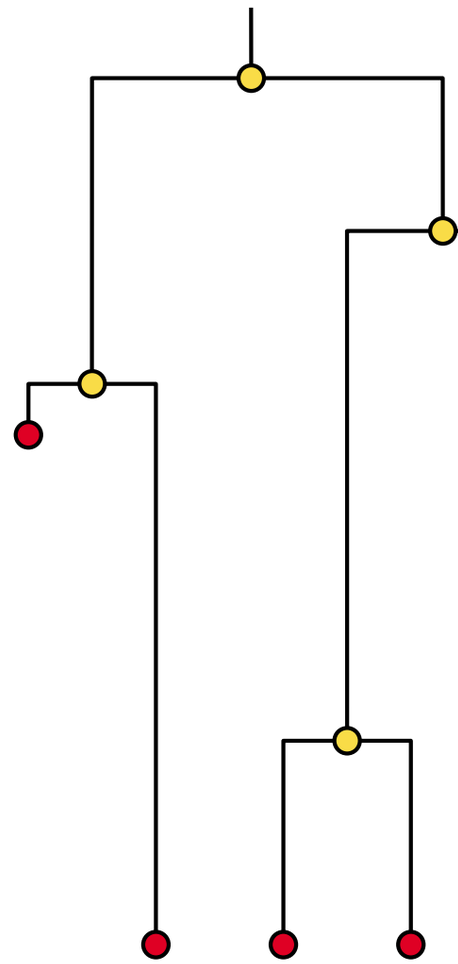


Growing population

Changing population size



Constant size



Growing population

Given a phylogeny, how can we learn about the evolutionary process that underlies it?

Generally, we want to know:

$$p(\text{model}|\text{data})$$

Bayes rule:

$$p(\text{model}|\text{data}) \propto p(\text{data}|\text{model})\ p(\text{model})$$

Often referred to as:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

λ – coalescent model

τ – phylogeny

D – sequence data

μ – mutation model

In this case, we have:

$$p(\lambda|\tau) \propto p(\tau|\lambda)\ p(\lambda)$$

However, we don't observe the tree directly:

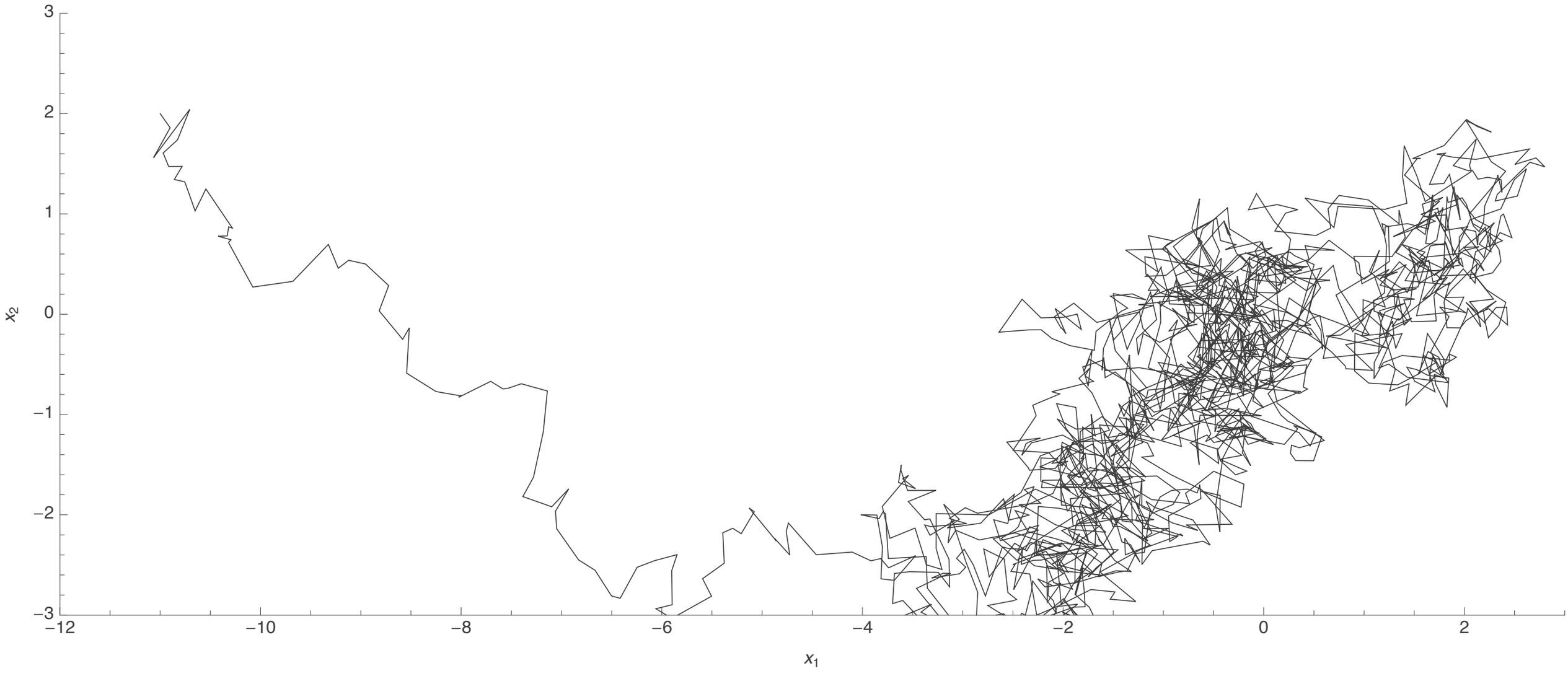
$$p(\tau, \mu|D) \propto p(D|\tau, \mu)\ p(\tau)\ p(\mu)$$

We integrate over uncertainty:

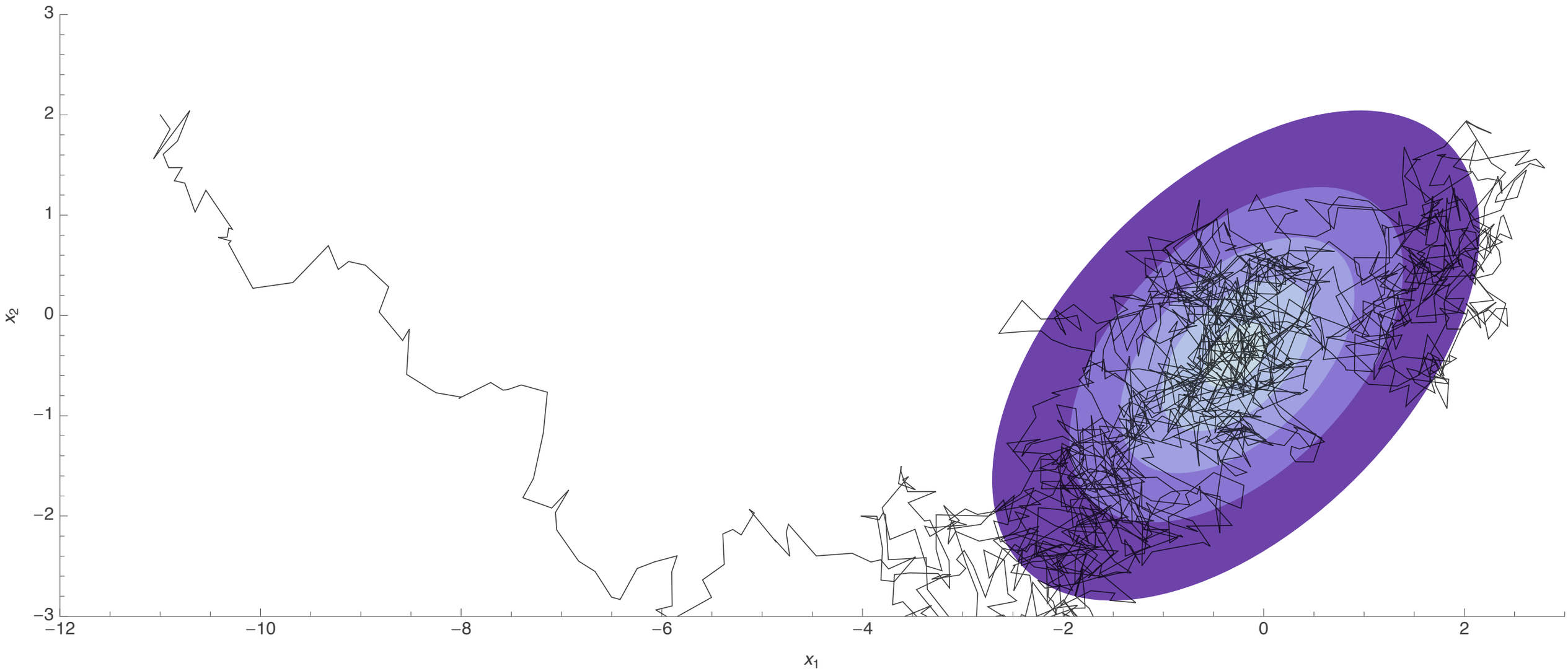
$$p(\lambda|D) \propto \int p(D|\tau, \mu)\ p(\tau|\lambda)\ p(\lambda)\ p(\mu)\ d\tau\ d\mu$$

BEAST: Bayesian Evolutionary Analysis by Sampling Trees

Integration through Markov chain Monte Carlo



Integration through Markov chain Monte Carlo



Metropolis-Hastings algorithm

Starting from state θ propose a new state θ^*. For the following, this proposal must to symmetric, i.e. $Q(\theta \rightarrow \theta^*) = Q(\theta^* \rightarrow \theta)$

If new state is more likely, always accept. If new state is less likely, accept with probability proportional to ratio of new state to old state.

Acceptance probability: $\min\left(1, \frac{p(\theta^*)}{p(\theta)}\right)$

Simple example: $p(x) = 0.2$ $p(y) = 0.8$

$A(x \rightarrow y) = 0.8/0.2 = 1$ $A(y \rightarrow x) = 0.2/0.8 = 0.25$

Mass moving from x to y: $p(x)\,A(x \rightarrow y) = 0.2 \times 1 = 0.2$

Mass moving from y to x: $p(y)\,A(y \rightarrow x) = 0.8 \times 0.25 = 0.2$

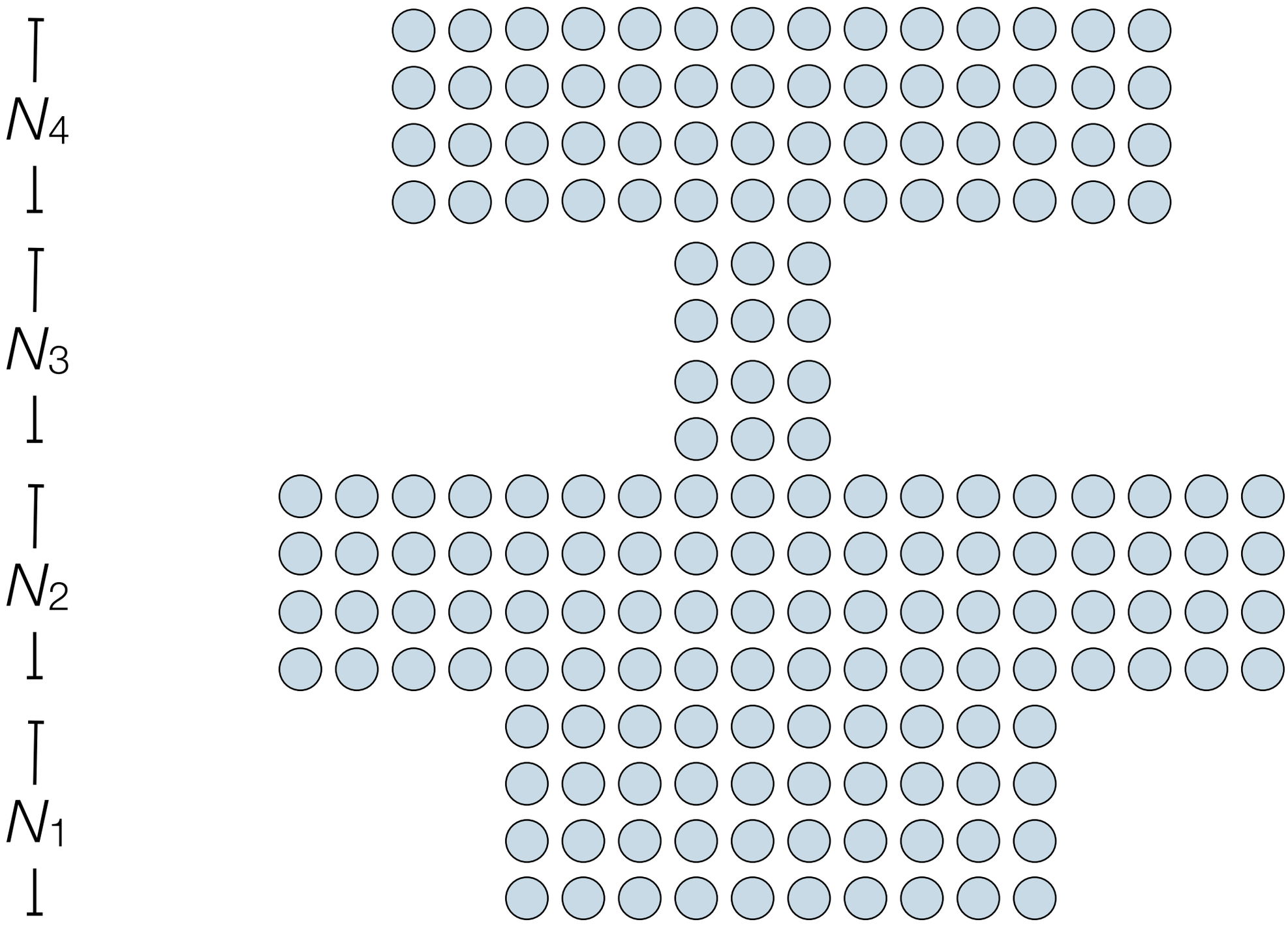
BEAST will produce samples from:

λ – coalescent model

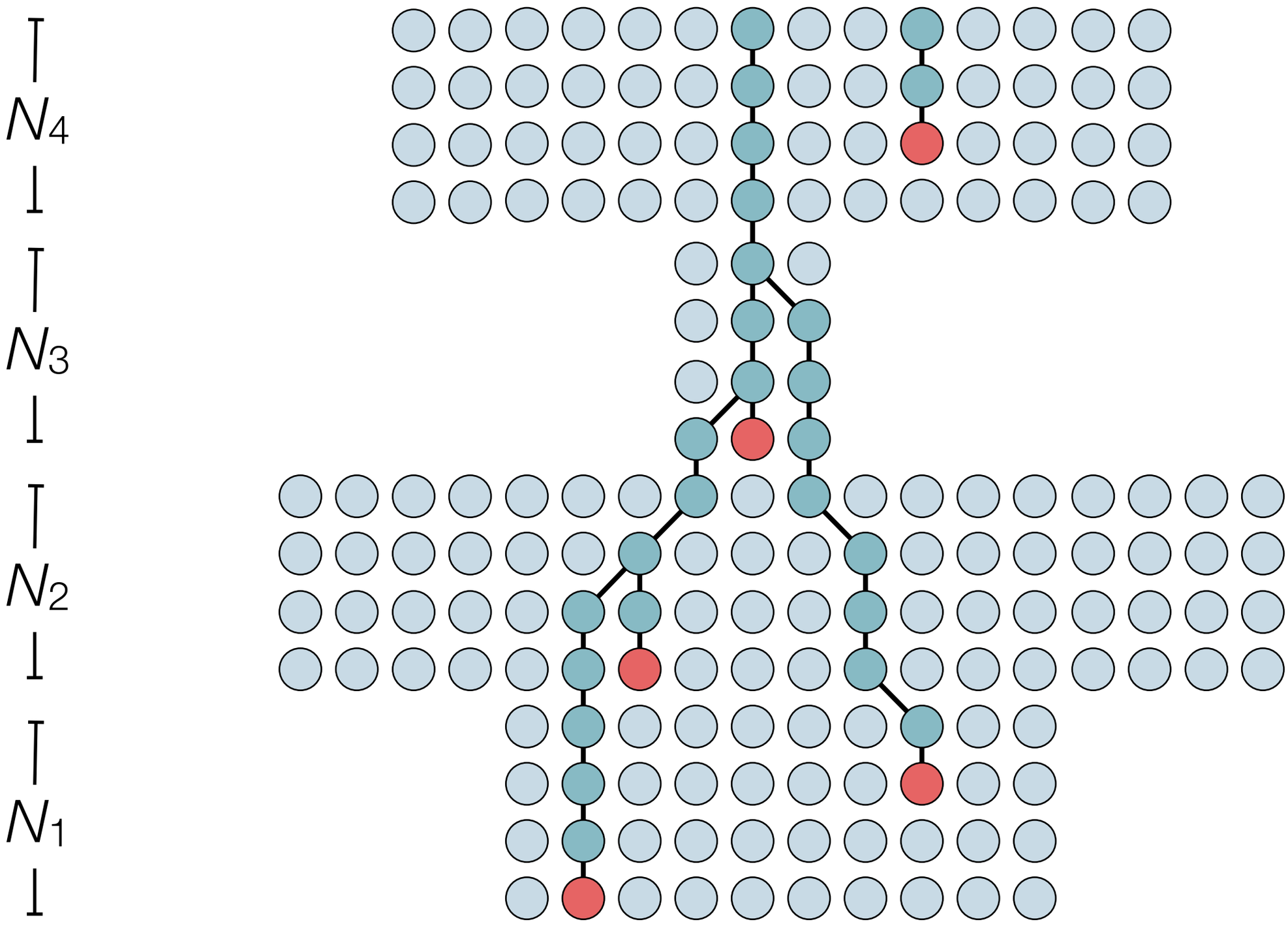
μ – mutation model

τ – phylogeny

Use a ‘skyline’ demographic model

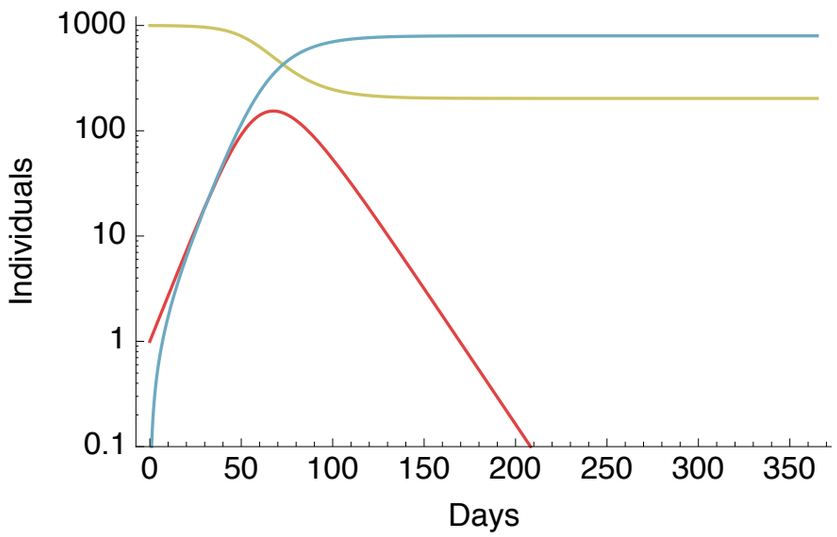


Use a ‘skyline’ demographic model



Practical part 1

Estimating R_0 from timeseries data



$$r(0) = \beta - \gamma$$

r = 0.20 per day for 1918 influenza

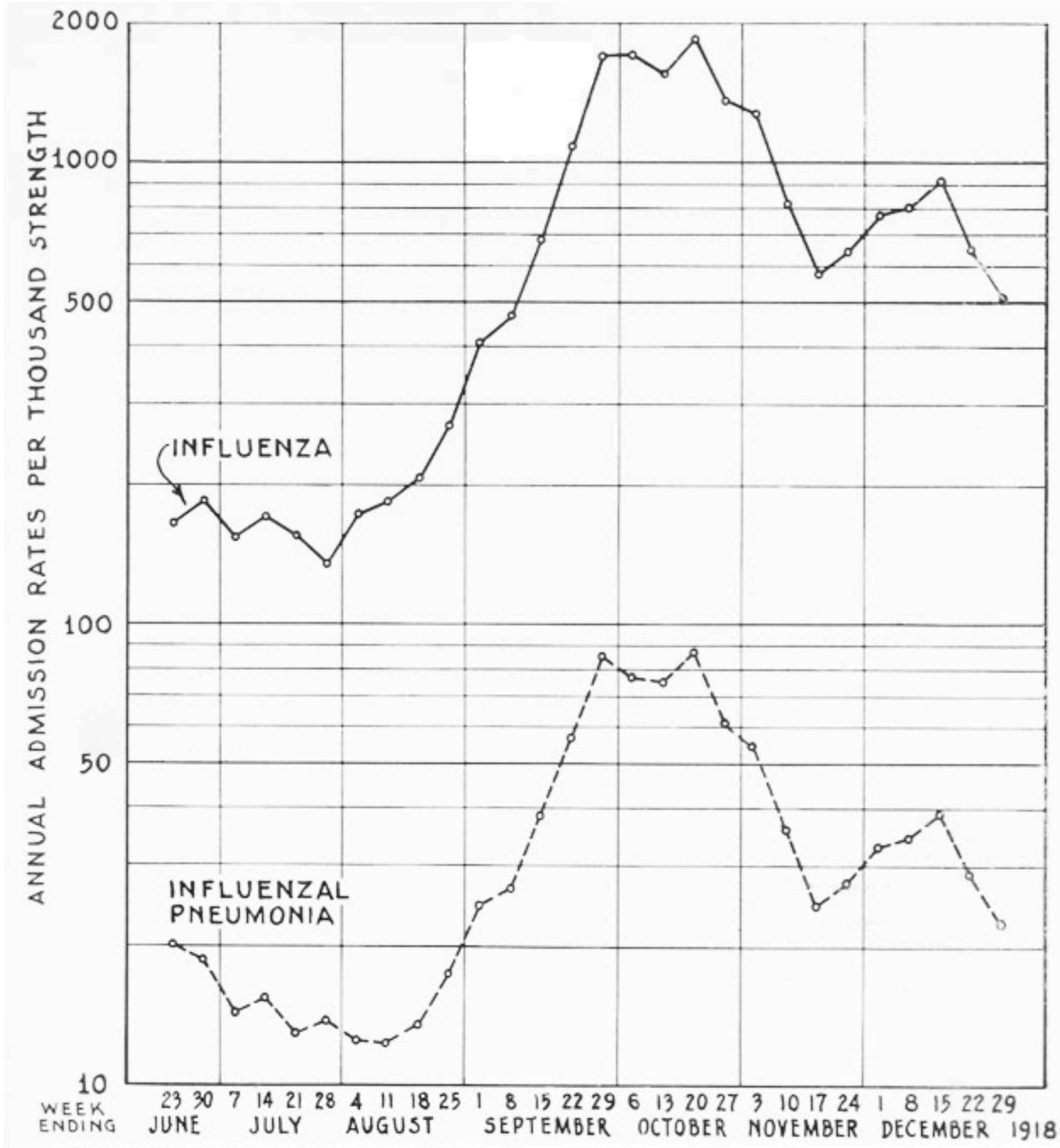
We know the approximate recovery rate

$$\gamma \approx 0.25$$

We can solve for β and hence R_0

$$\beta = r + \gamma \approx 0.45$$

$$R_0 = \frac{\beta}{\gamma} \approx \frac{0.45}{0.25} \approx 1.8$$

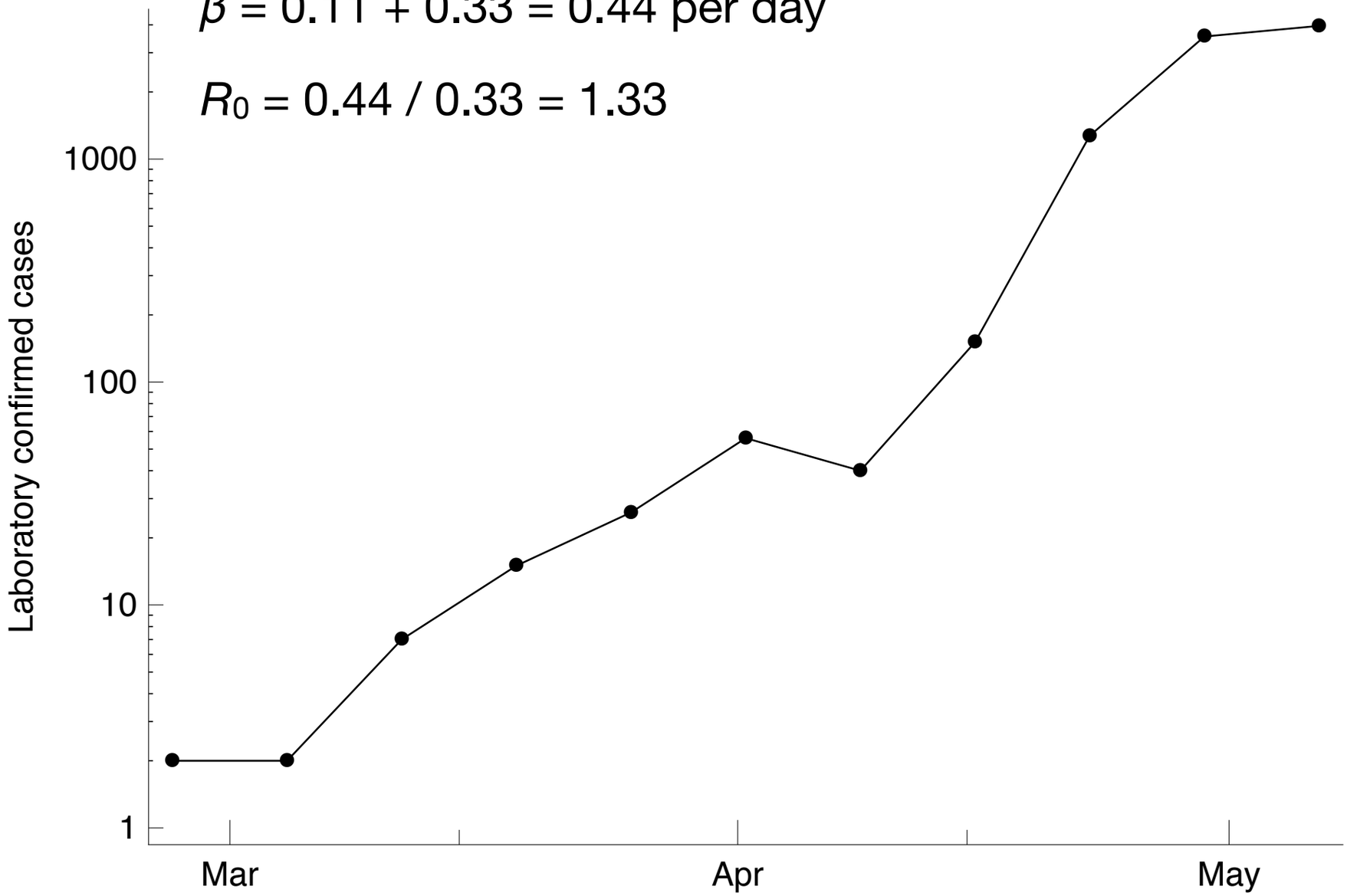


Growth rate of pandemic H1N1

r = 0.11 per day

β = 0.11 + 0.33 = 0.44 per day

R_0 = 0.44 / 0.33 = 1.33



Generation time τ of infection

At the beginning of the epidemic, new infections emerge at rate β.

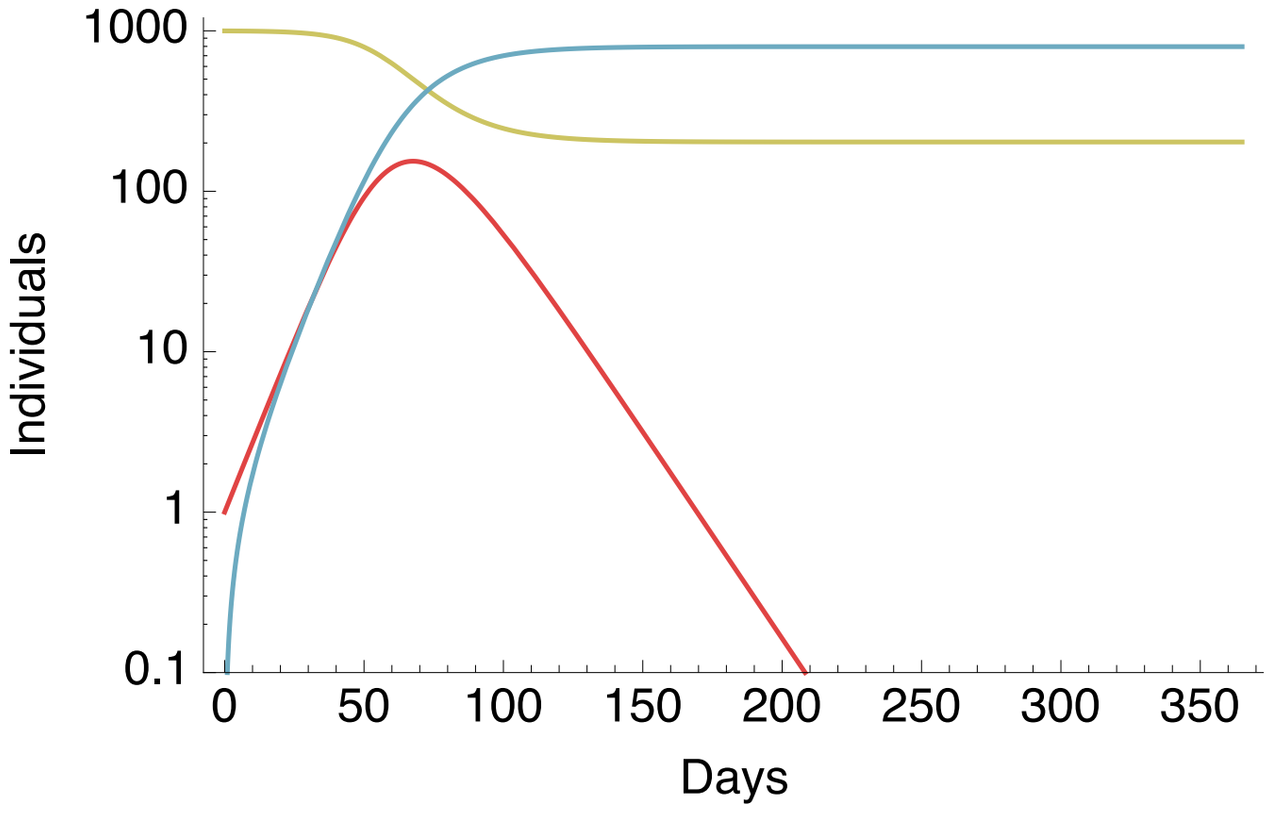
$$\tau = \frac{1}{2\beta S(0)} = \frac{1}{2 \times 0.36} = 1.39$$

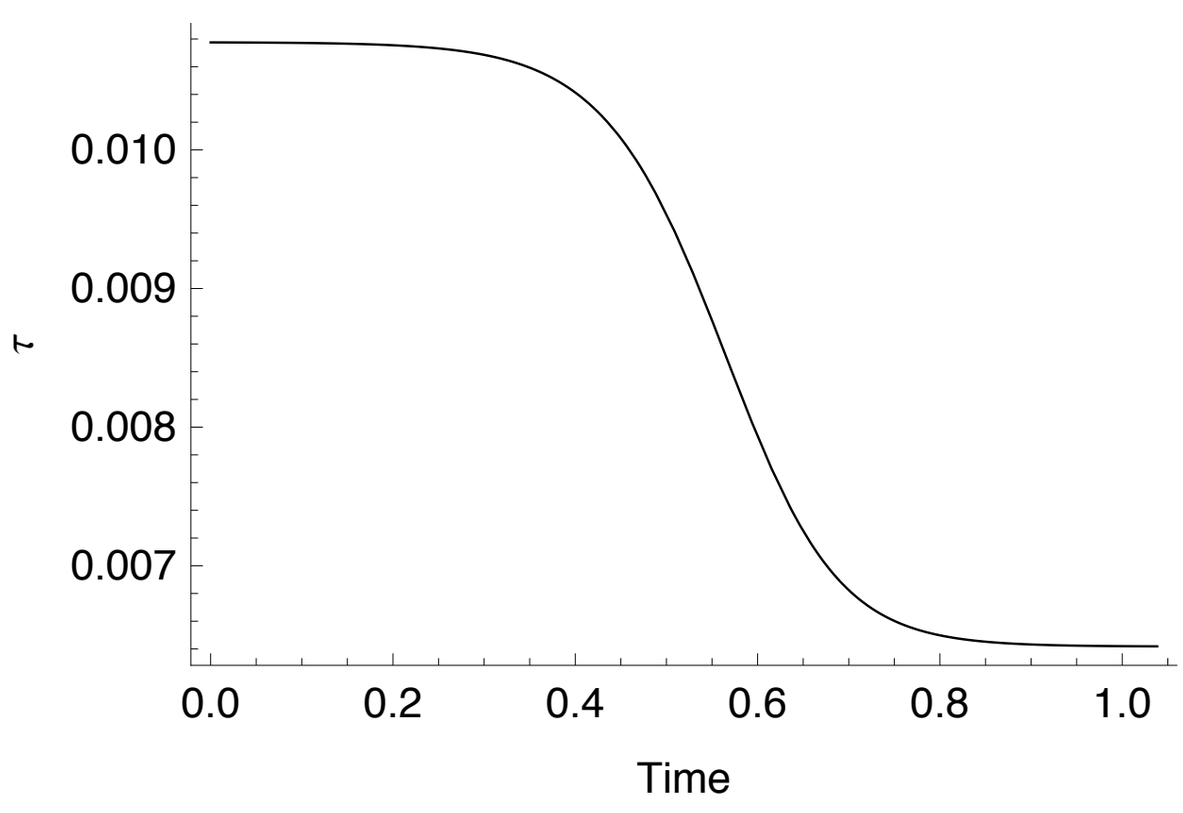
Final susceptible fraction:

$$S(\infty) = e^{-R_0(1-S(\infty))}$$

At the end of the epidemic:

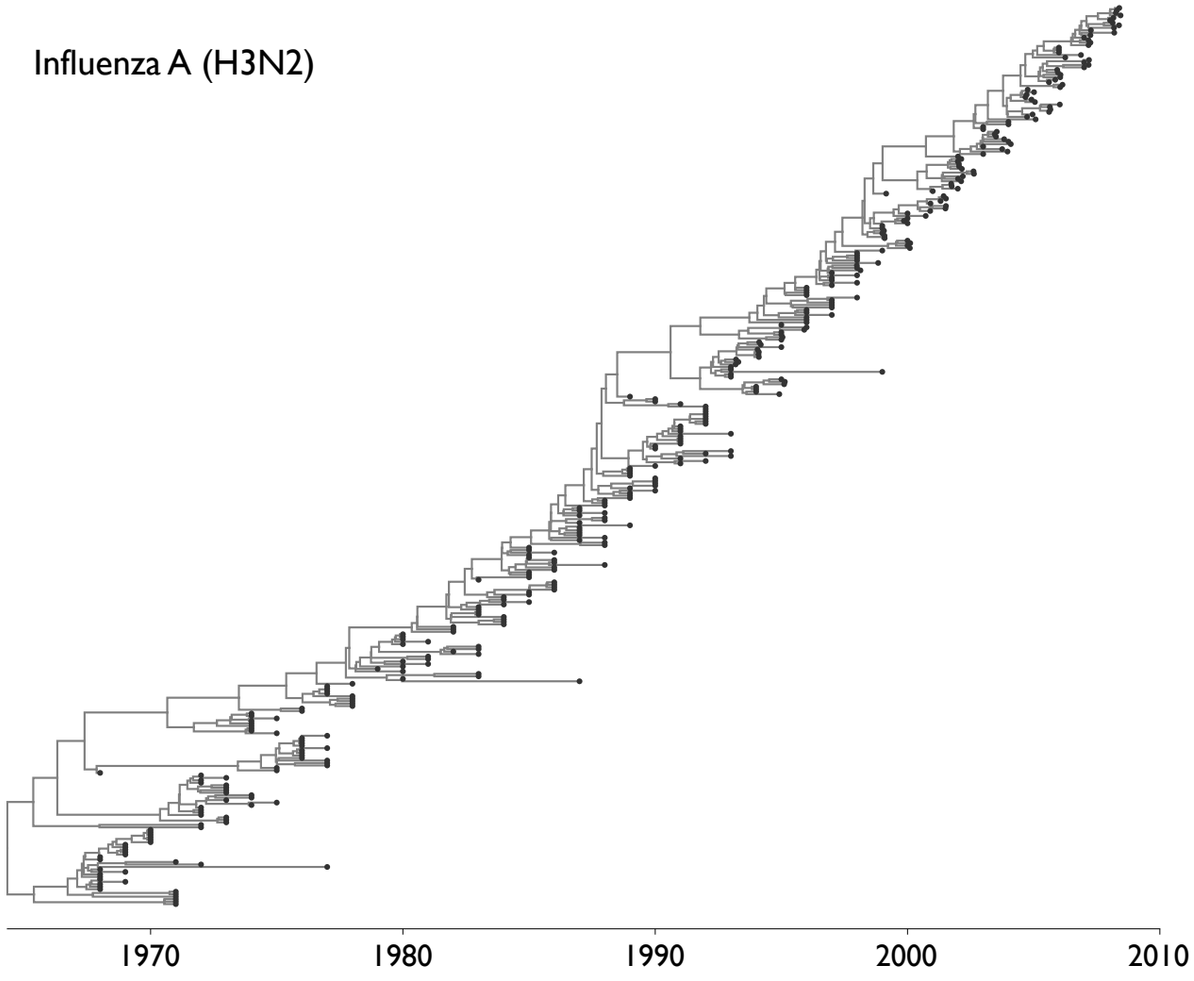
$$\tau = \frac{1}{2\beta S(\infty)} = \frac{1}{2 \times 0.36 \times 0.84} = 1.65$$





Effective population sizes of flu vs measles

Influenza A (H3N2)



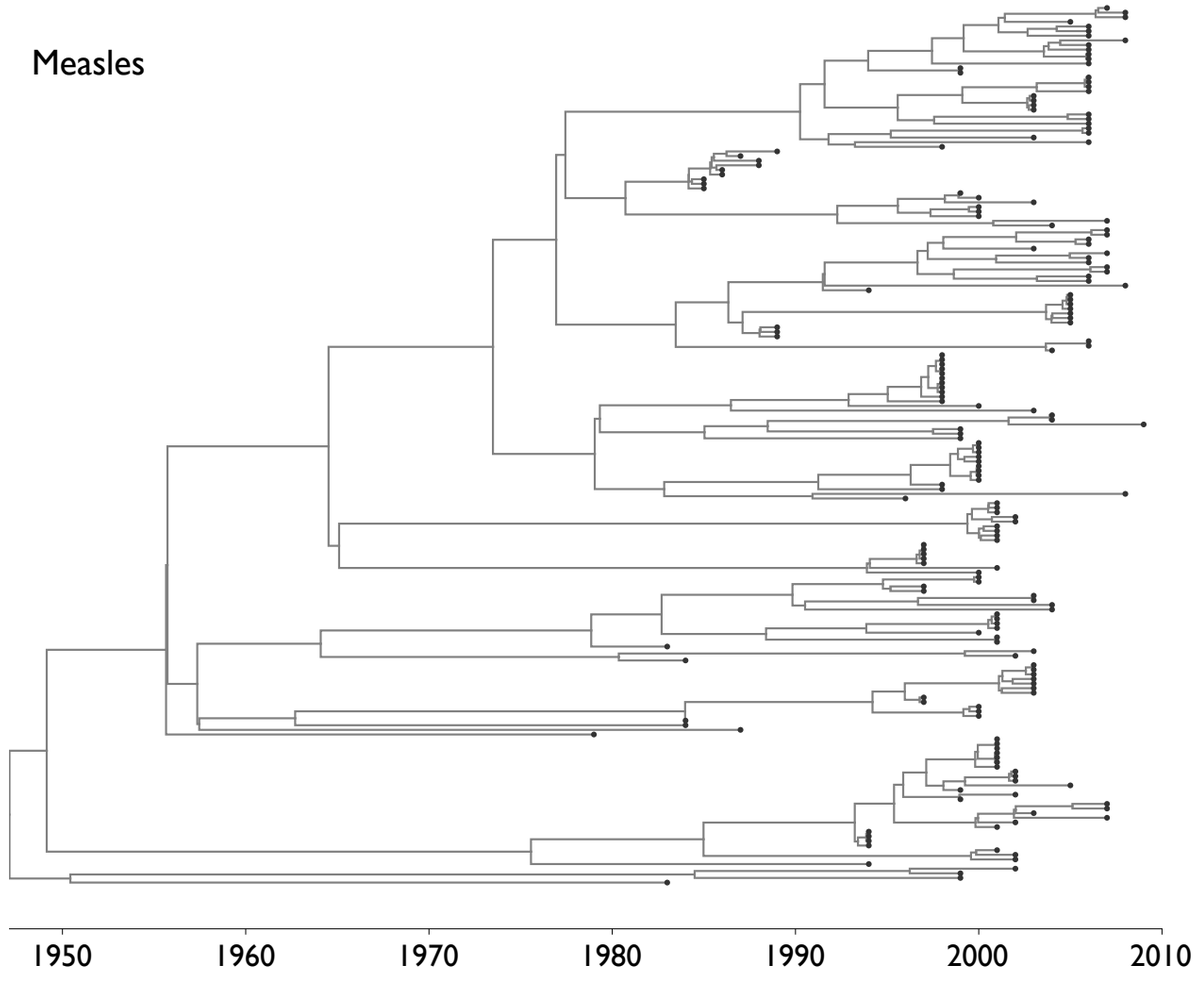
N_e = 7.2 years

N_e = 1050 infections (duration of infection of 5 days)

N = 70 million infections (prevalence)

Off by a factor of 6,700

Measles



N_e = 124.6 years

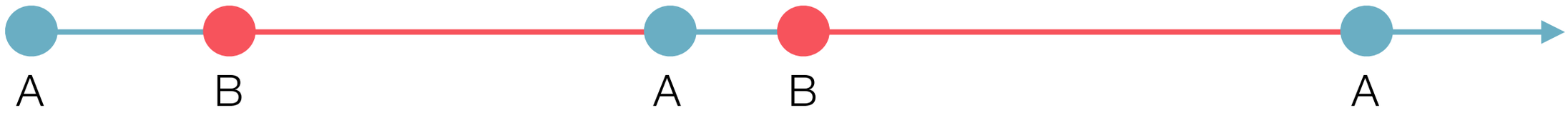
N_e = 8270 infections (duration of infection of 11 days)

N = 0.9 million infections (prevalence)

Off by a factor of 110

Practical part 2

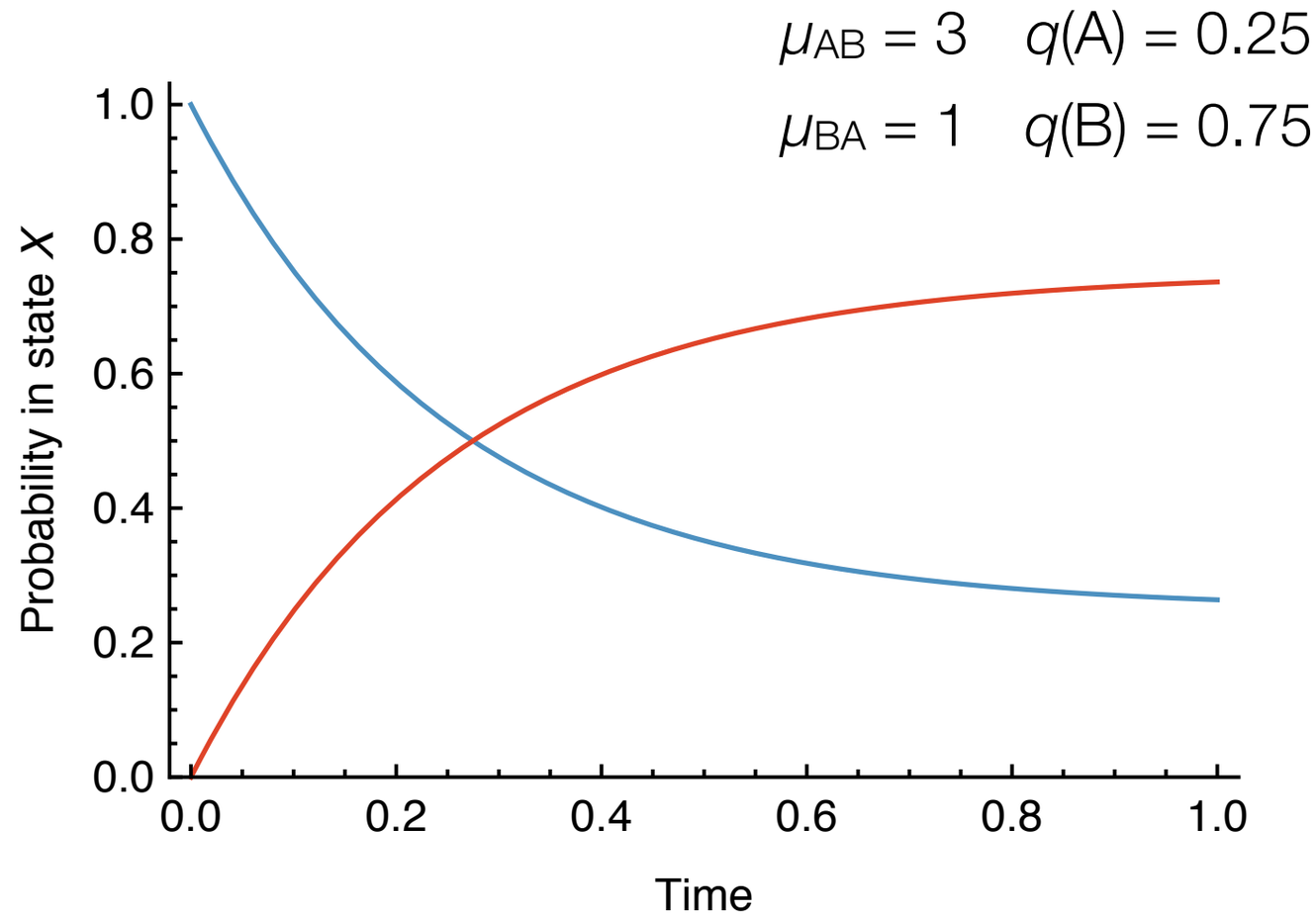
Continuous time Markov chains (CTMCs)



	A	B
A		μ_{AB}
B	μ_{BA}	

$$p_{t\to\infty}(A) = \frac{\mu_{BA}}{\mu_{AB} + \mu_{BA}}$$

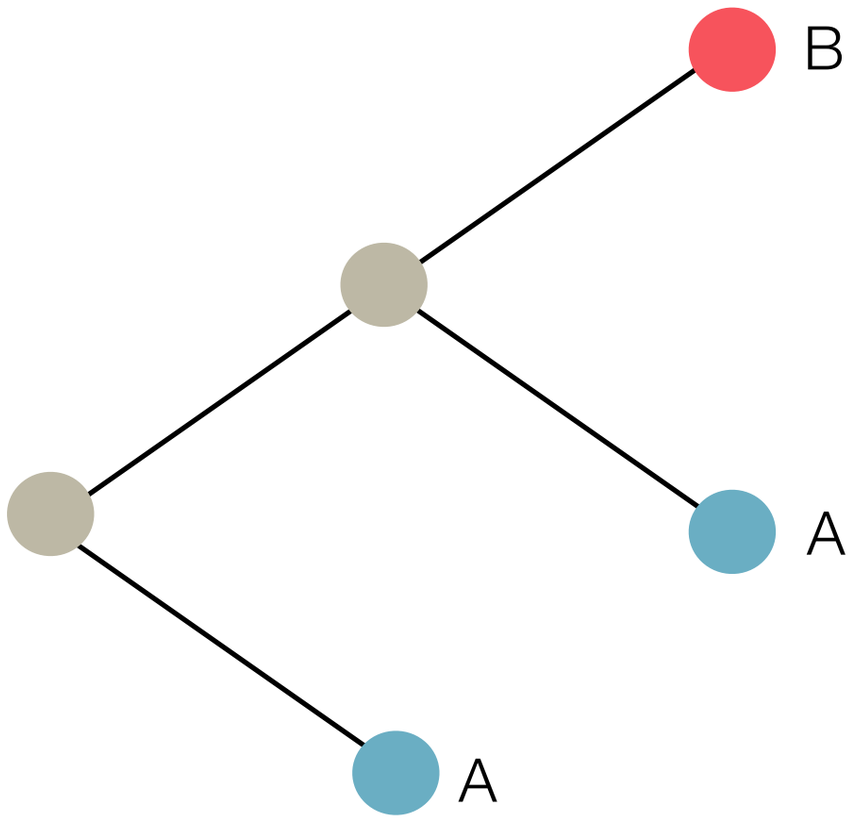
$$p_{t\to\infty}(B) = \frac{\mu_{AB}}{\mu_{AB} + \mu_{BA}}$$



CTMCs on trees

Transition matrix with $\mu_{AB} = 3$ $\mu_{BA} = 1$ $t = 0.2$

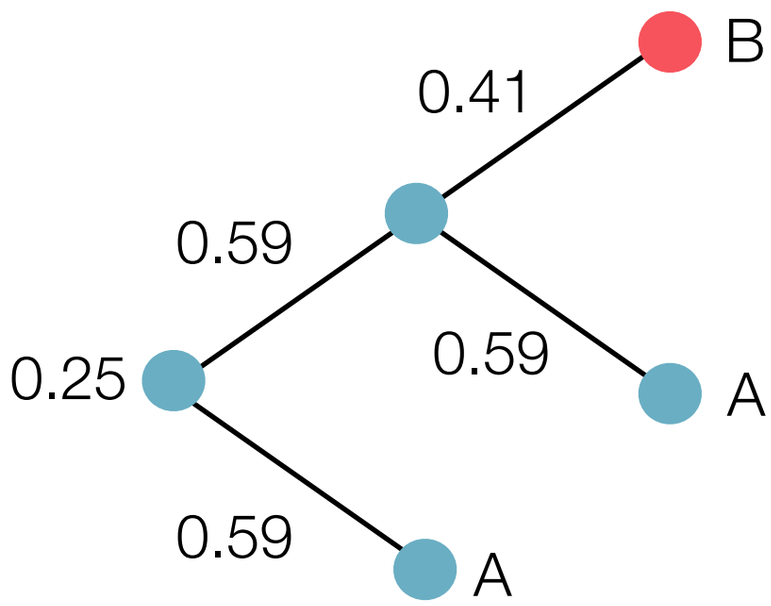
	A	B
A	0.59	0.41
B	0.14	0.86

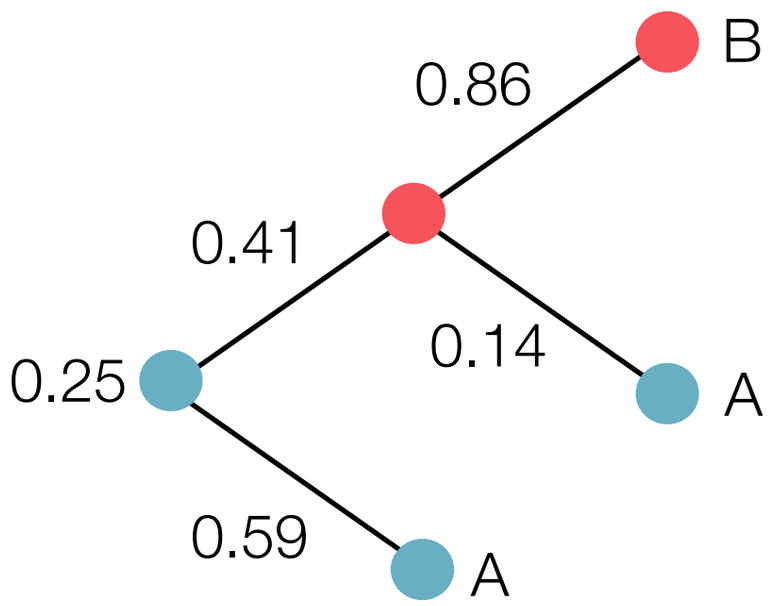


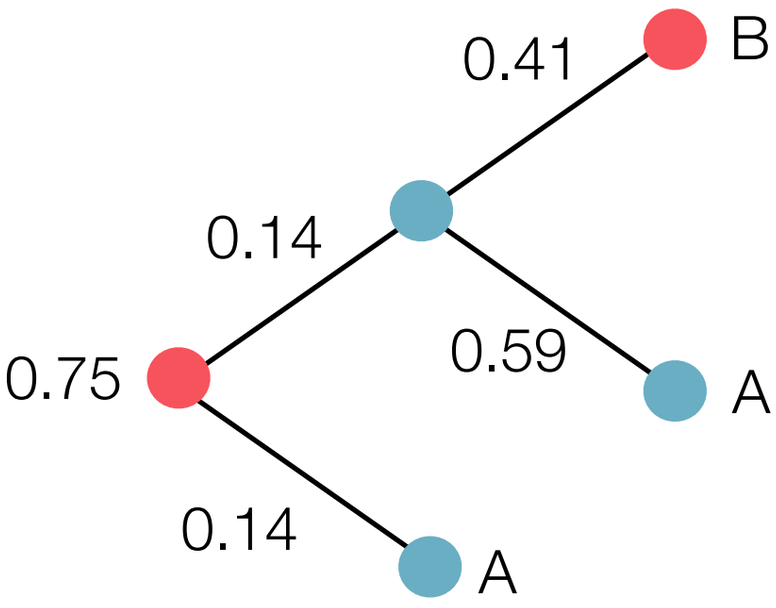
Integrate over internal states

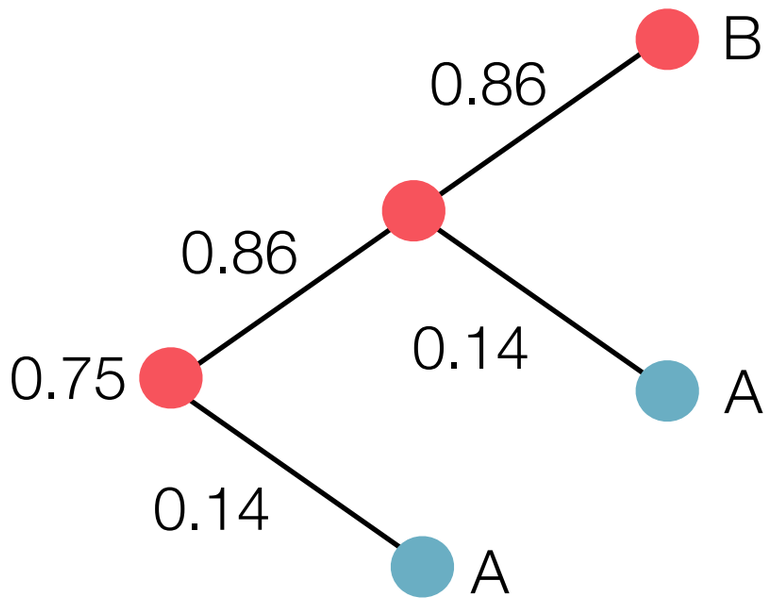
Transition matrix with $\mu_{AB} = 3$ $\mu_{BA} = 1$ $t = 0.2$

	A	B
A	0.59	0.41
B	0.14	0.86



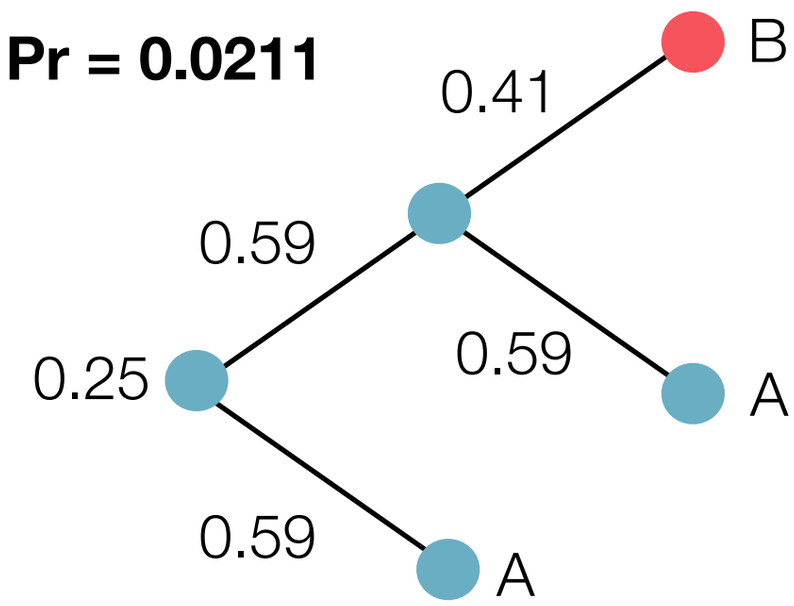


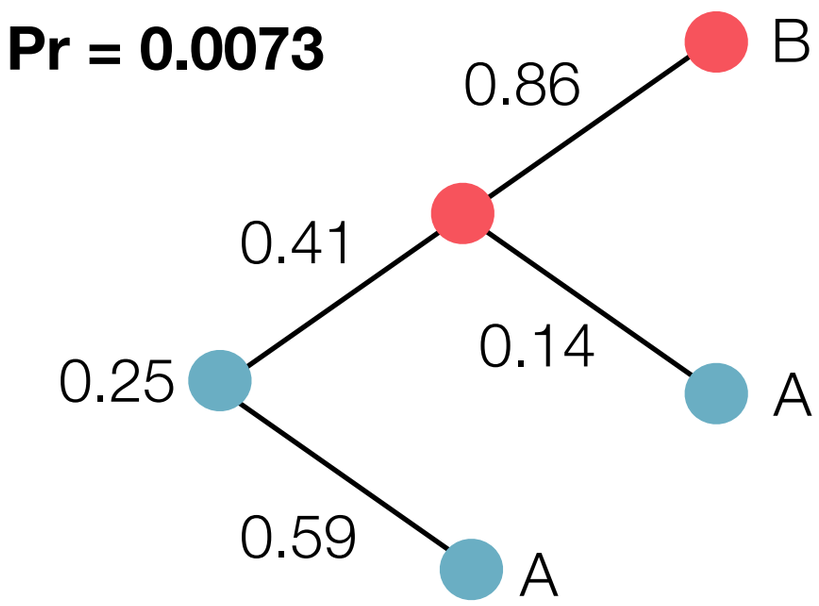


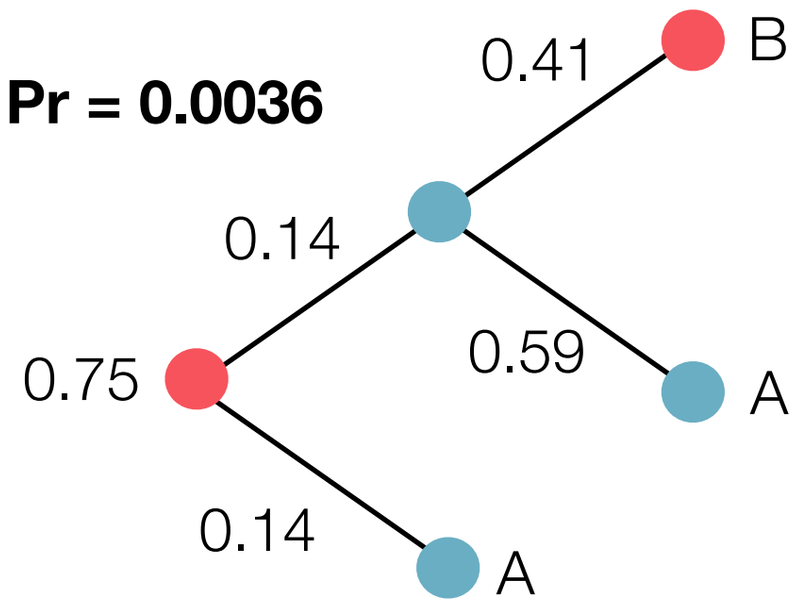


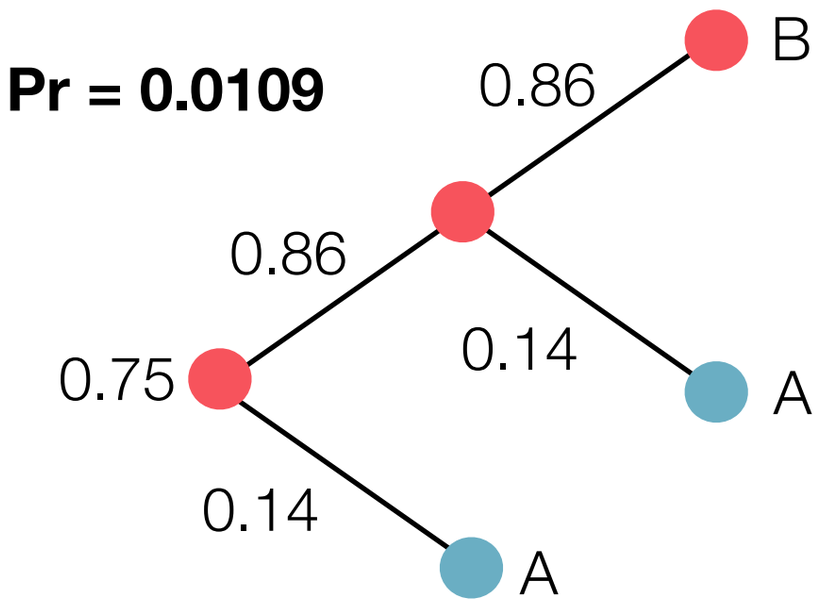
Integrate over internal states

Transition matrix with $\mu_{AB} = 3$ $\mu_{BA} = 1$ $t = 0.2$





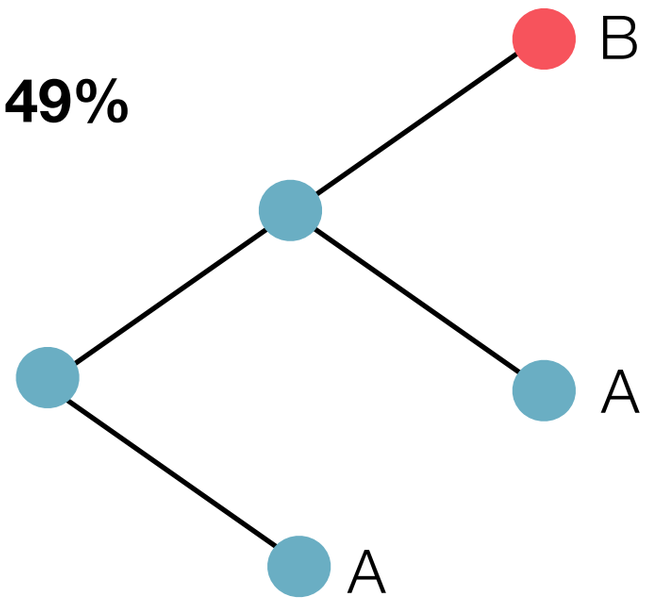




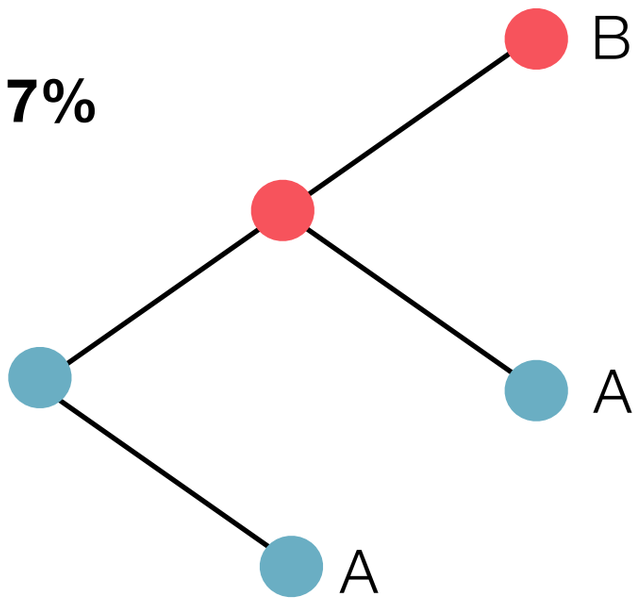
Integrate over internal states

$$p(D|\tau, \mu) = 0.0211 + 0.0073 + 0.0036 + 0.0109 = 0.0429$$

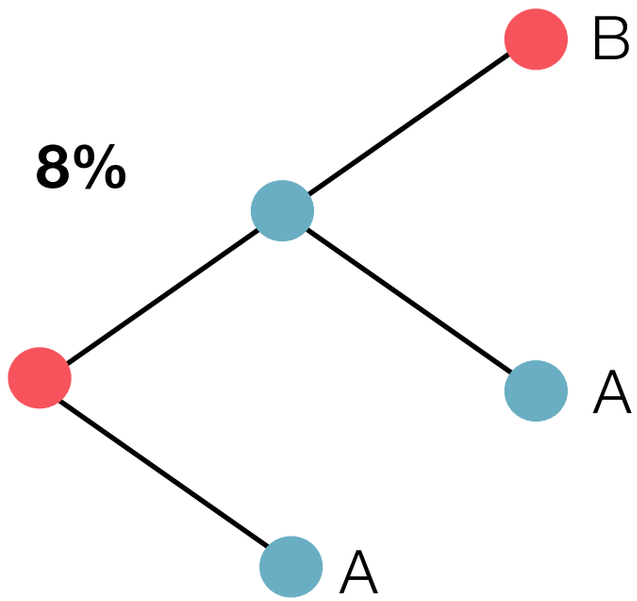
49%



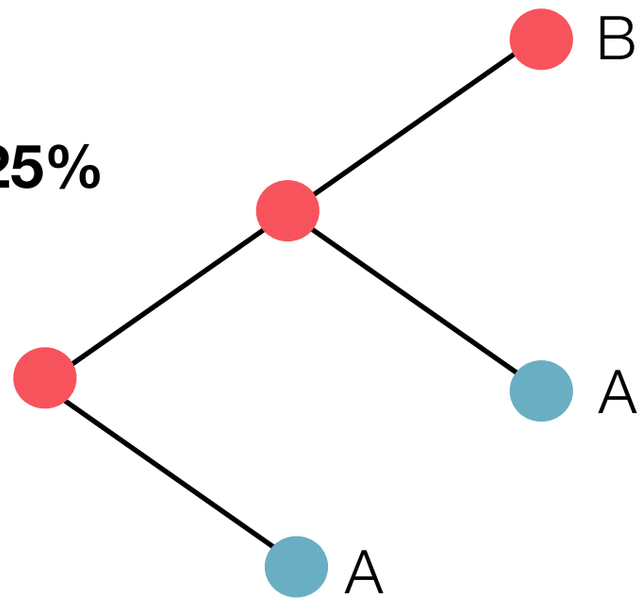
17%



8%



25%



Practical part 3