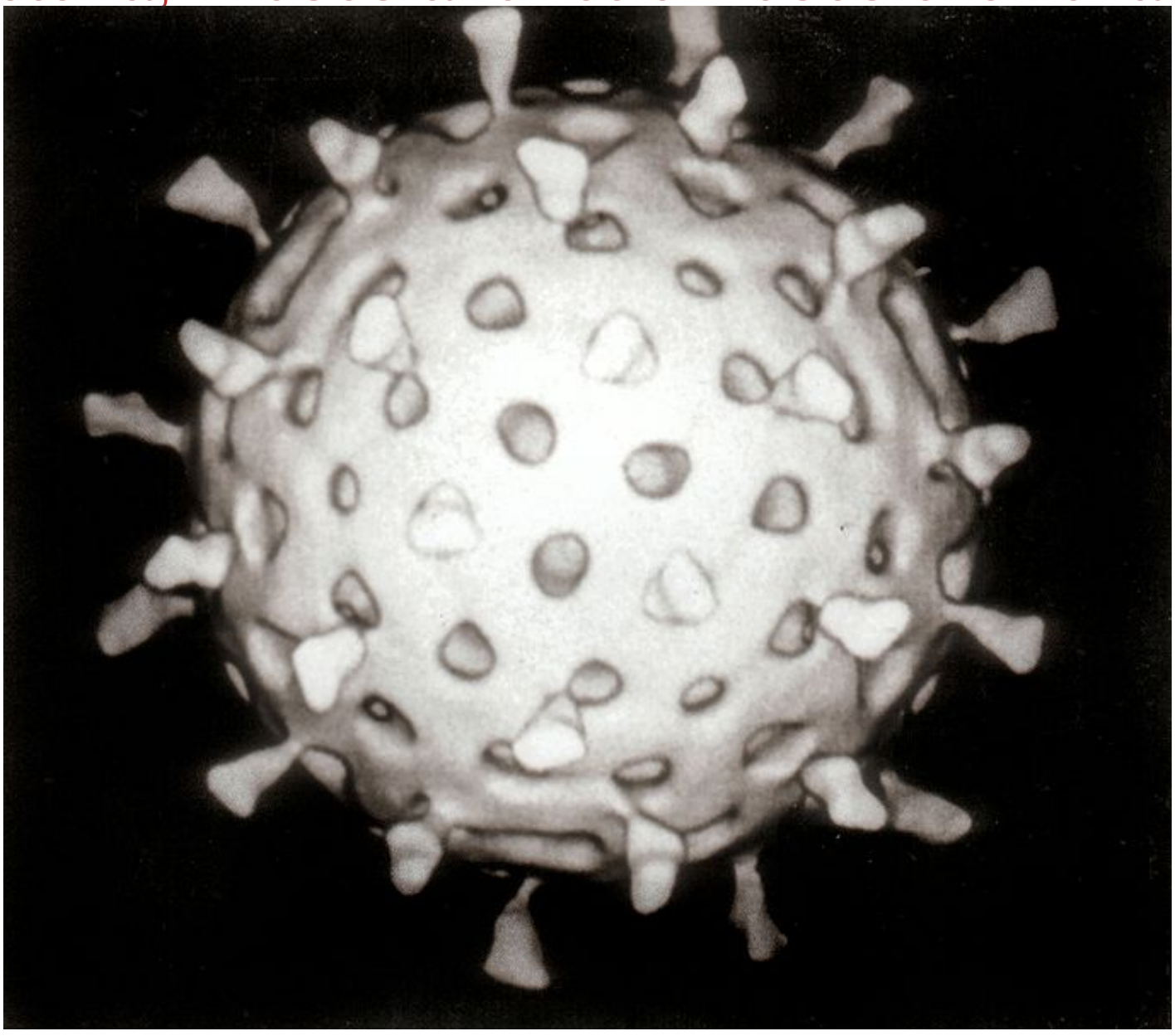
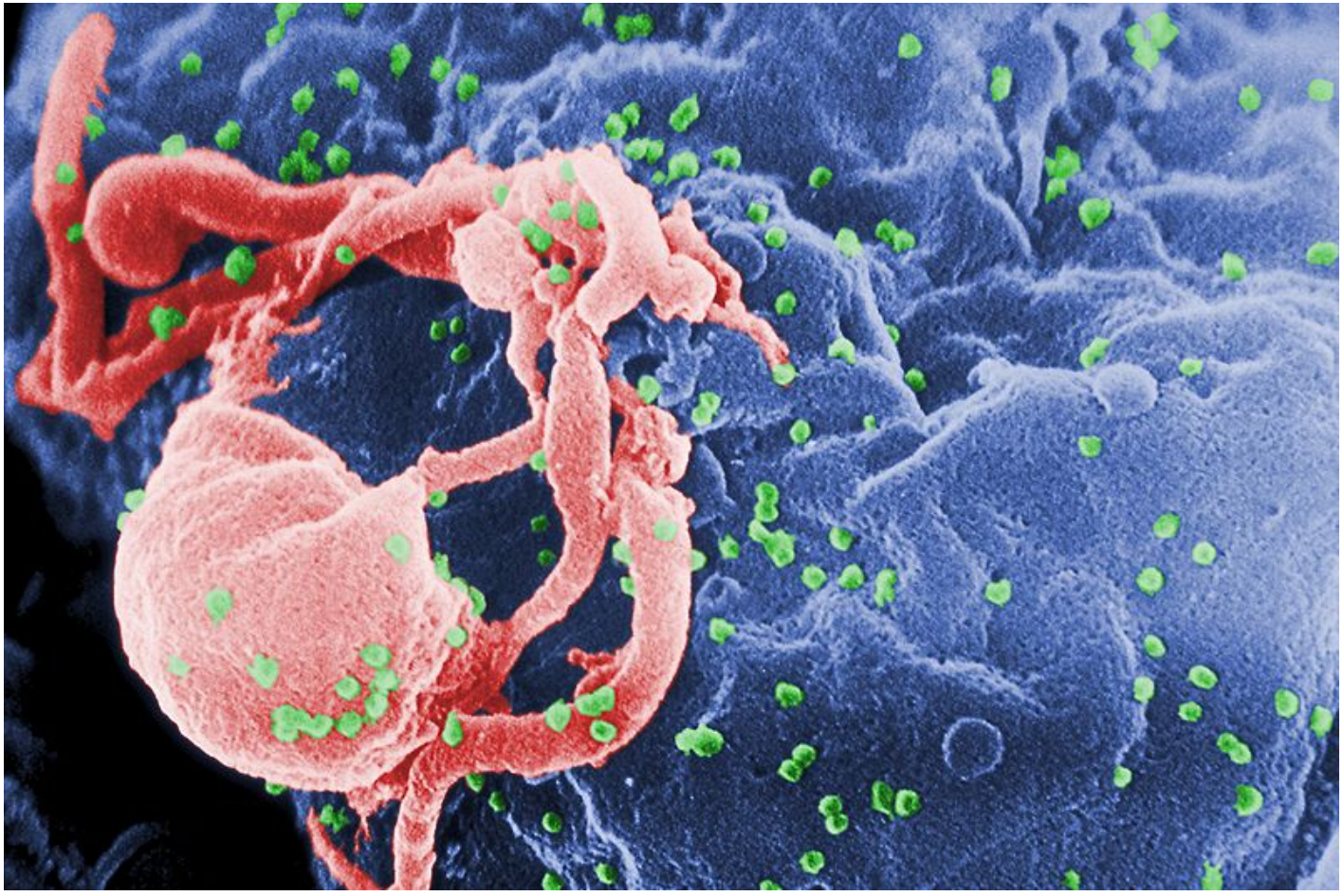
Bacteria, viruses and retroviruses evolve fast



Rate of evolution of HIV is maybe 6 orders of magnitude greater than in humans



See as much evolution in an HIV in a few years as in entire human history

Figure S3 (a) PB2 segment

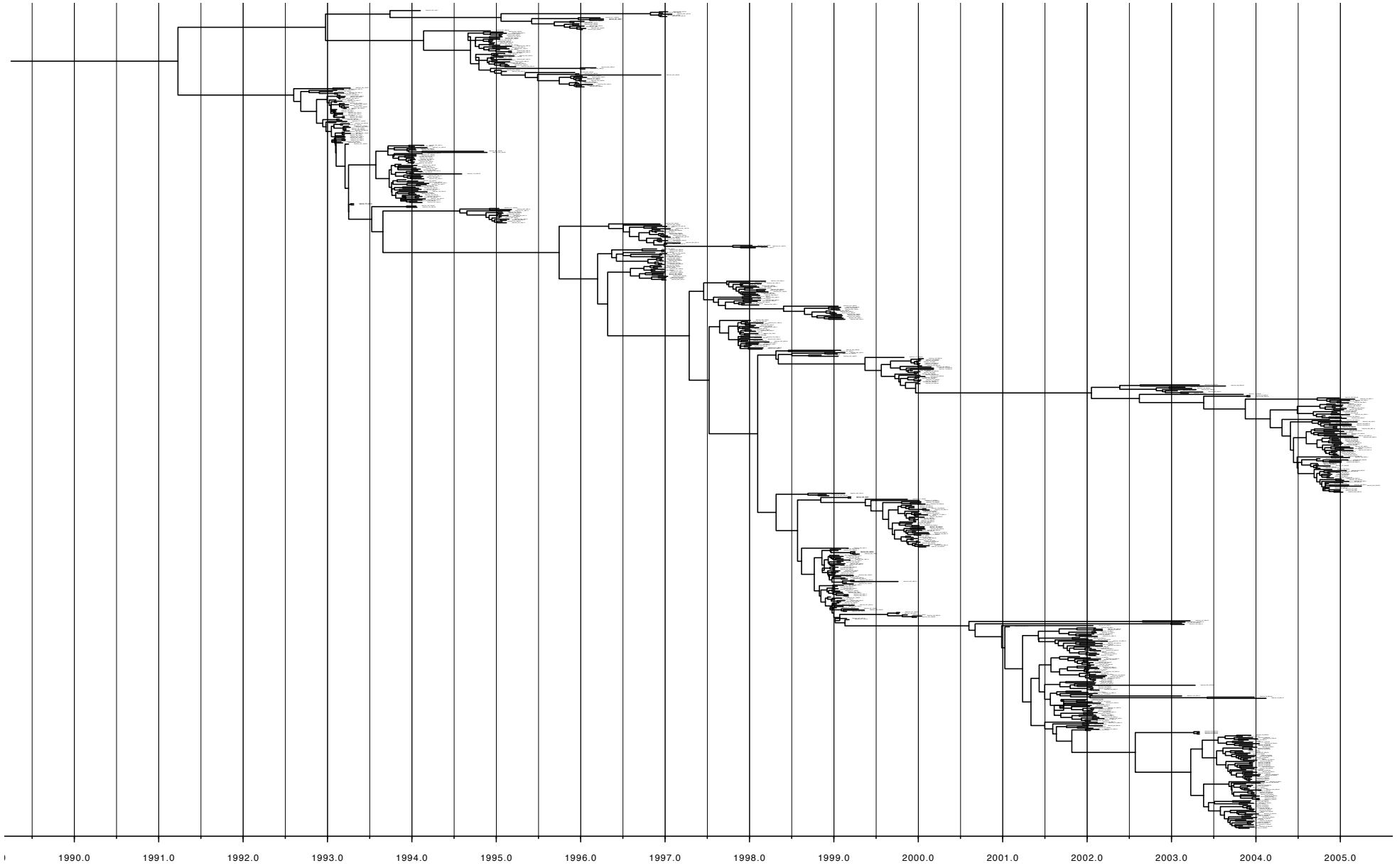
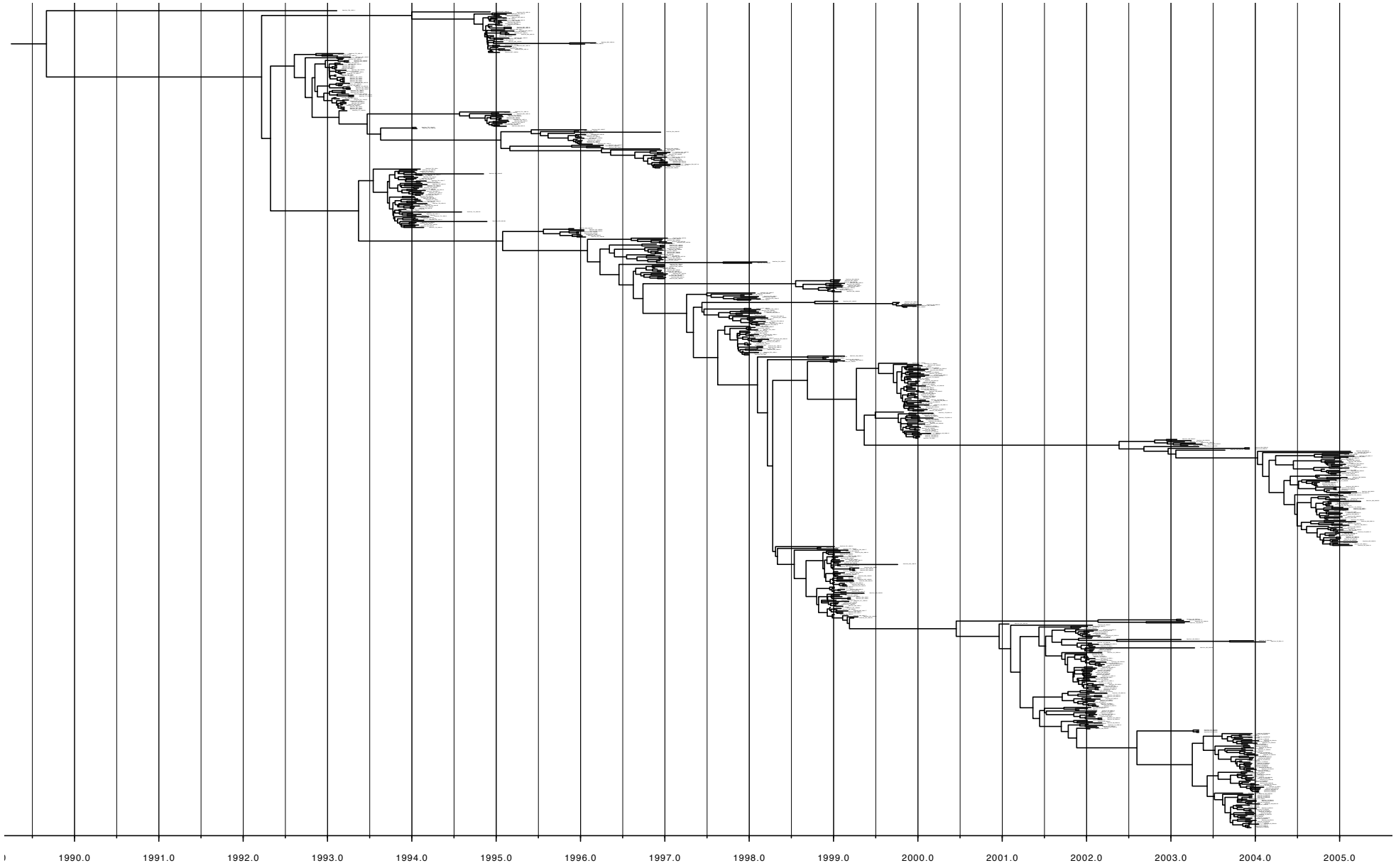
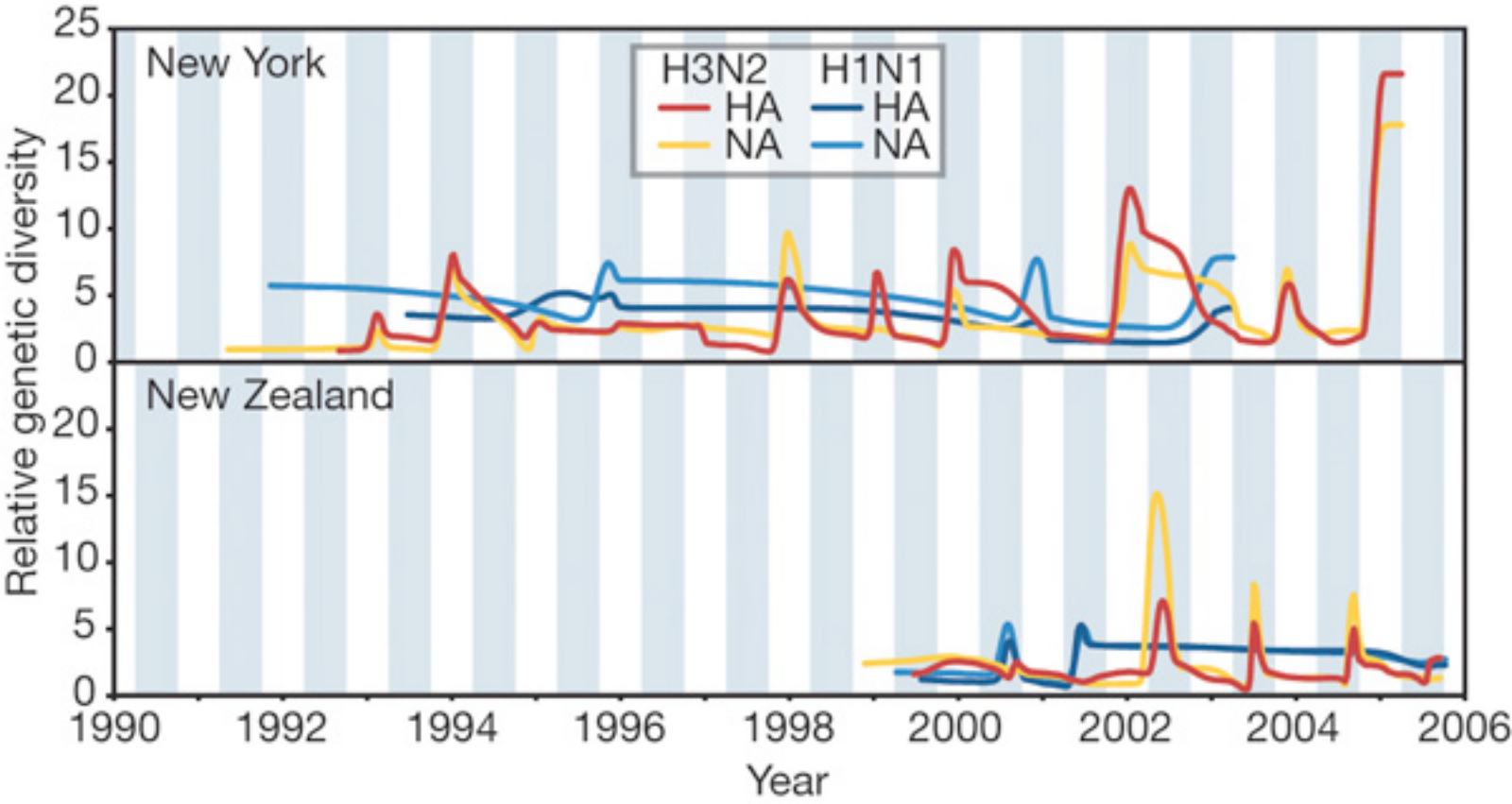
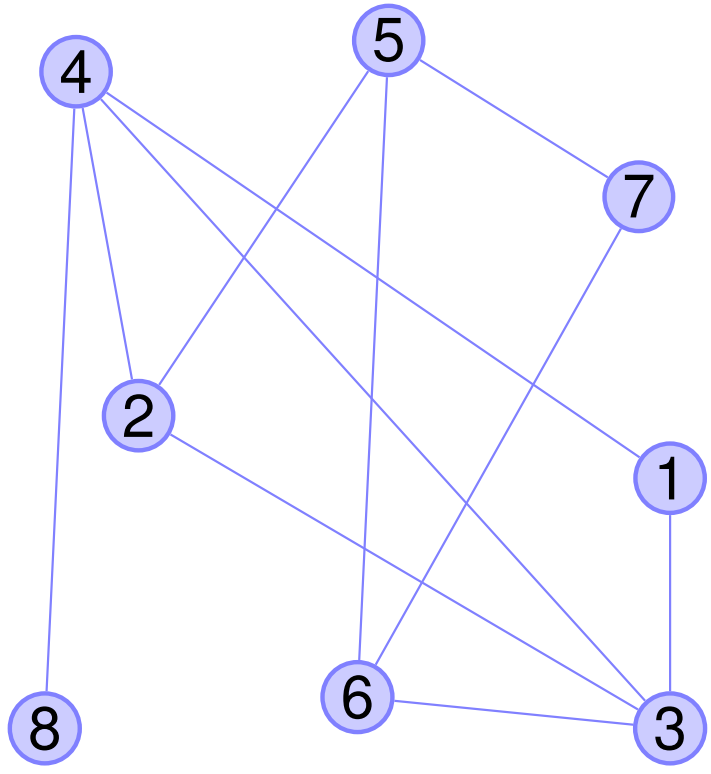


Figure S3 (b) PB1 segment



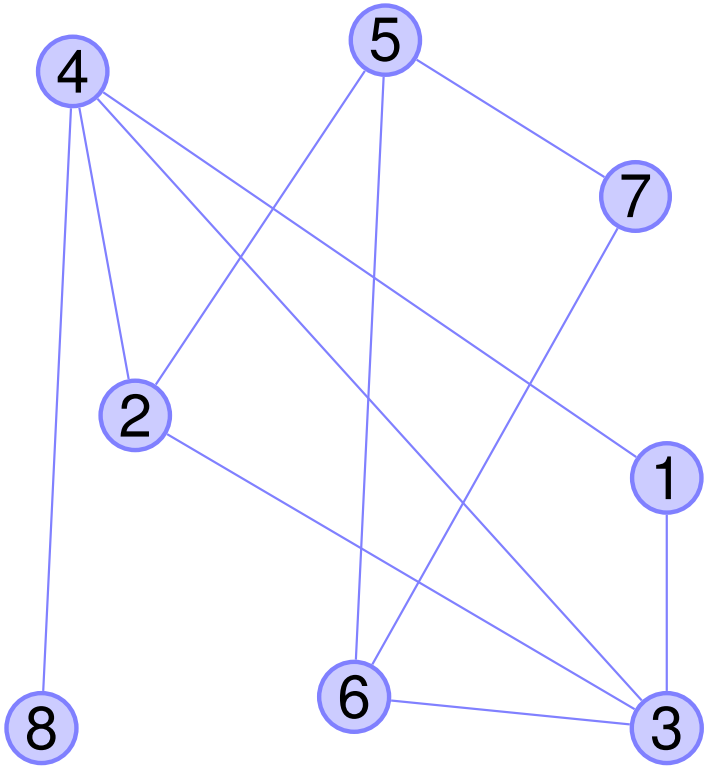

New York
New Zealand
H3N2
HA
NA
H1N1
HA
NA
Relative genetic diversity
0
5
10
15
20
25
1990
1992
1994
1996
1998
2000
2002
2004
2006
Year

Networks are a natural way to model heterogeneity of contacts among hosts



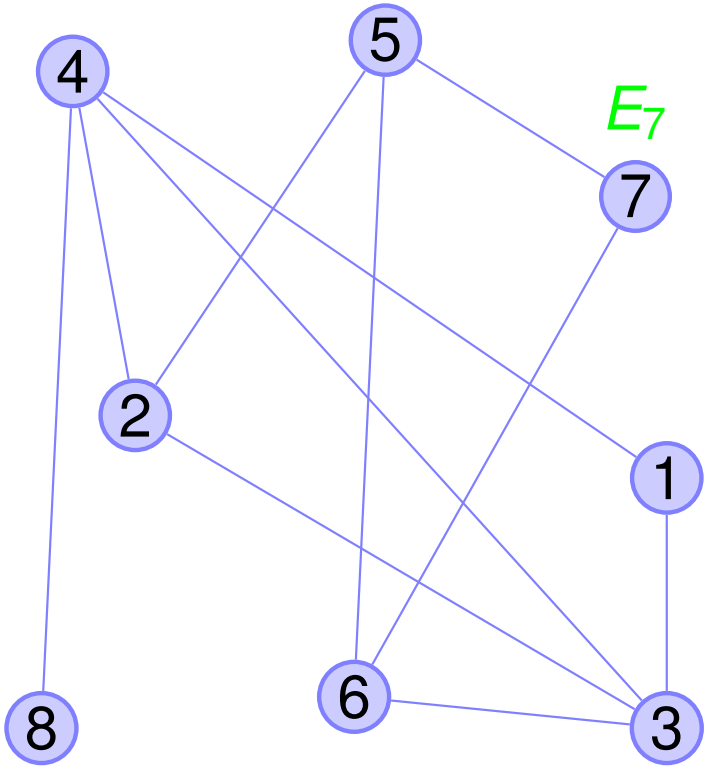
- Nodes represent hosts, edges represent contacts
- We want to use available data to estimate good models of networks

Contact networks and transmission trees



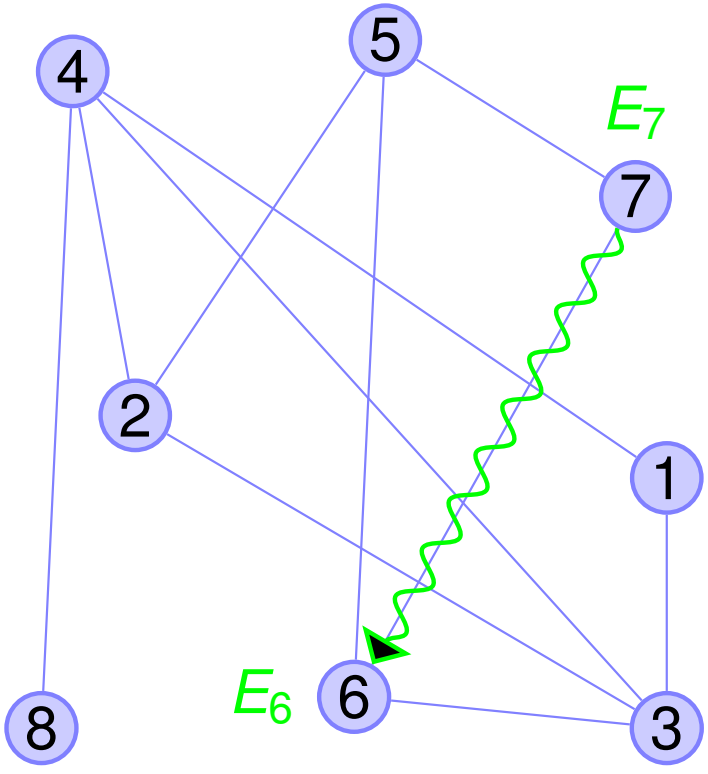
- Assume a *contact network* $\mathcal{G}$ on N individuals: A “contact” is necessary for disease transmission.

Contact networks and transmission trees



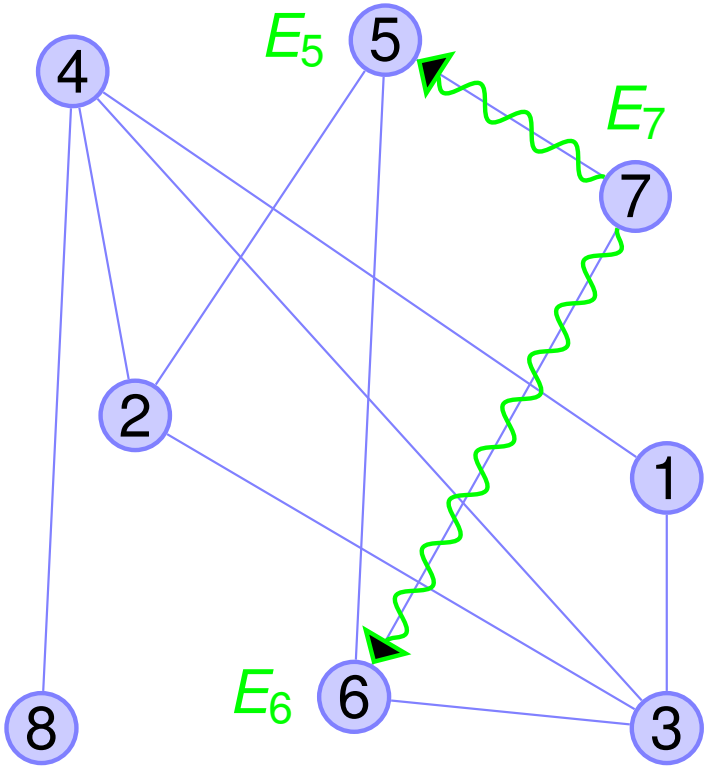
- Assume a *contact network* $\mathcal{G}$ on N individuals: A “contact” is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Contact networks and transmission trees



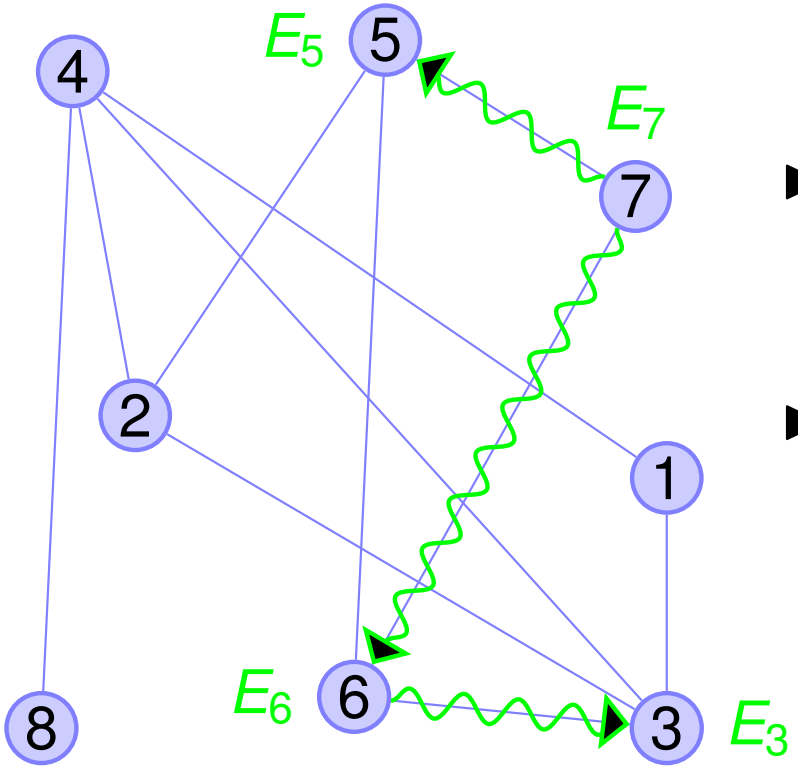
- Assume a *contact network* $\mathcal{G}$ on N individuals: A "contact" is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Contact networks and transmission trees



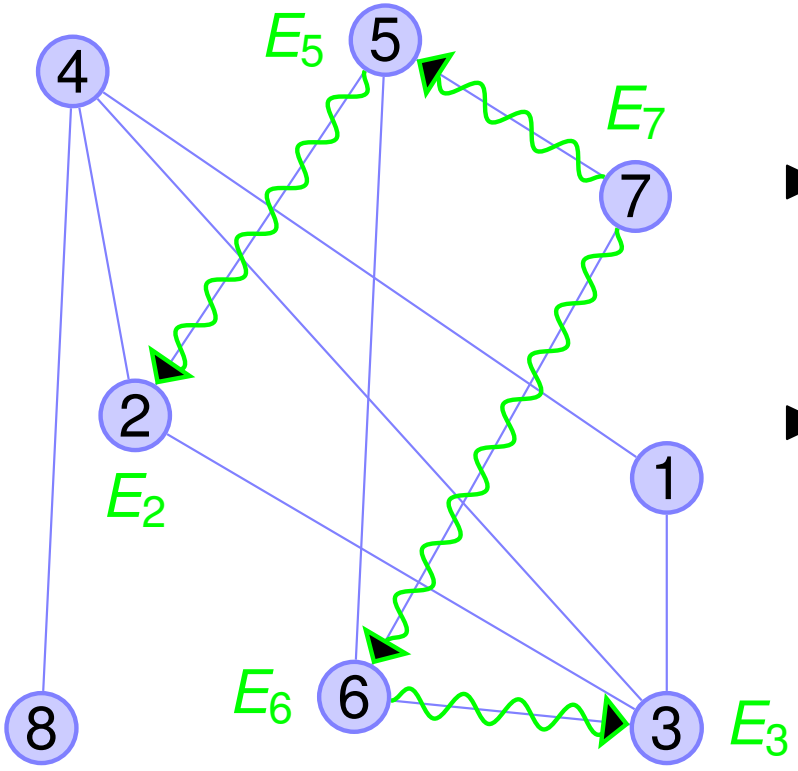
- Assume a *contact network* $\mathcal{G}$ on N individuals: A “contact” is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Contact networks and transmission trees



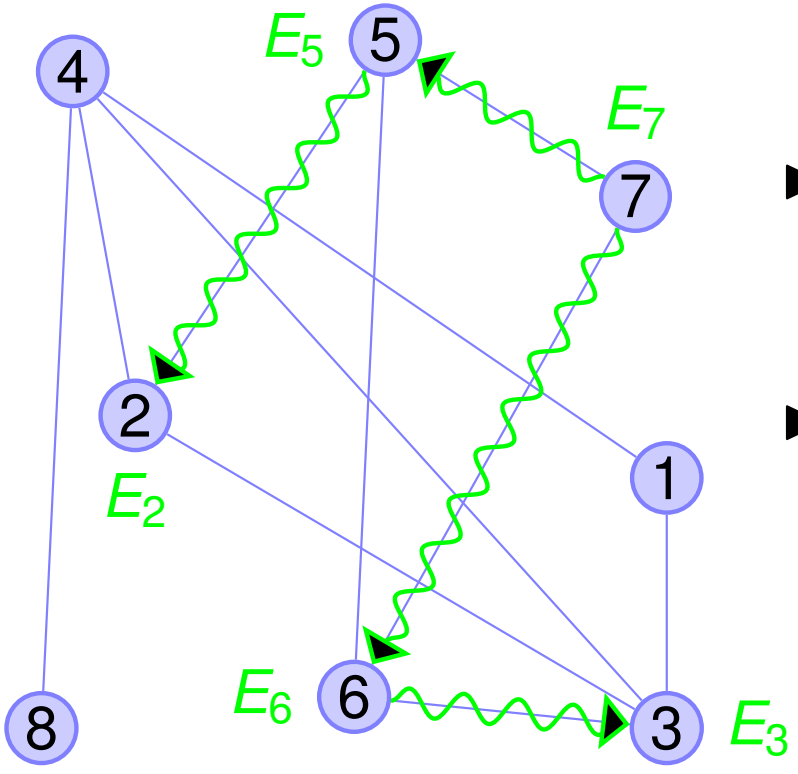
- Assume a *contact network* $\mathcal{G}$ on N individuals: A "contact" is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Contact networks and transmission trees



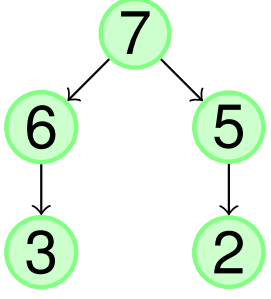
- Assume a *contact network* $\mathcal{G}$ on N individuals: A "contact" is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Contact networks and transmission trees



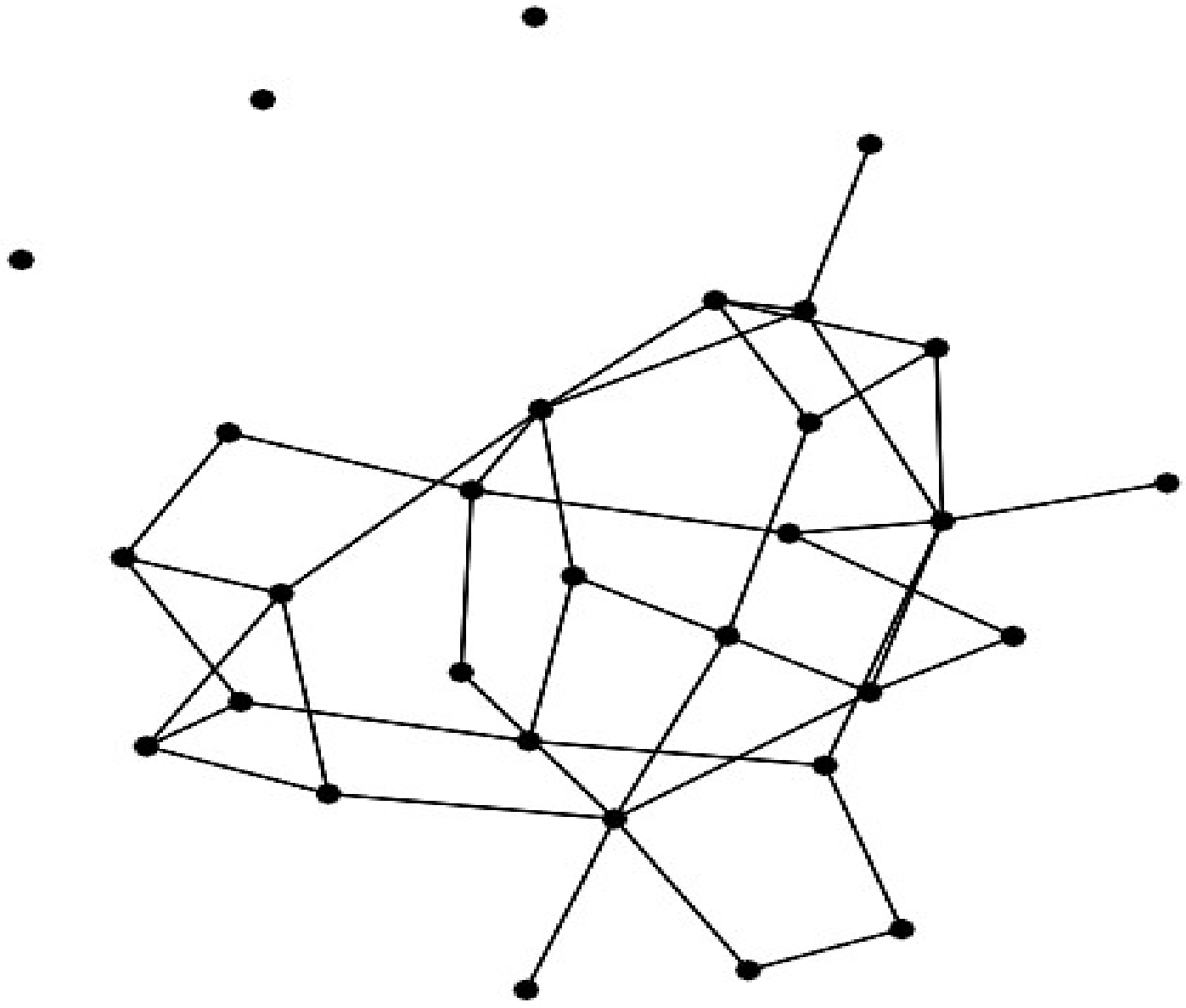
- Assume a *contact network* $\mathcal{G}$ on N individuals: A "contact" is necessary for disease transmission.
- Beginning with the first infected, disease is spread at exponential rate β, defining a subtree of the contact network called the *transmission tree* $\mathcal{P}$.

Transmission tree, $\mathcal{P}$:



A network model: Bernoulli (Erdős-Rényi) network

For any nodes i and j, edge (i, j) exists with probability p.



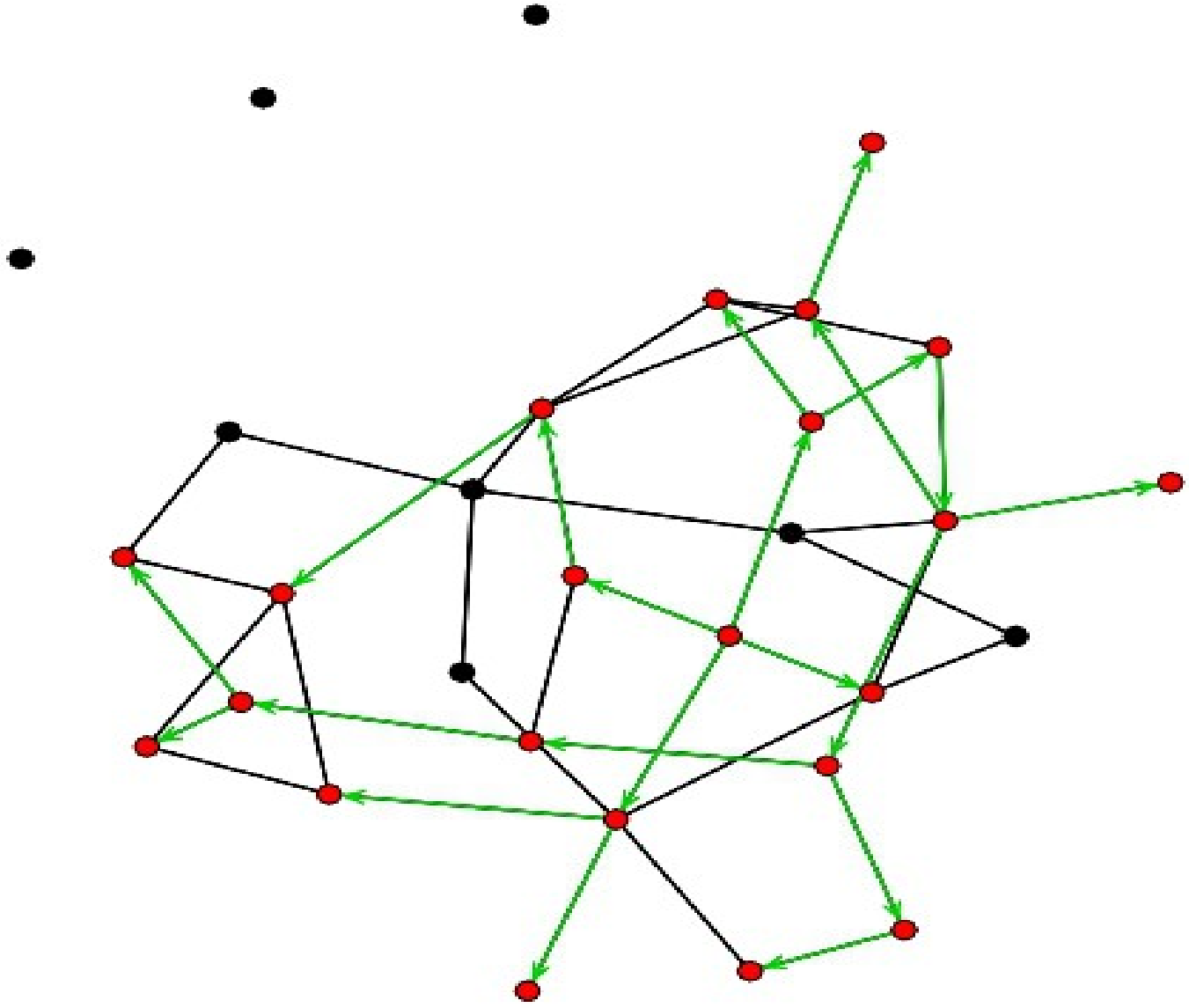
Modelling epidemics: Stochastic compartmental models

- The host population are considered to occupy different classes based on current disease status.
- The SEIR model has four classes:

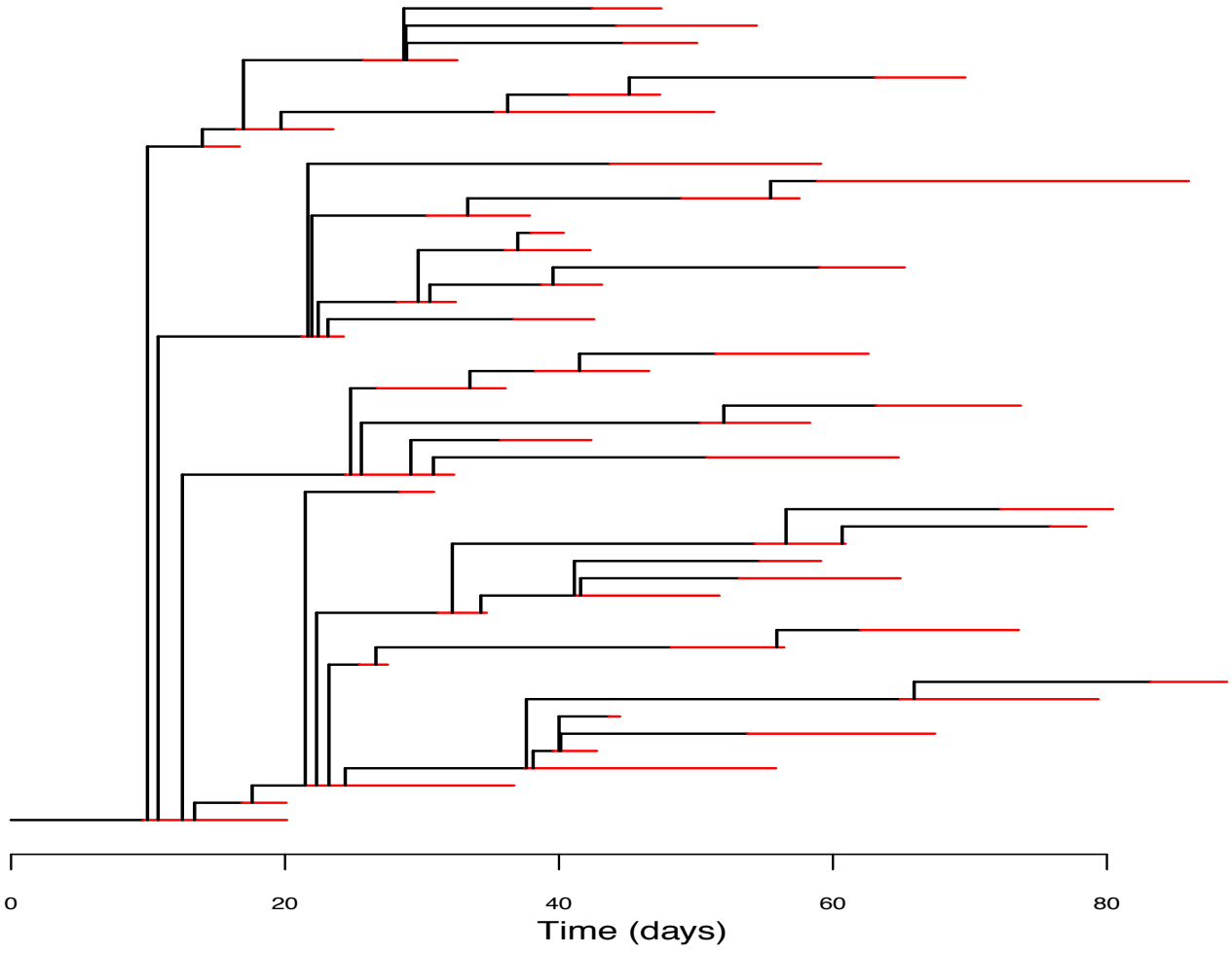
  $$\text{Susceptible} \rightarrow \text{Exposed} \rightarrow \text{Infectious} \rightarrow \text{Removed}$$

- The $S \rightarrow E$ transitions have exponential waiting times, while $E \rightarrow I$ and $I \rightarrow R$ have gamma distributed waiting times.

Modelling epidemics: Stochastic compartmental models

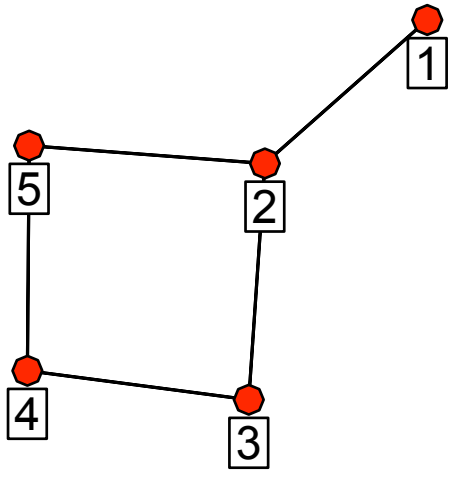


The path the epidemic follows describes a tree. The virus mutates down the tree and we collect genomes from the leaves.

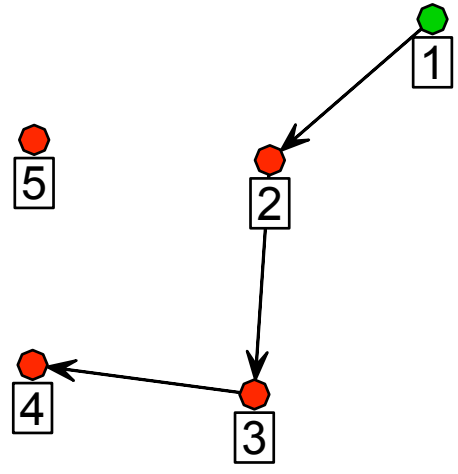


Can also use event time data where known

Node	Exposure Time	Infective Time	Removal Time
1	0.0	6.4	15.1
2	8.1	12.3	16.7
3	13.5	22.9	41.2
4	38.6	48.0	56.9



Contact Network



Transmission Tree

Parameters of interest

- Infection rate β
- Parameters θ and k of latent and recovery times
- η are the parameters of the network model
- Use Bayesian techniques to estimate posterior distributions for these parameters.
- To calculate likelihood, need also to estimate $\mathcal{P}$ and $\mathcal{G}$

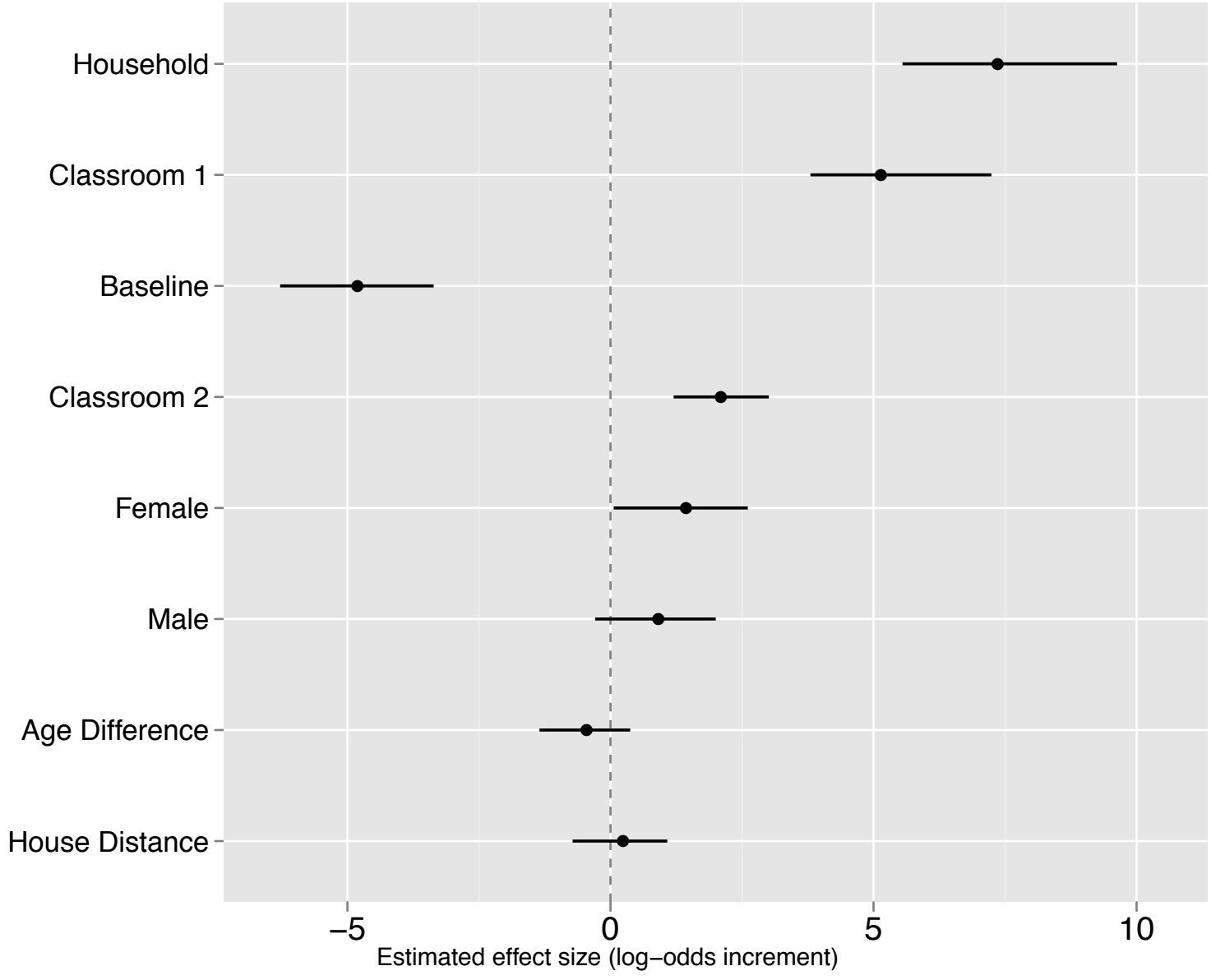
Hagelloch Measles Data

- We consider (unusually complete) data from a measles epidemic that spread through the small town of Hagelloch, Germany, in 1861.
- The data contain proxies for the Infective and Recovery times.
- We have to infer the Exposure times.

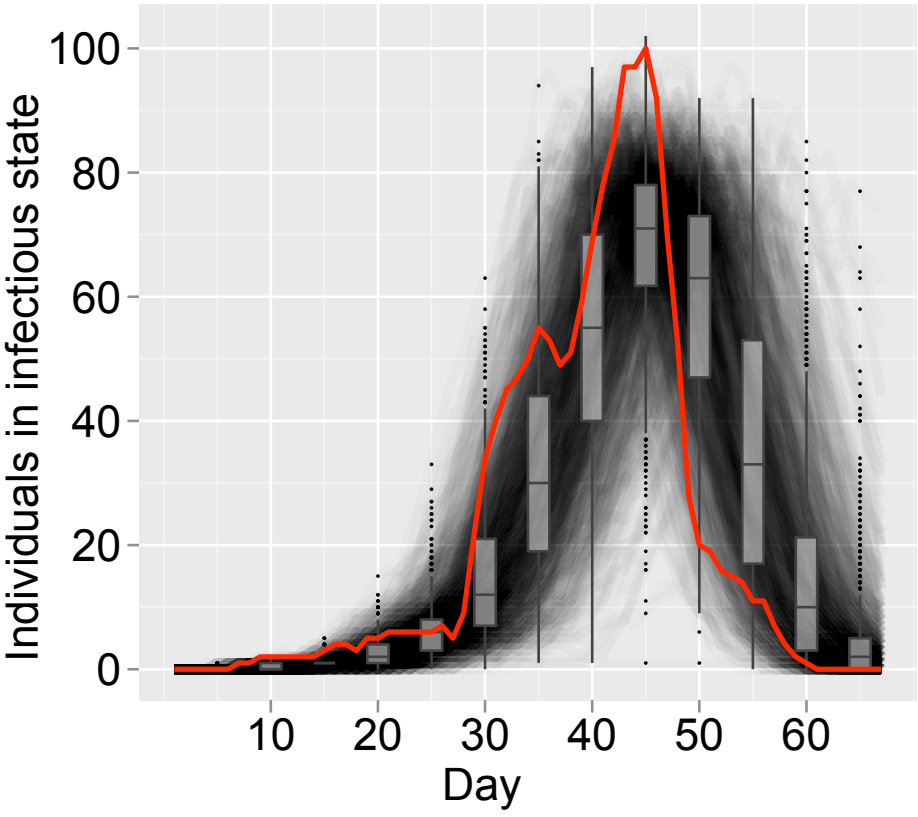
Node i	E_i	I_i	R_i
1	?	6.4	15.1
2	?	12.3	16.7
3	?	22.9	41.2
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Assume 188 individuals infected comprise the complete susceptible population so $N = 188$.

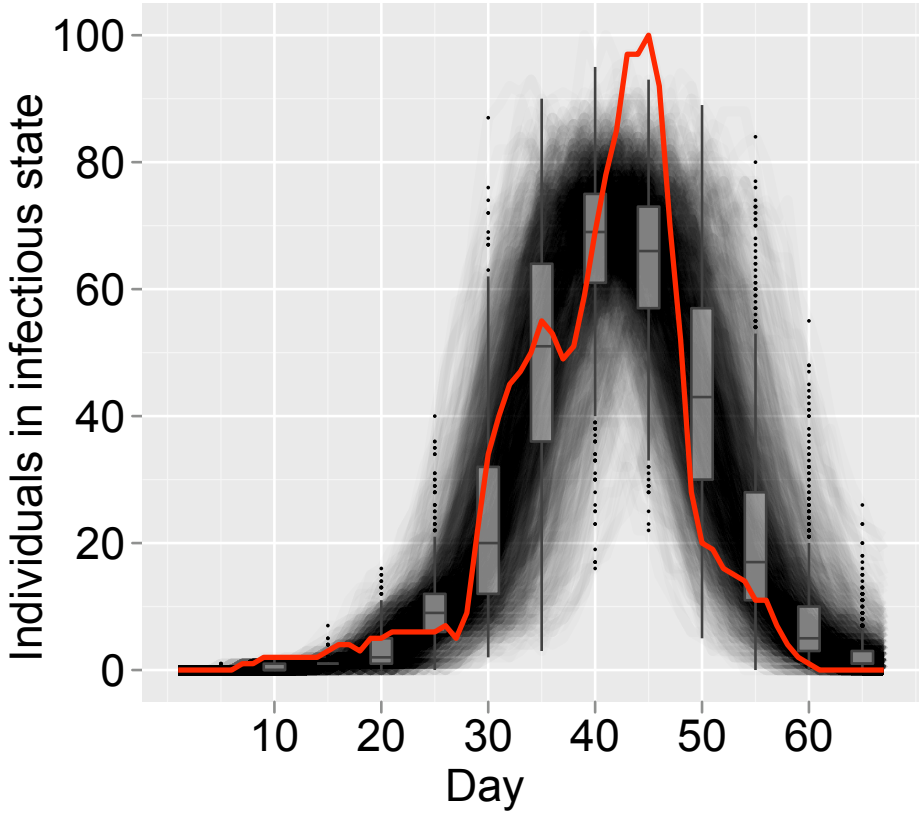
Posterior Estimates for η



Model checking — simulations (black) vs real (red)



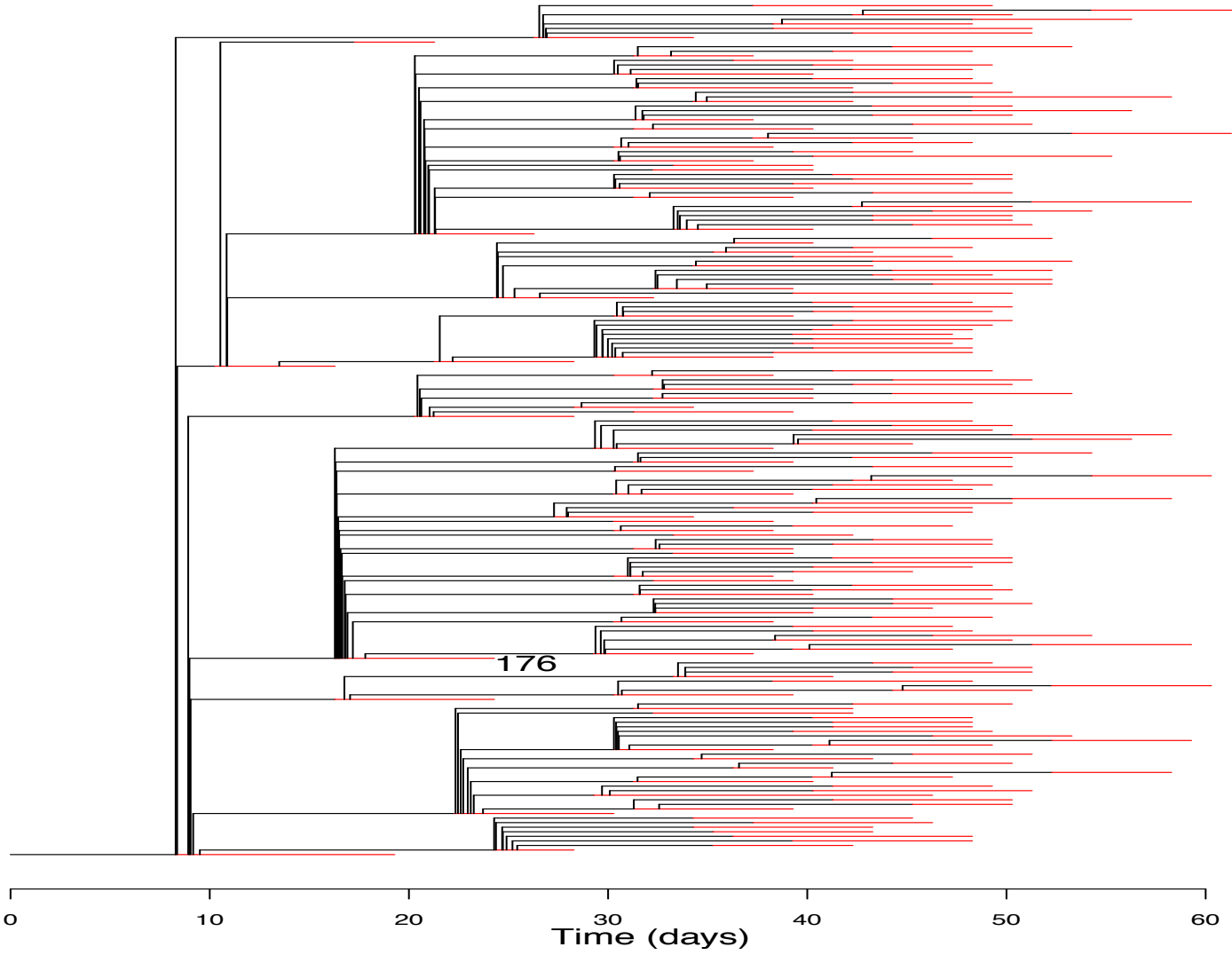
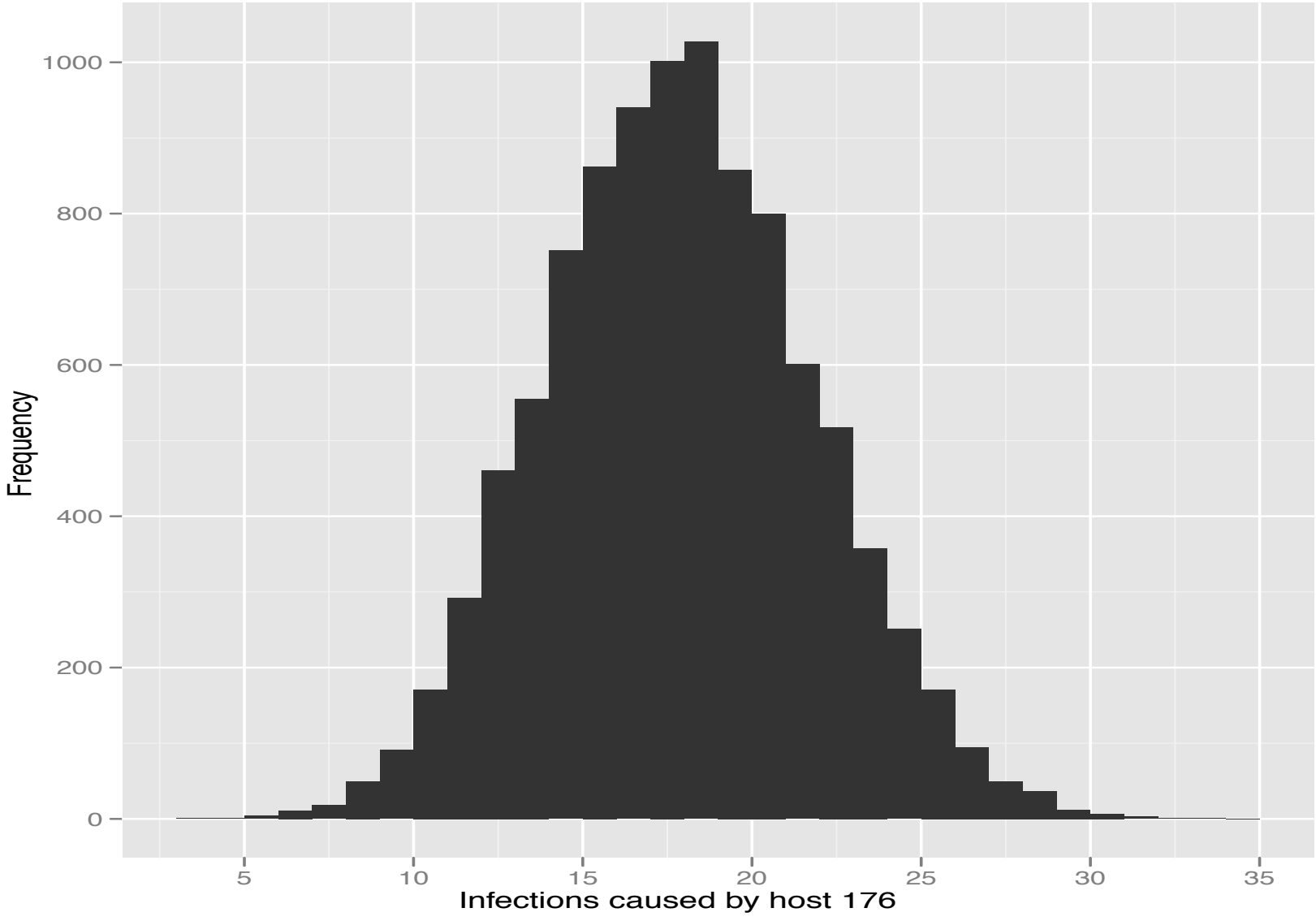
One edge probability



Multiple edge probabilities

Time data tell us little about who infected whom (tree topology)

An typical posterior sample of $\mathcal{P}$



Time data tell us little about who infected whom (tree topology)