
A case study in object-oriented programming

Dirk Husmeier

Biomathematics and Statistics Scotland

Edinburgh, United Kingdom

Email: dirk@bioss.ac.uk

<http://www.bioss.ac.uk/~dirk>

SERAD

SERAD

Searching for Evidence of Recombination in Alignments of DNA

SERAD

Searching for Evidence of Recombination in Alignments of DNA

Husmeier, Wright (2001)

Journal of Computational Biology 8, 401-427.

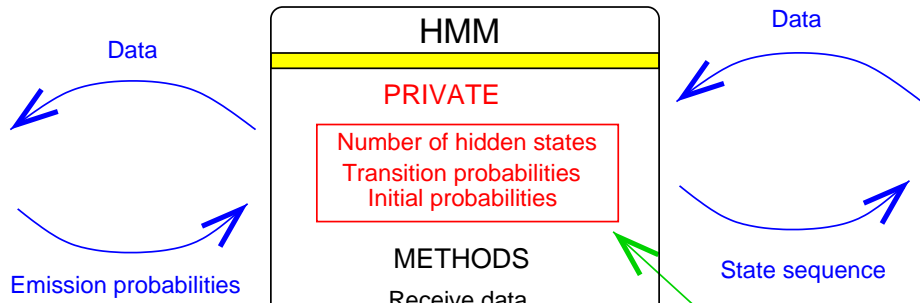
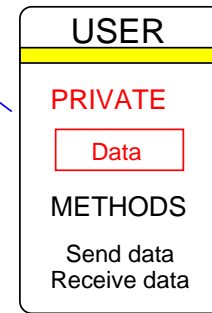
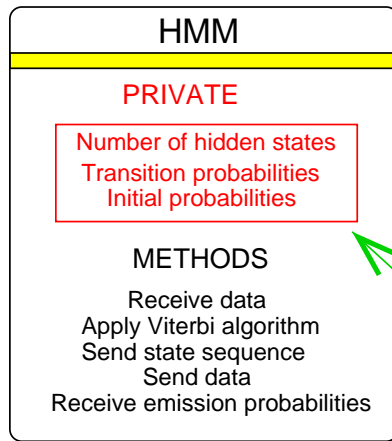
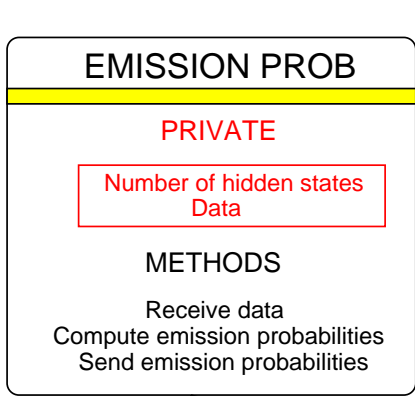
SERAD

Searching for Evidence of Recombination in Alignments of DNA

Husmeier, Wright (2001)

Journal of Computational Biology 8, 401-427.

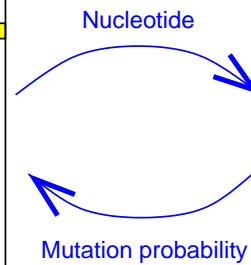
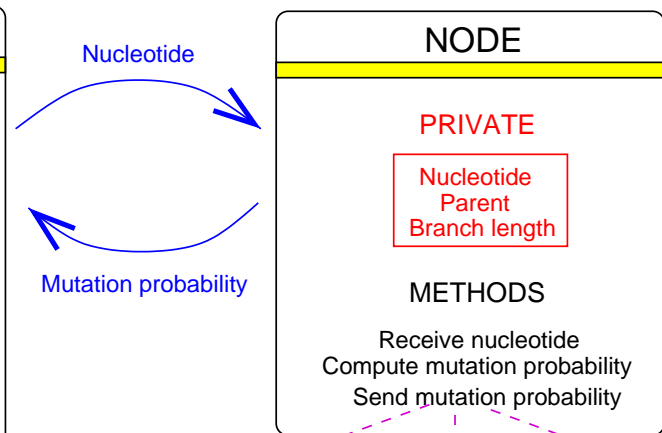
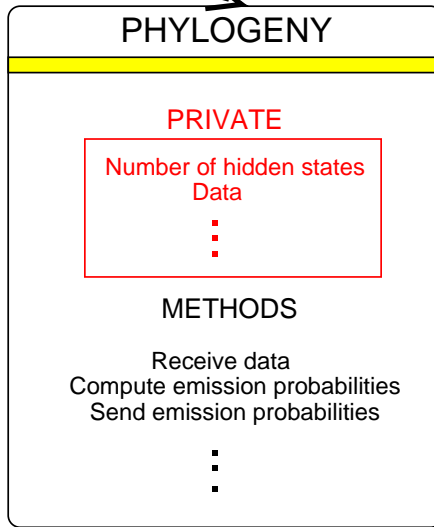
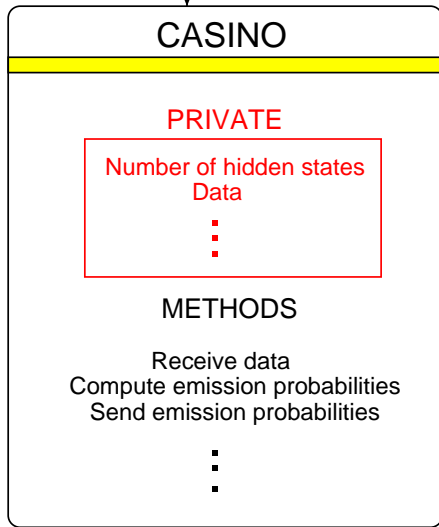
- HMMs
- Phylogenetic trees
- Recombination
- Object-oriented programming implementation



Encapsulation

Inheritance

Polymorphism



Jukes-Cantor
No extra parameter

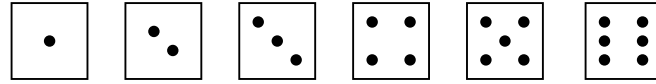
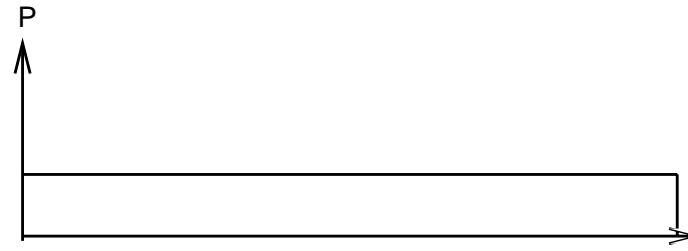
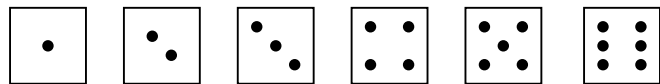
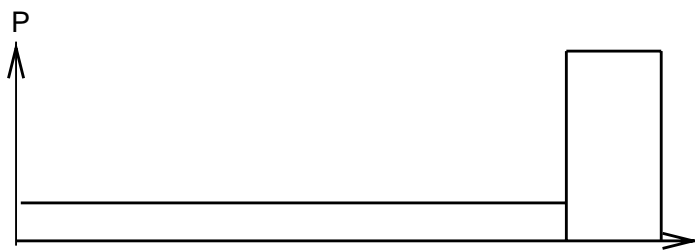
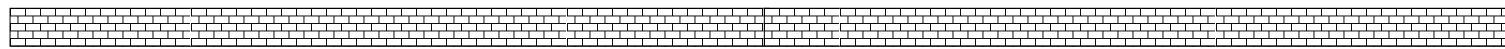
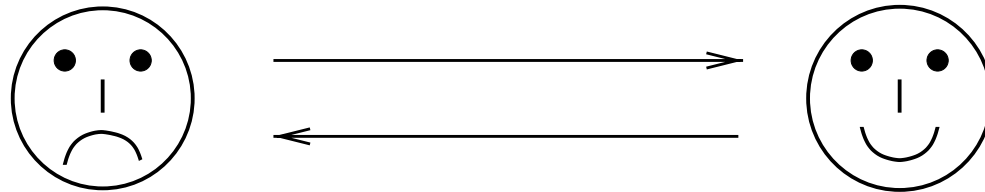
Kimura
1 extra parameter

HKY85
4 extra parameters

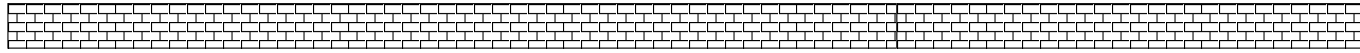
Overloading

Hidden Markov Models

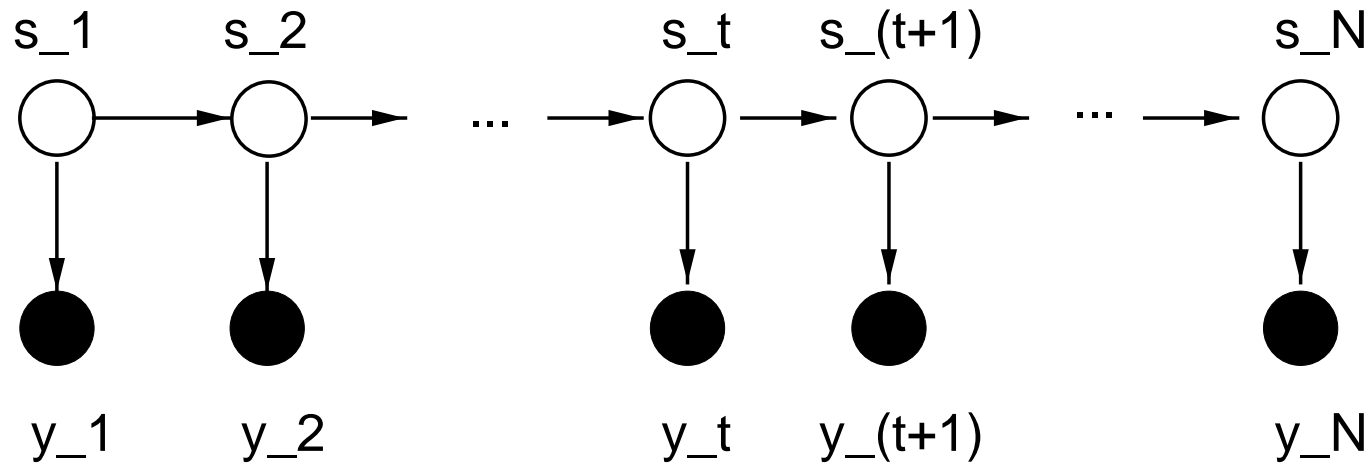
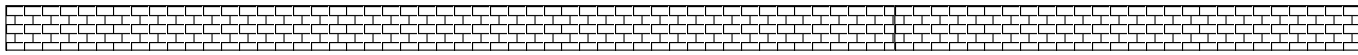
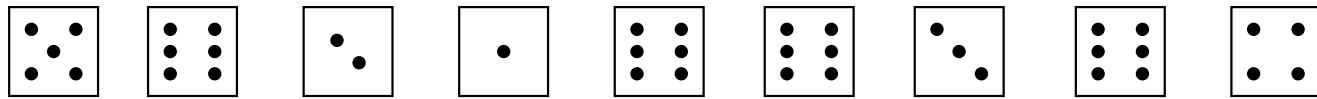
Example: The occasionally corrupt casino



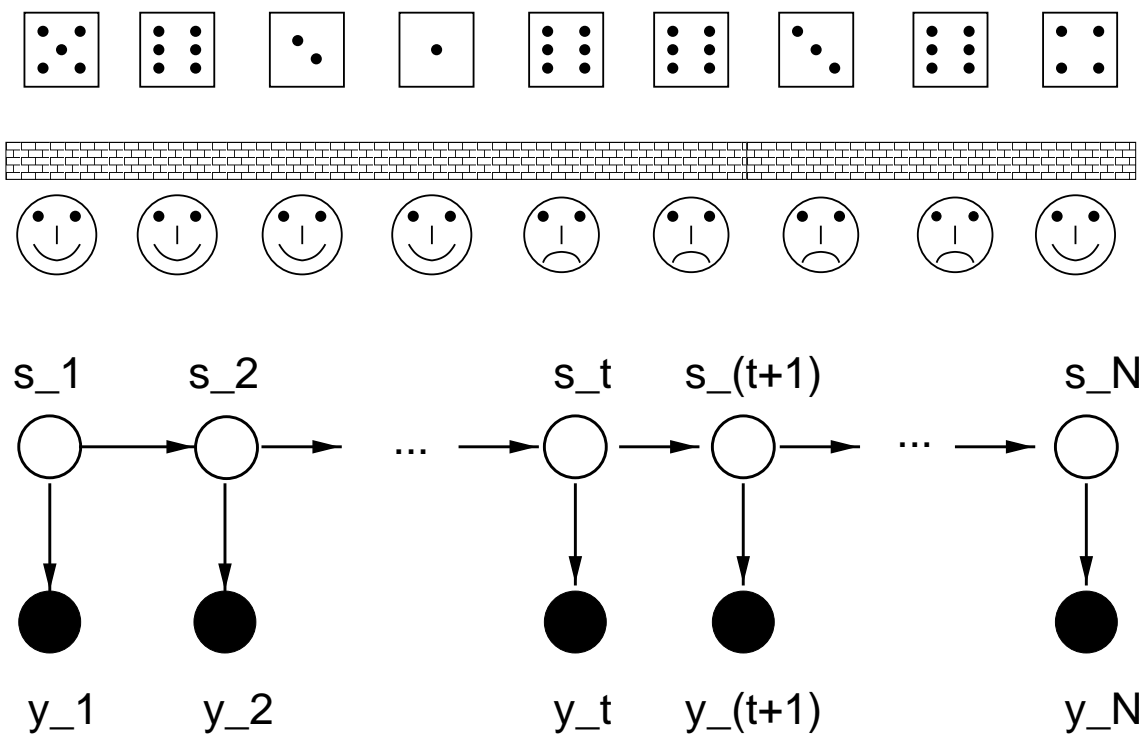
Example: HMM



Example: HMM



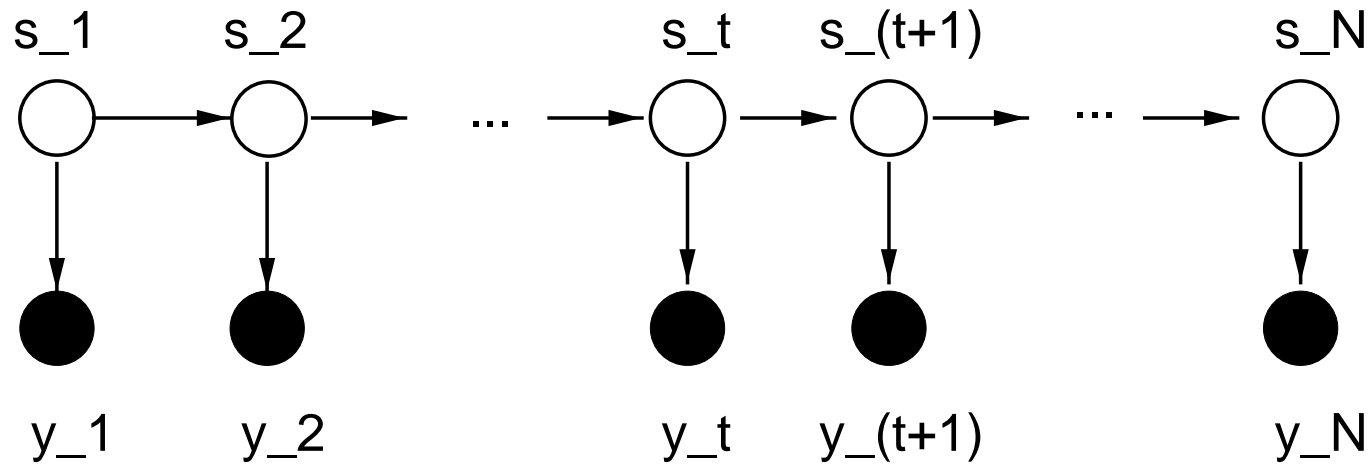
The most likely state sequence



Find the mode of $P(S_1, \dots, S_N | y_1, \dots, y_N)$

Problem: $(S_1, \dots, S_N) : 2^N$ different sequences.

Factorisation in HMMs

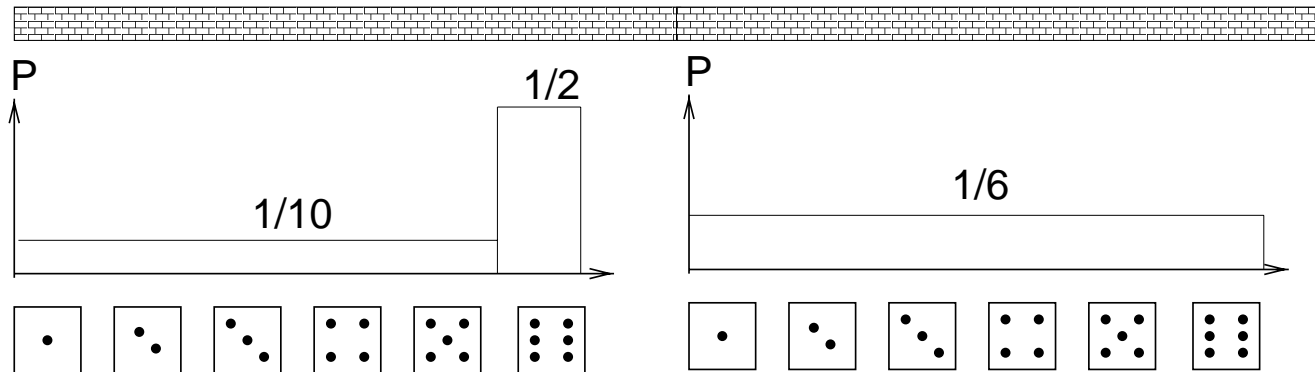
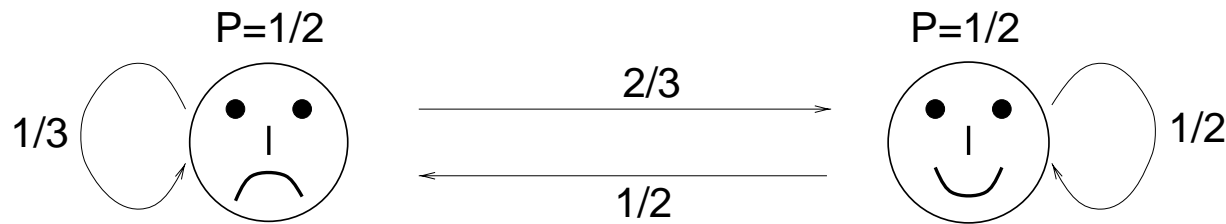


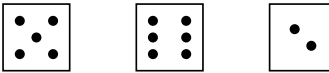
$$P(y_1, \dots, y_N, S_1, \dots, S_N) = \prod_{t=1}^N P(y_t | S_t) \prod_{t=2}^N P(S_t | S_{t-1}) P(S_1)$$

Viterbi algorithm $\longrightarrow P(S_1, \dots, S_N | y_1, \dots, y_N)$

Computational complexity and example

Computation complexity: $|\mathcal{H}|^N \longrightarrow N \times |\mathcal{H}|^2$

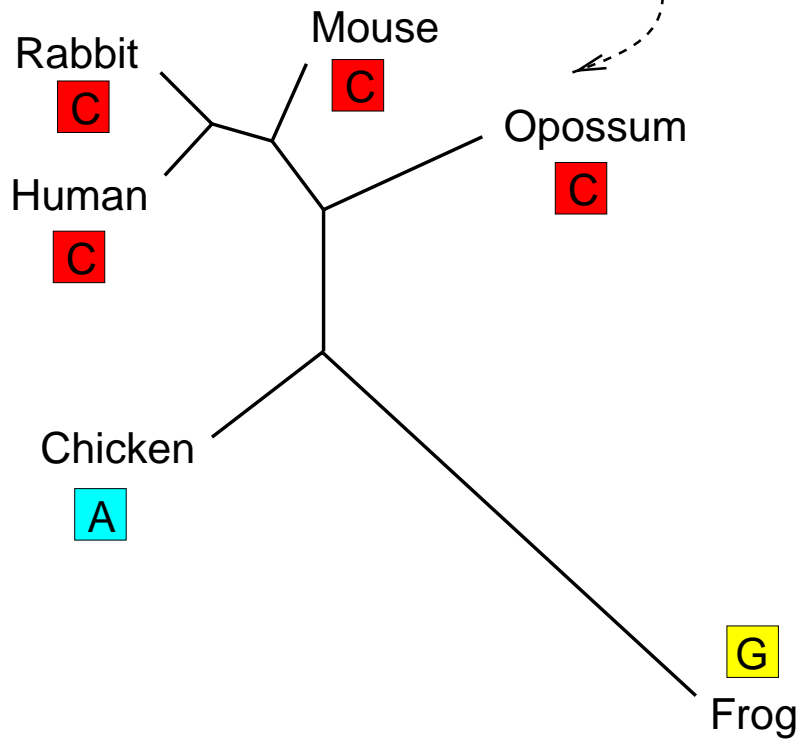


Observed sequence: 

Phylogenetic trees

↓

Frog	G	C	T	T	G	A	C	T	T	C	T	G	A	G	G	T	T
Chicken	G	C	G	T	A	A	C	T	T	C	A	C	A	T	G	A	T
Human	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T
Rabbit	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T
Mouse	G	C	G	T	C	A	C	T	T	G	A	C	A	G	G	C	T
Opossum	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T

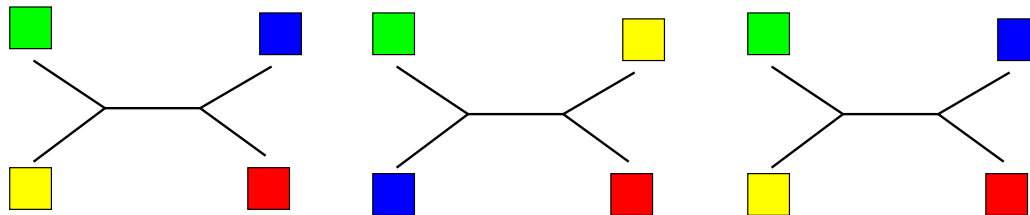
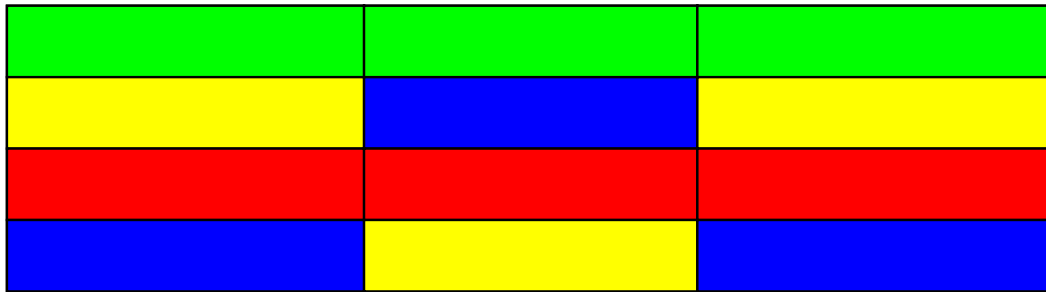
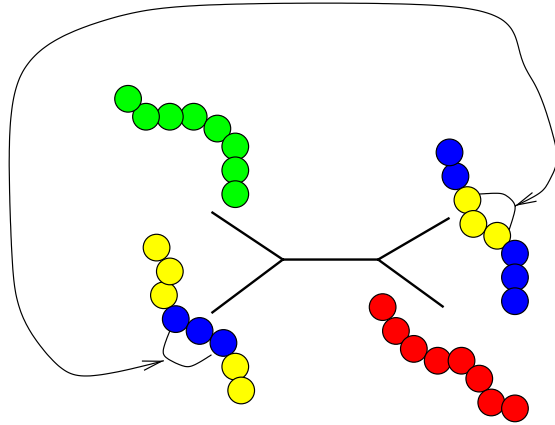


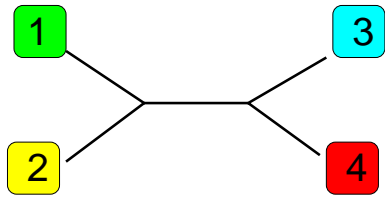
--> Likelihood

Topology
Branch lengths

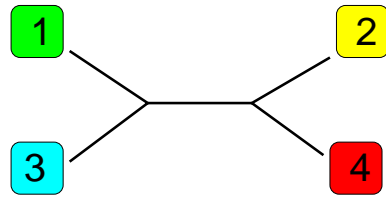
Recombination

Recombination

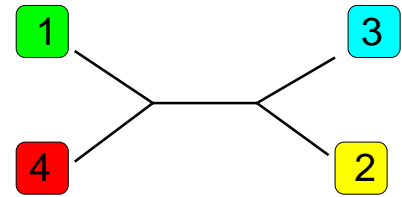




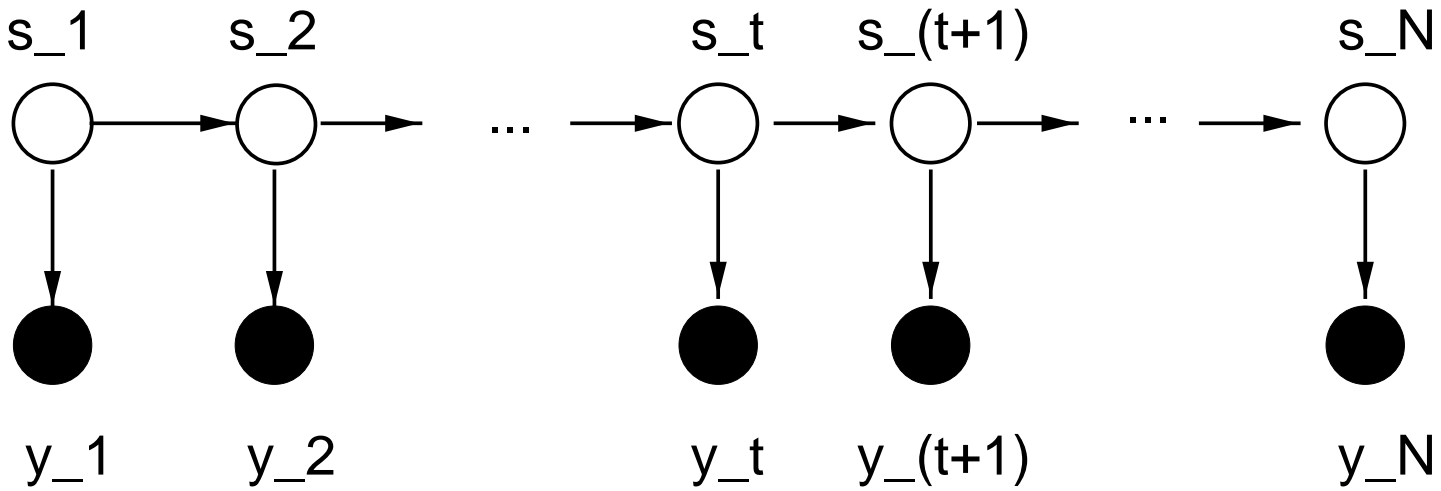
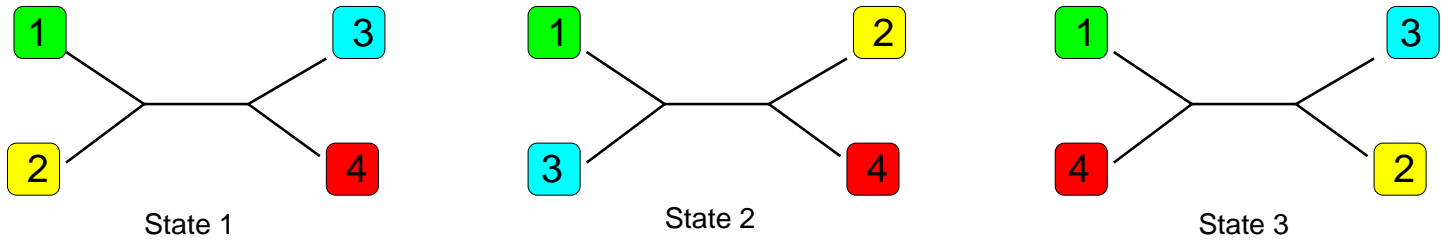
State 1



State 2



State 3



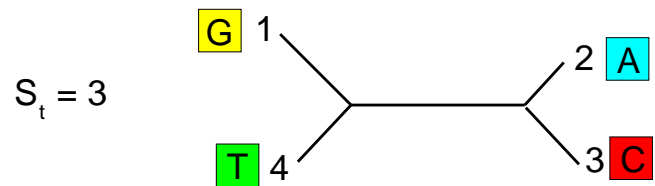
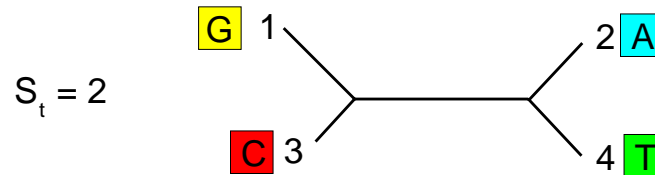
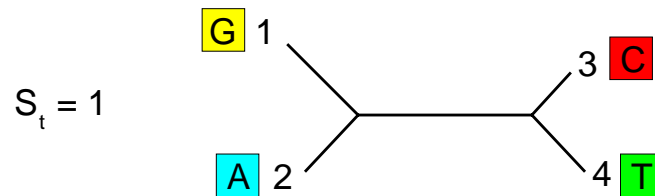
AGCATCGTTCTATTTTACCGGCTCCCG
 TGTGTCGCTCAAGATTGCCATCGCGCG
 TGTGTCGTTGGTCTAGATTGCCATCGCGCG
 TGTATCGCTCTAGTTTGCCAGCTCCCG

Emission probabilities

↓

Strain 1	G	C	T	T	G	A	C	T	T	C	T	G	A	G	G	T	T
Strain 2	G	C	G	T	A	A	C	T	T	C	A	C	A	T	G	A	T
Strain 3	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T
Strain 4	G	C	G	C	T	A	C	T	T	G	A	G	A	C	G	C	T

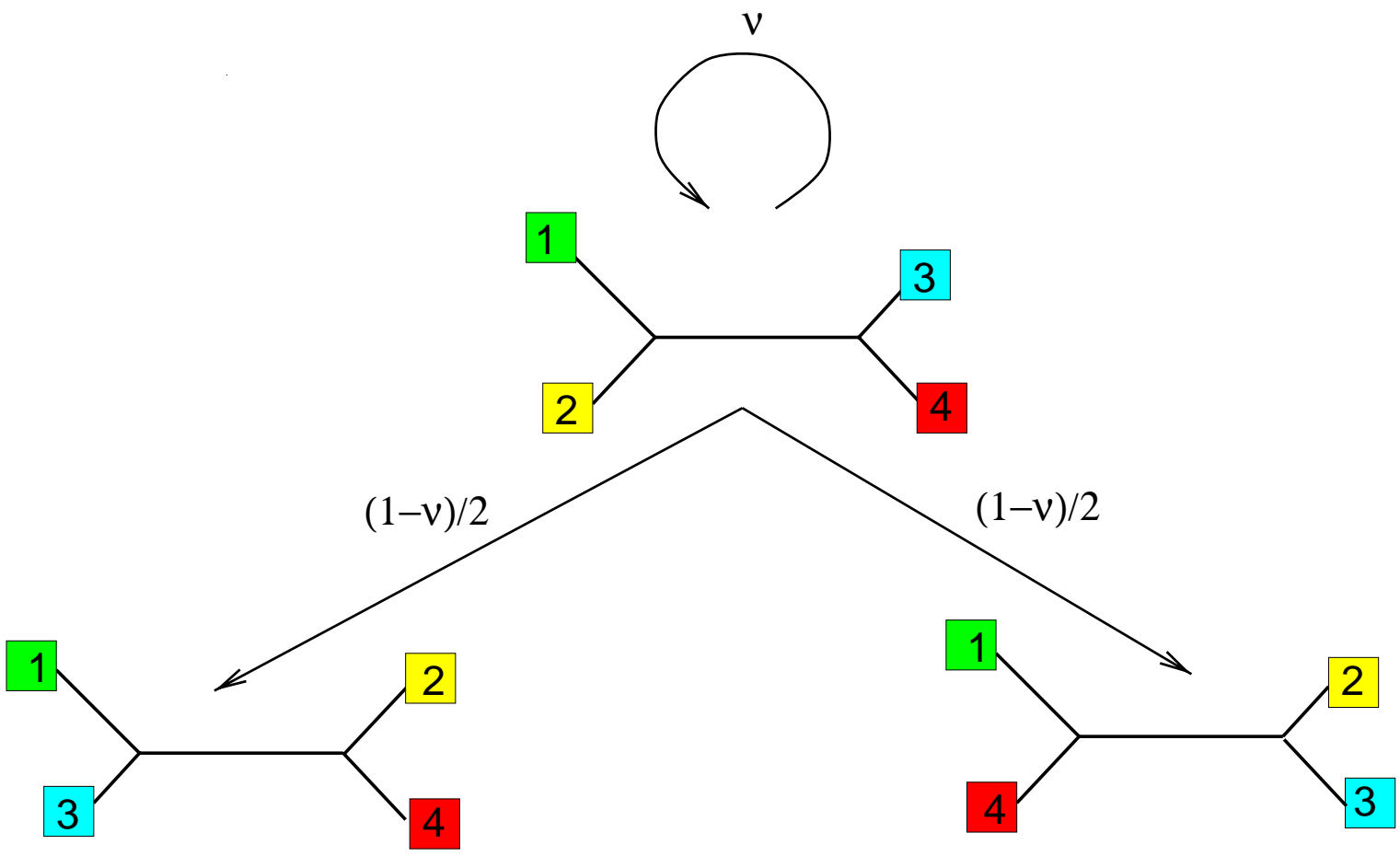
y_t



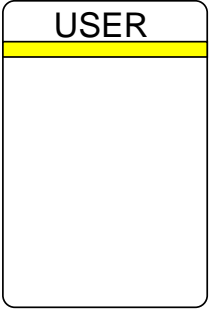
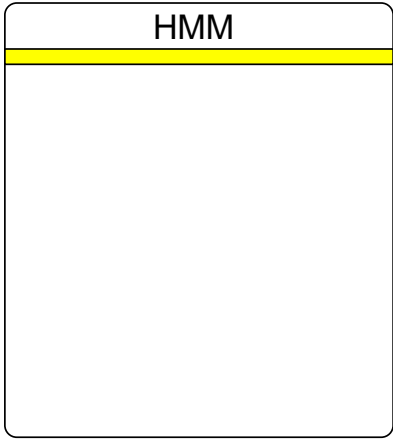
--> $P(y_t | S_t, w)$

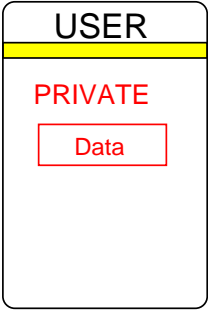
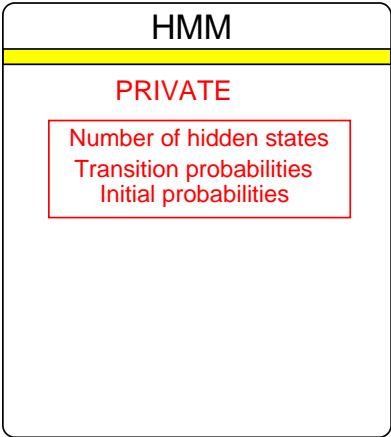
Topology	S_t
Branch lengths	w

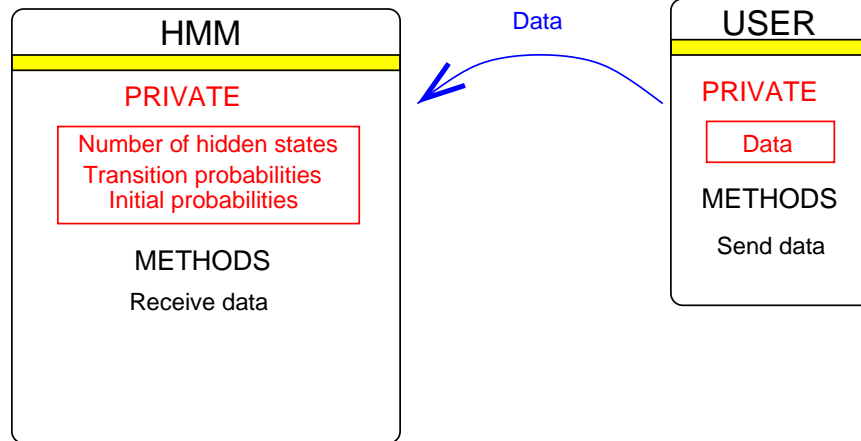
Transition probabilities

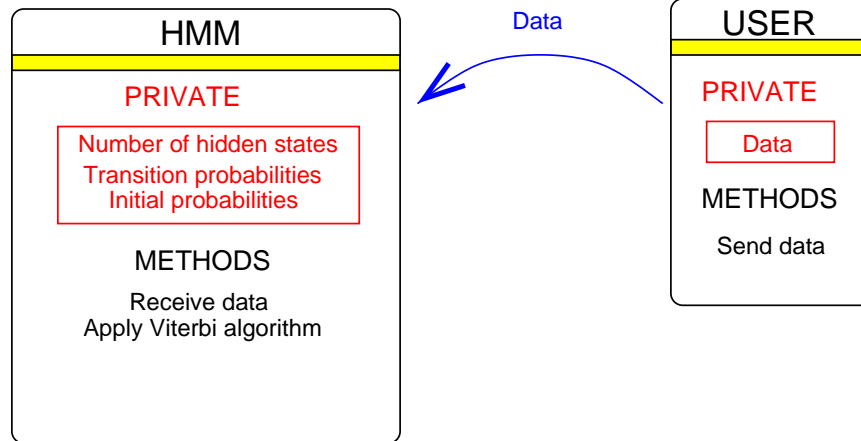


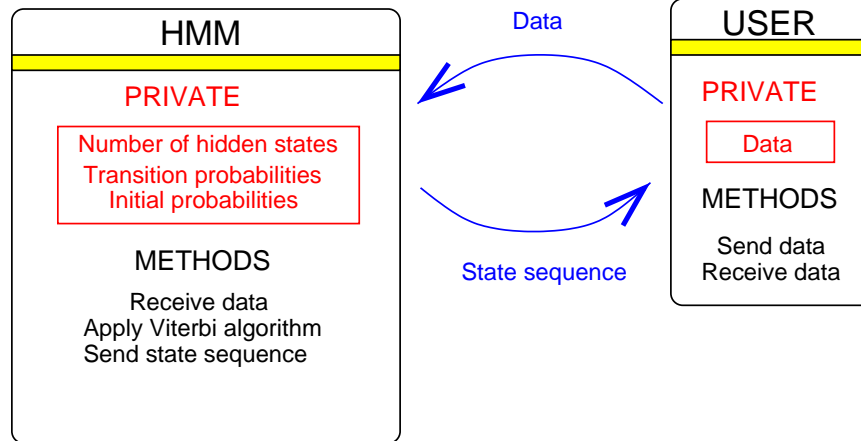
Object-oriented programming

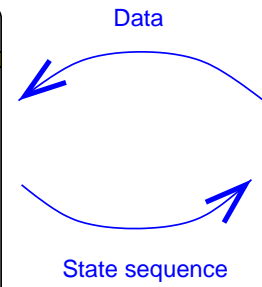
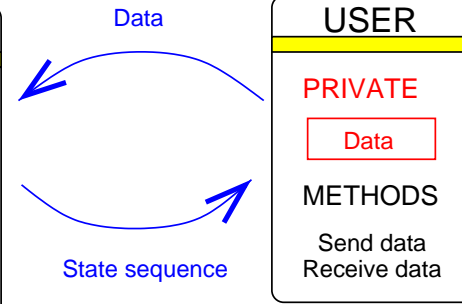
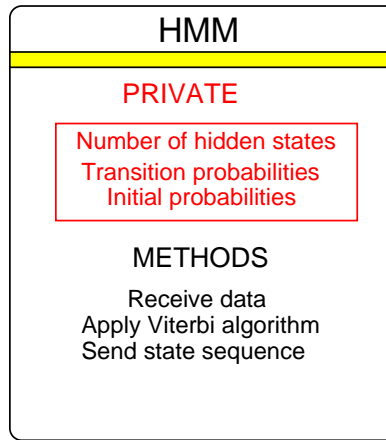
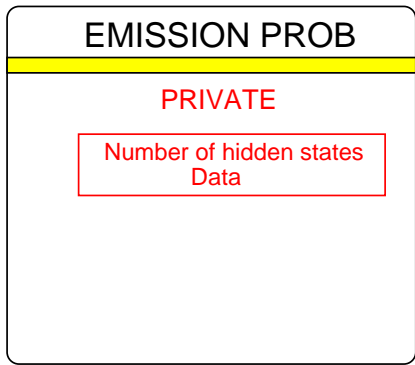


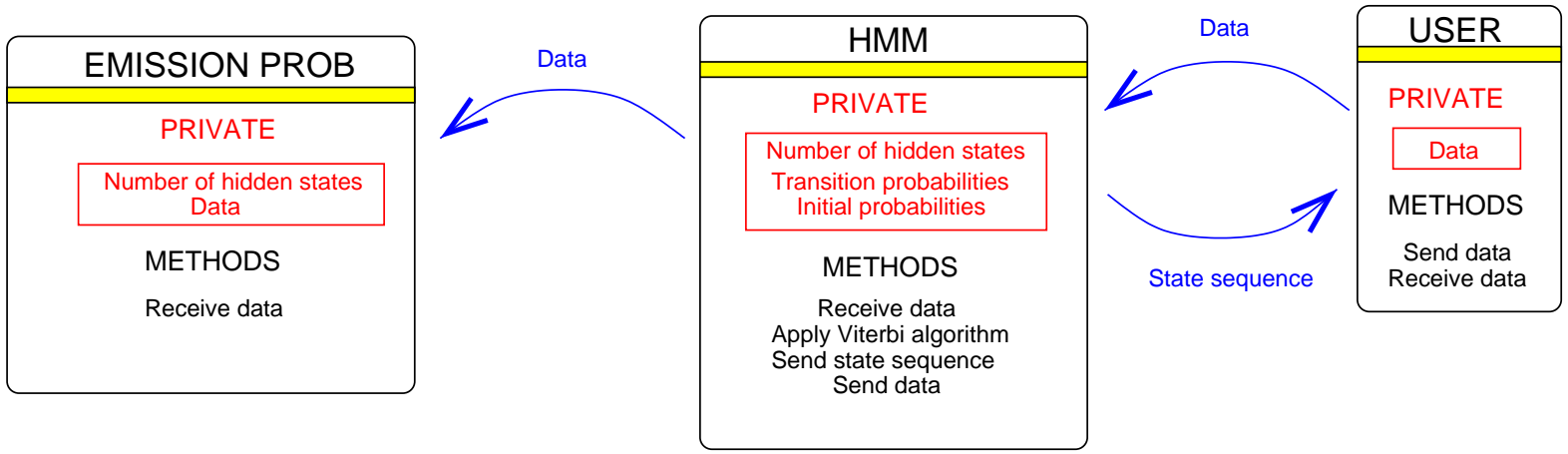


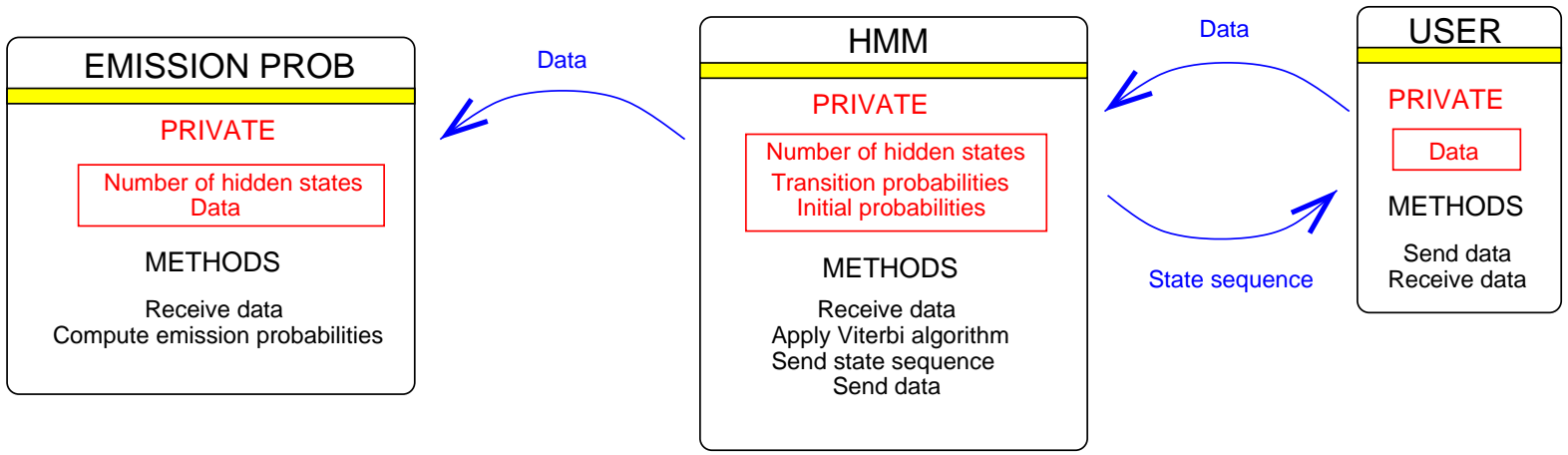


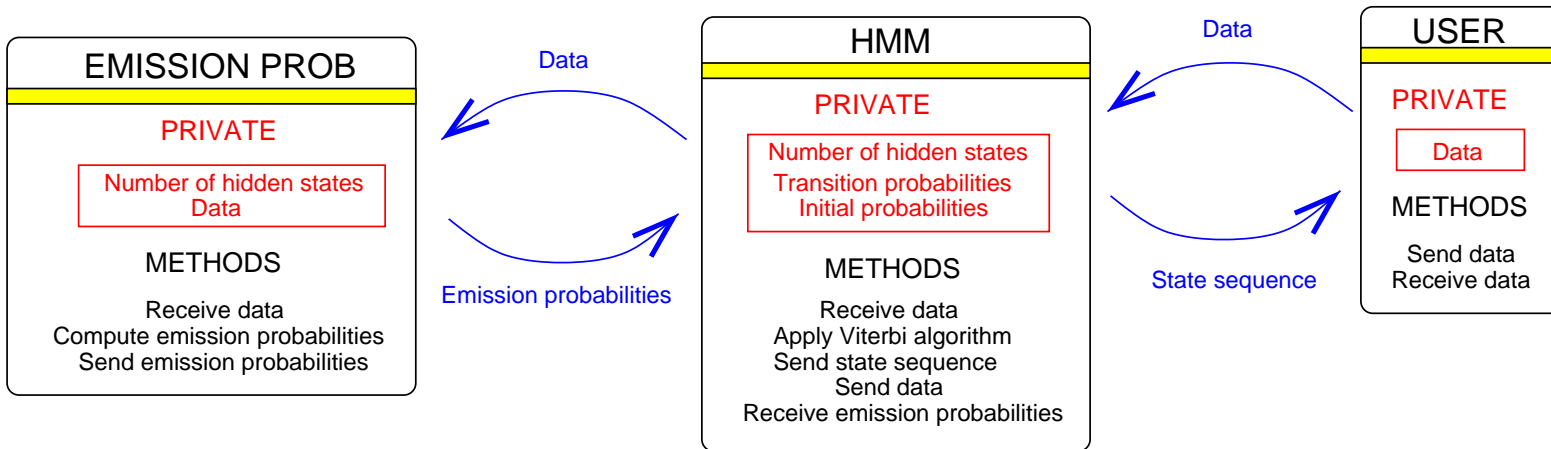


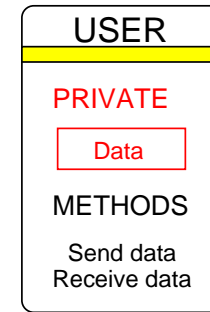
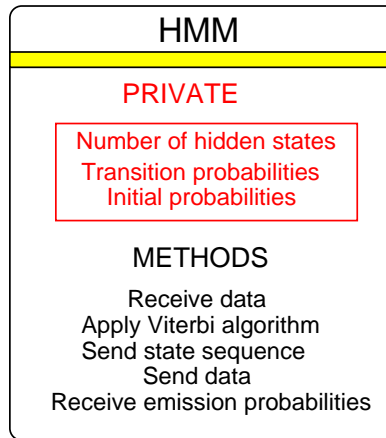
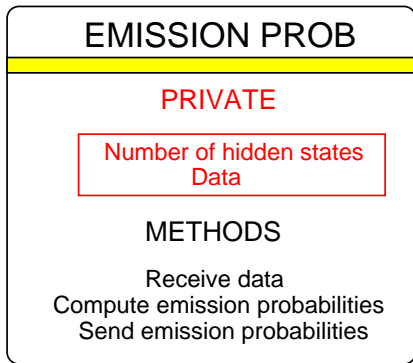


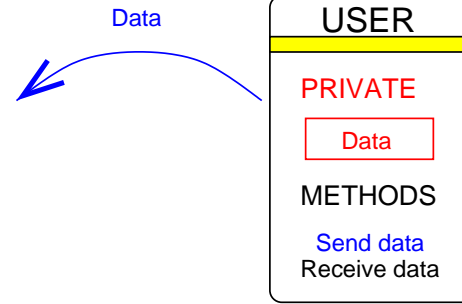
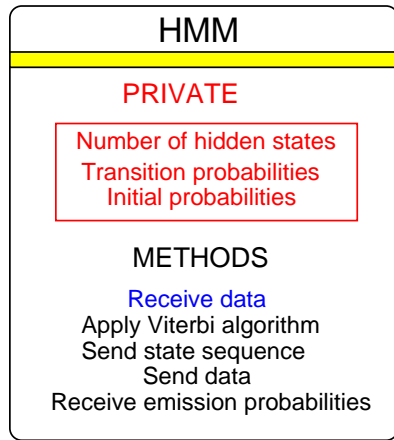
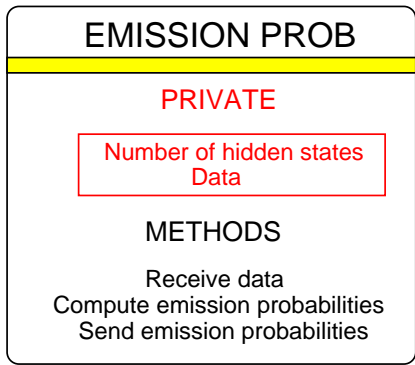


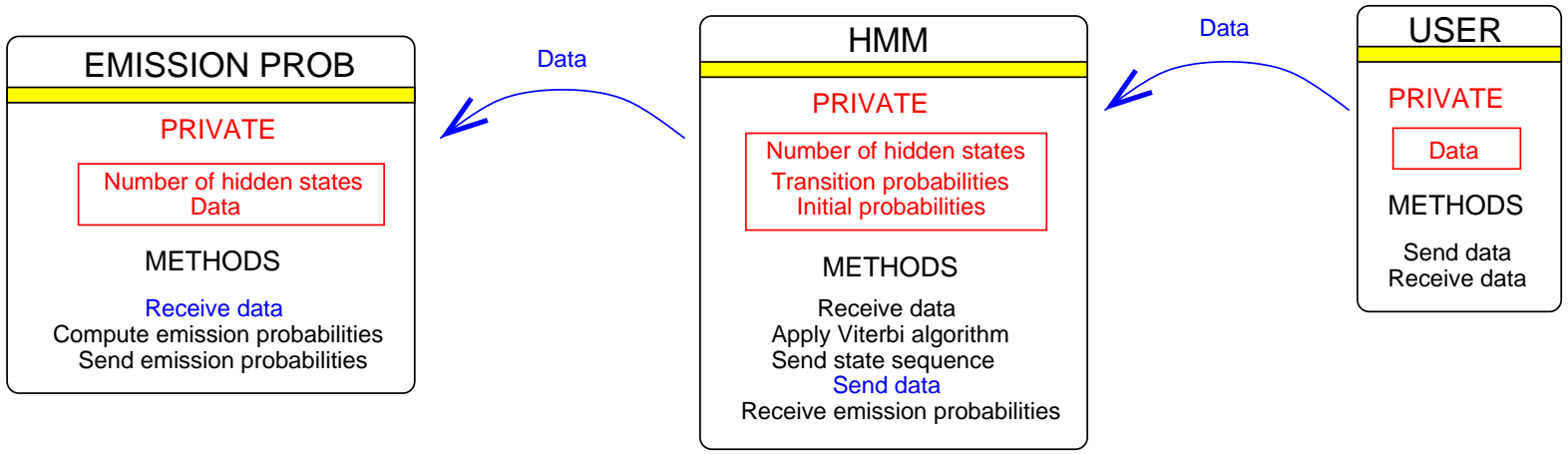


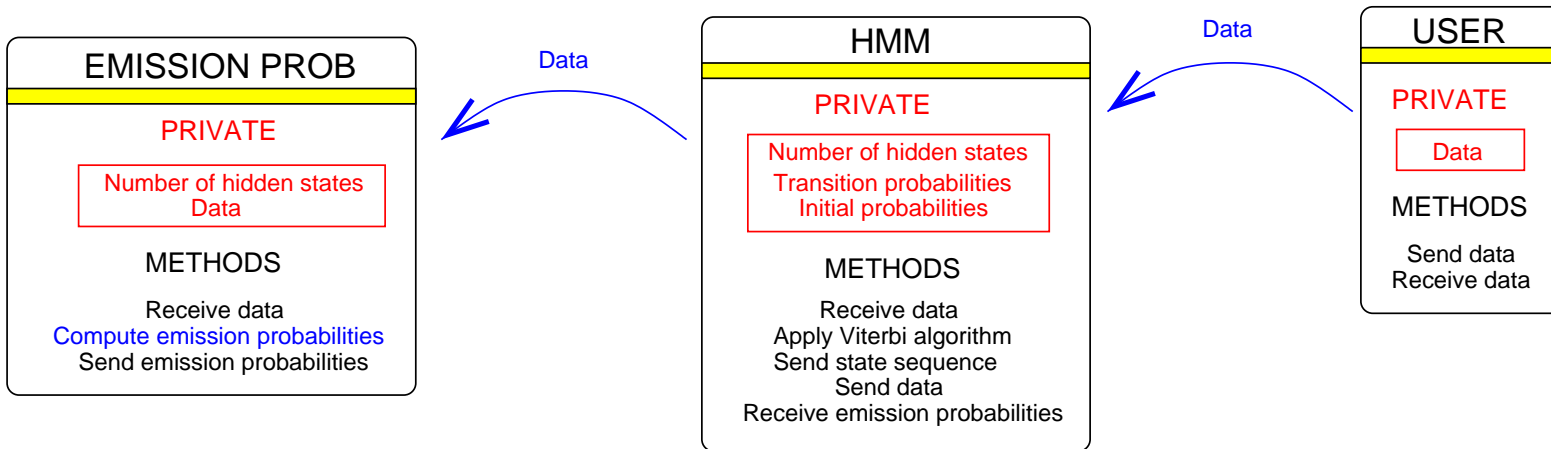


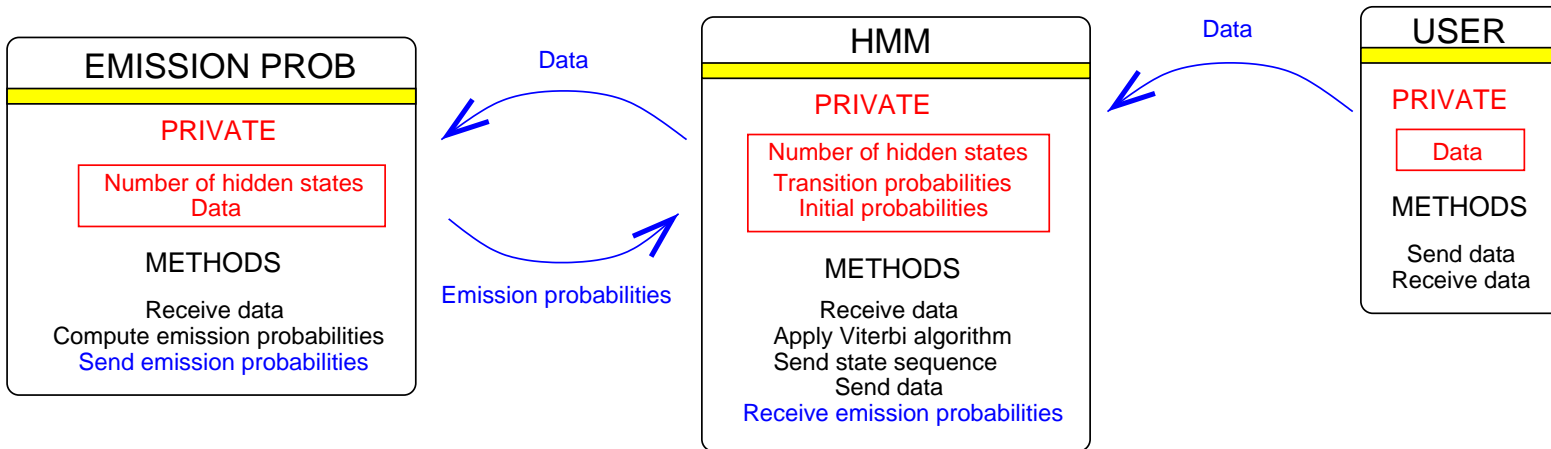


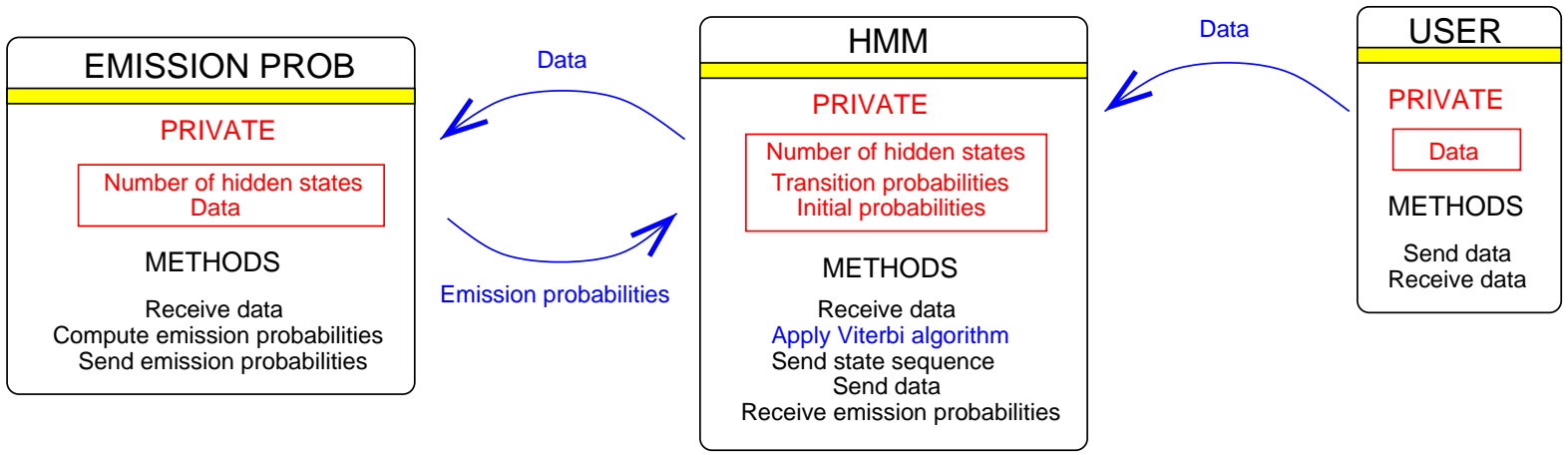


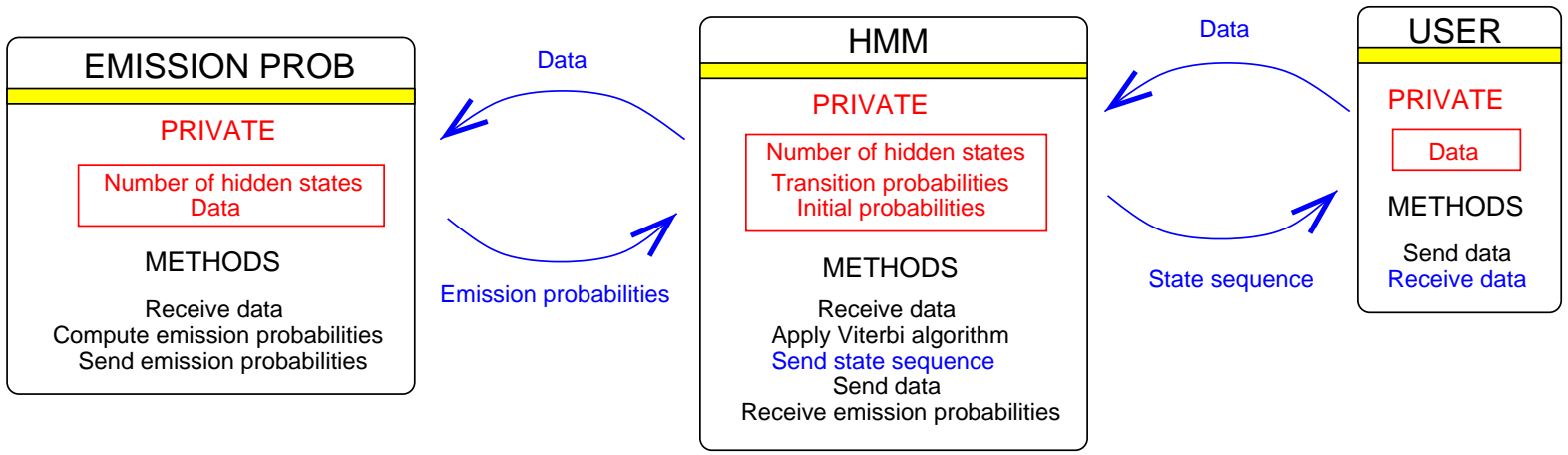


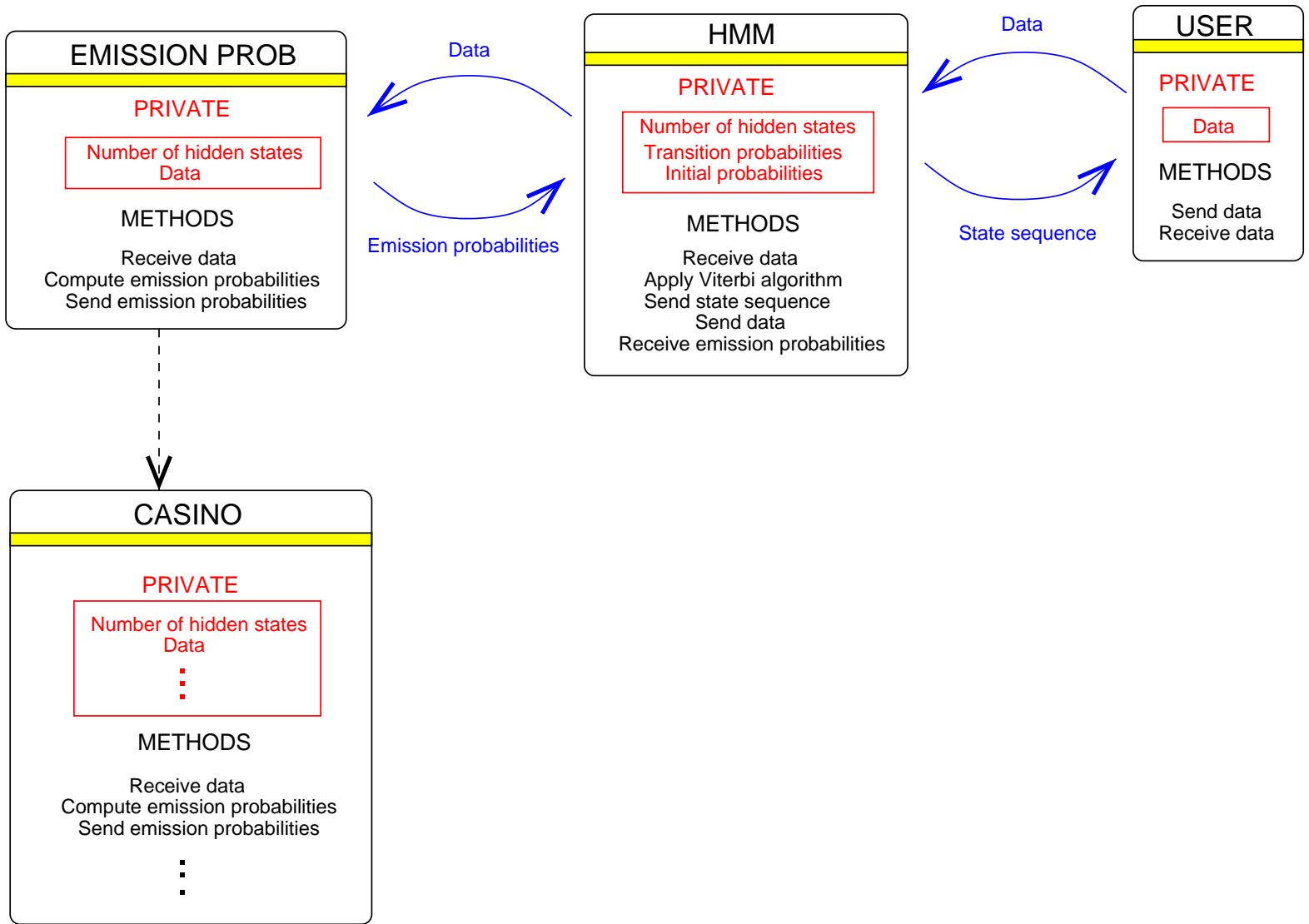


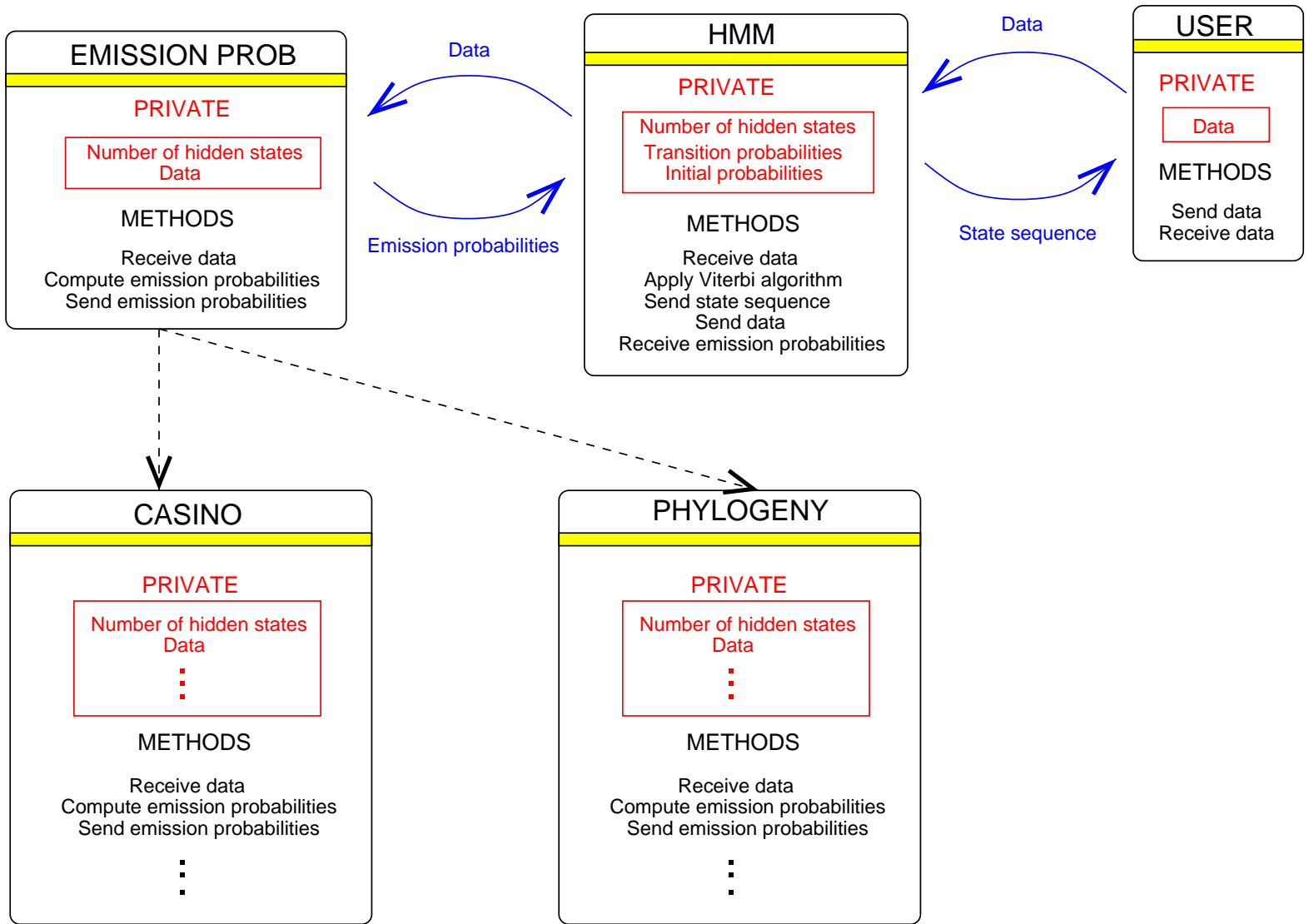


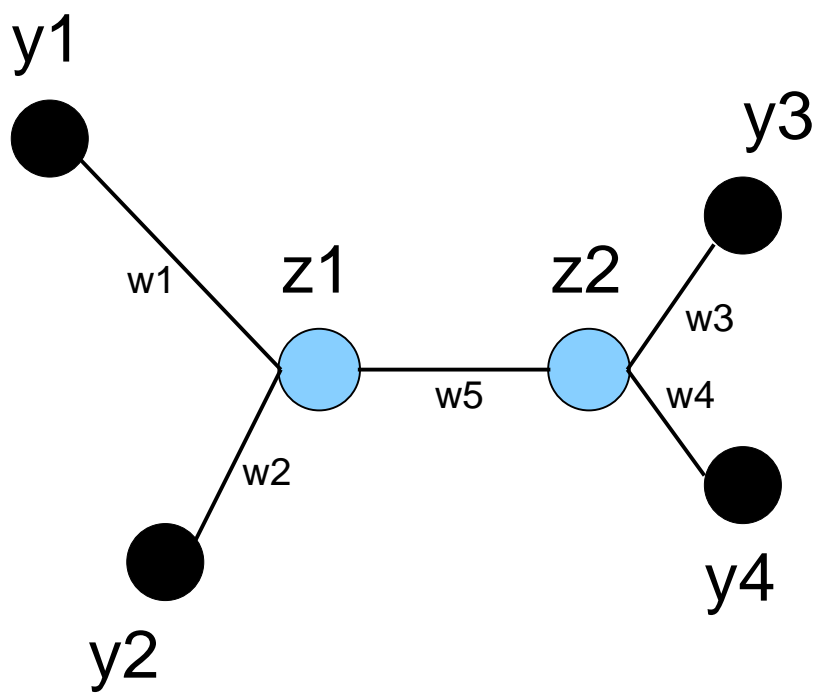




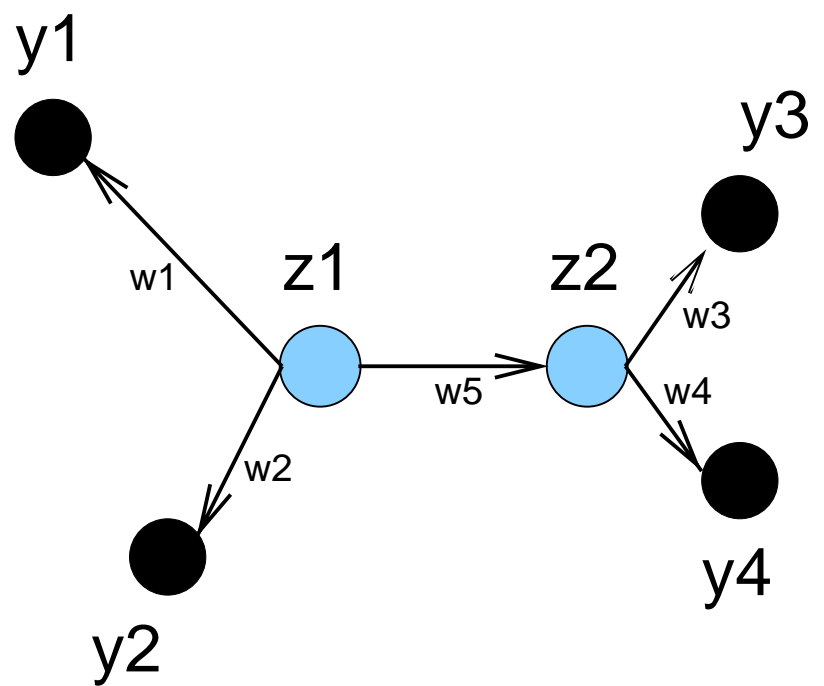




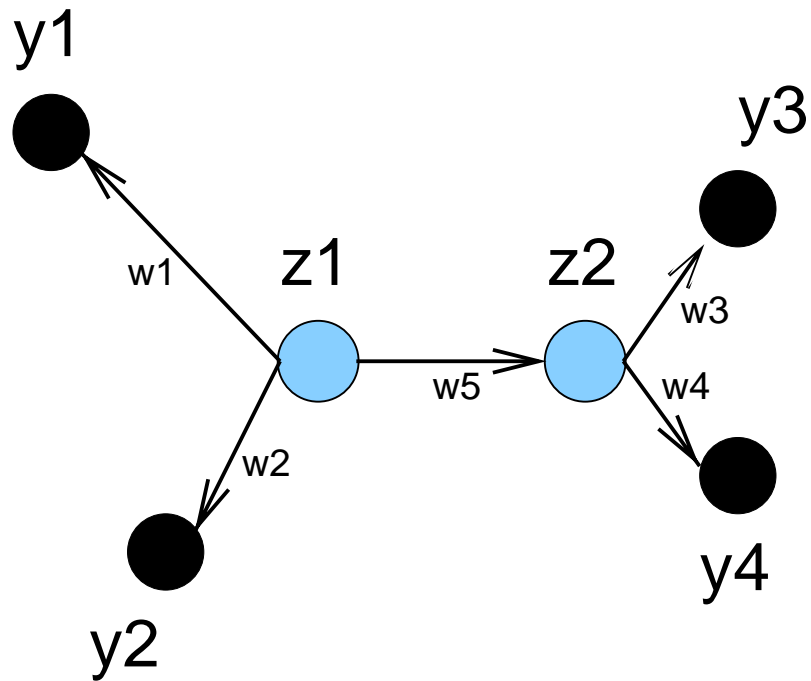




$$P(y_1, y_2, y_3, y_4, z_1, z_2 | \mathbf{w})$$

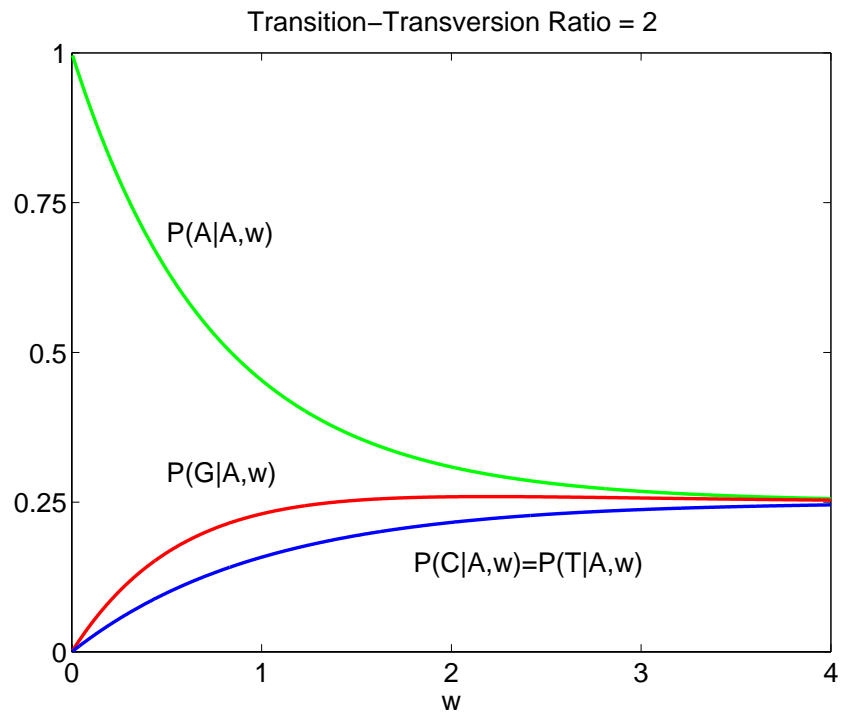
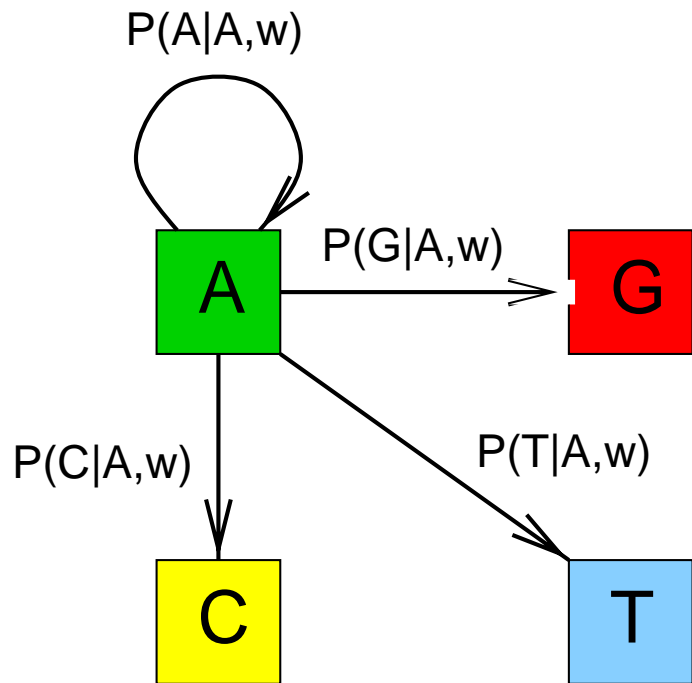


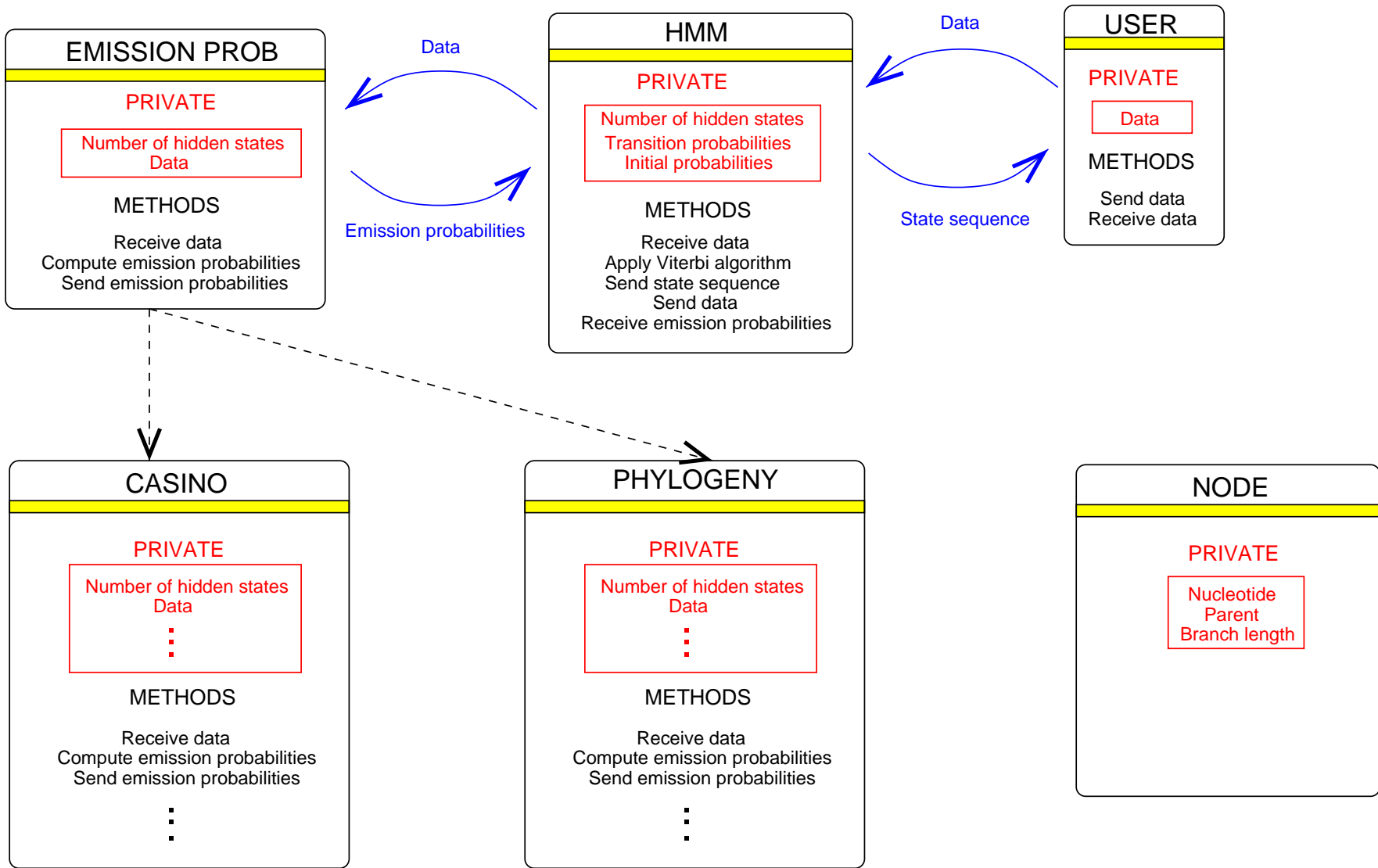
$$P(y_1, y_2, y_3, y_4, z_1, z_2 | \mathbf{w})$$

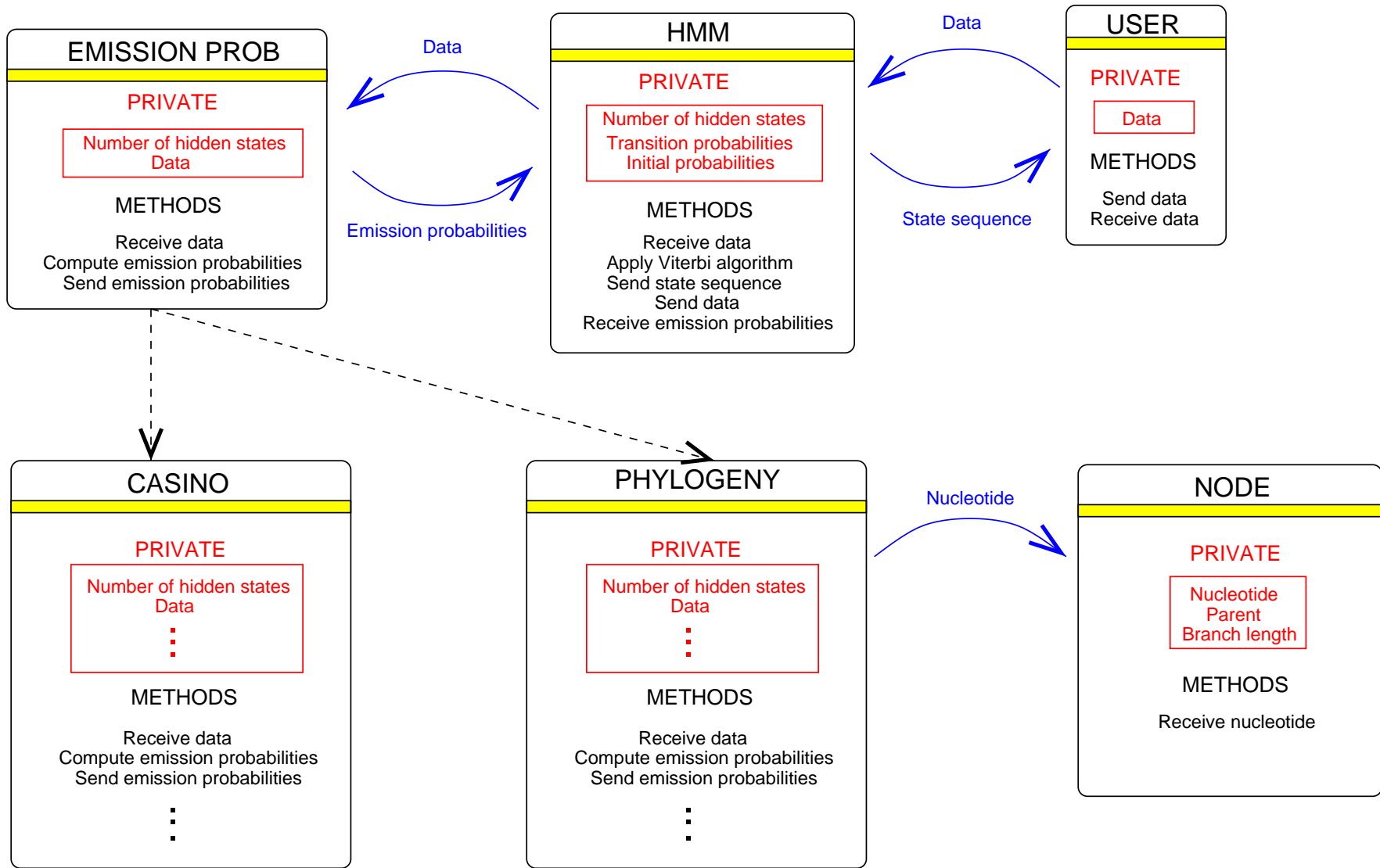


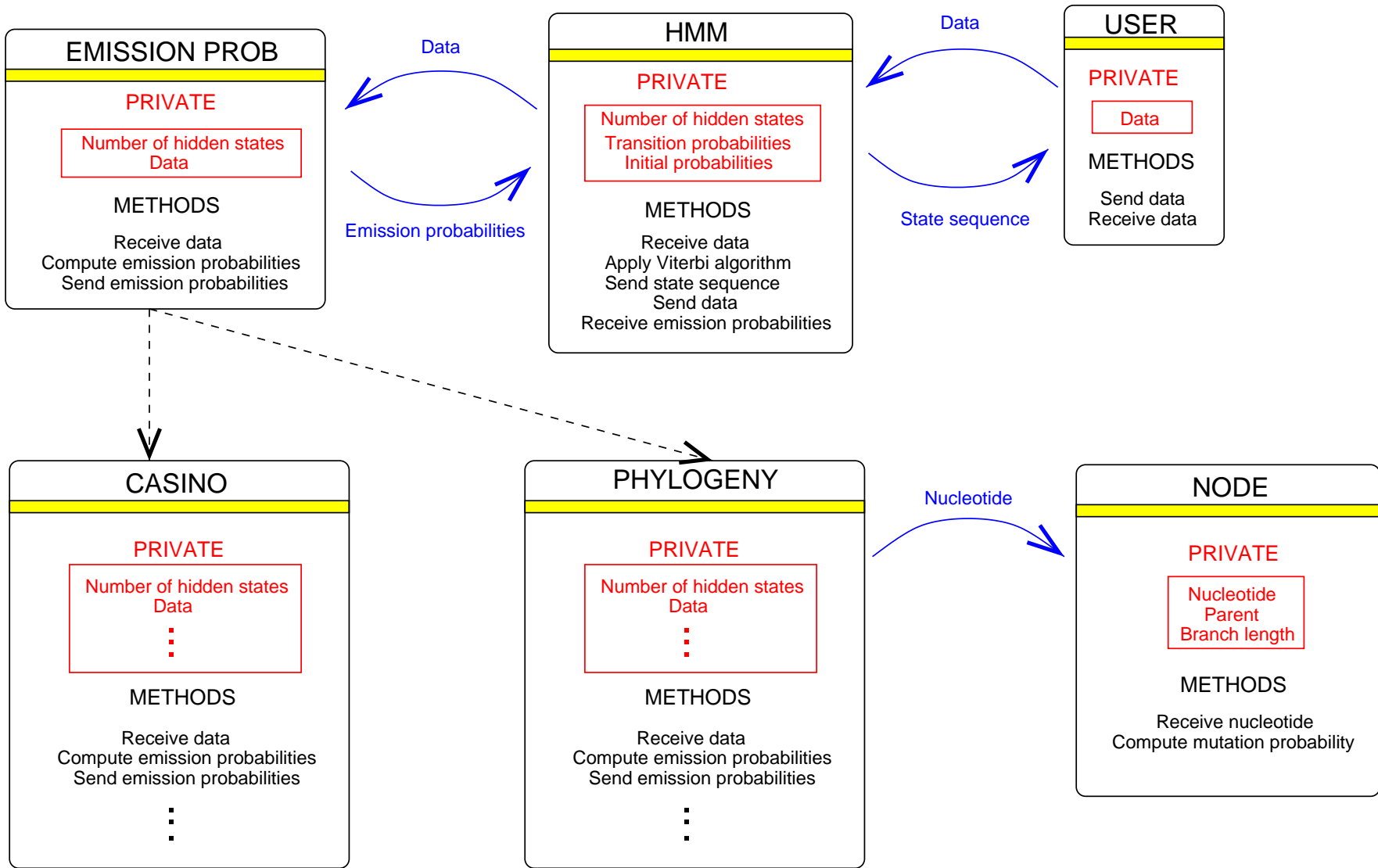
$$P(y_1, y_2, y_3, y_4, z_1, z_2 | \mathbf{w})$$

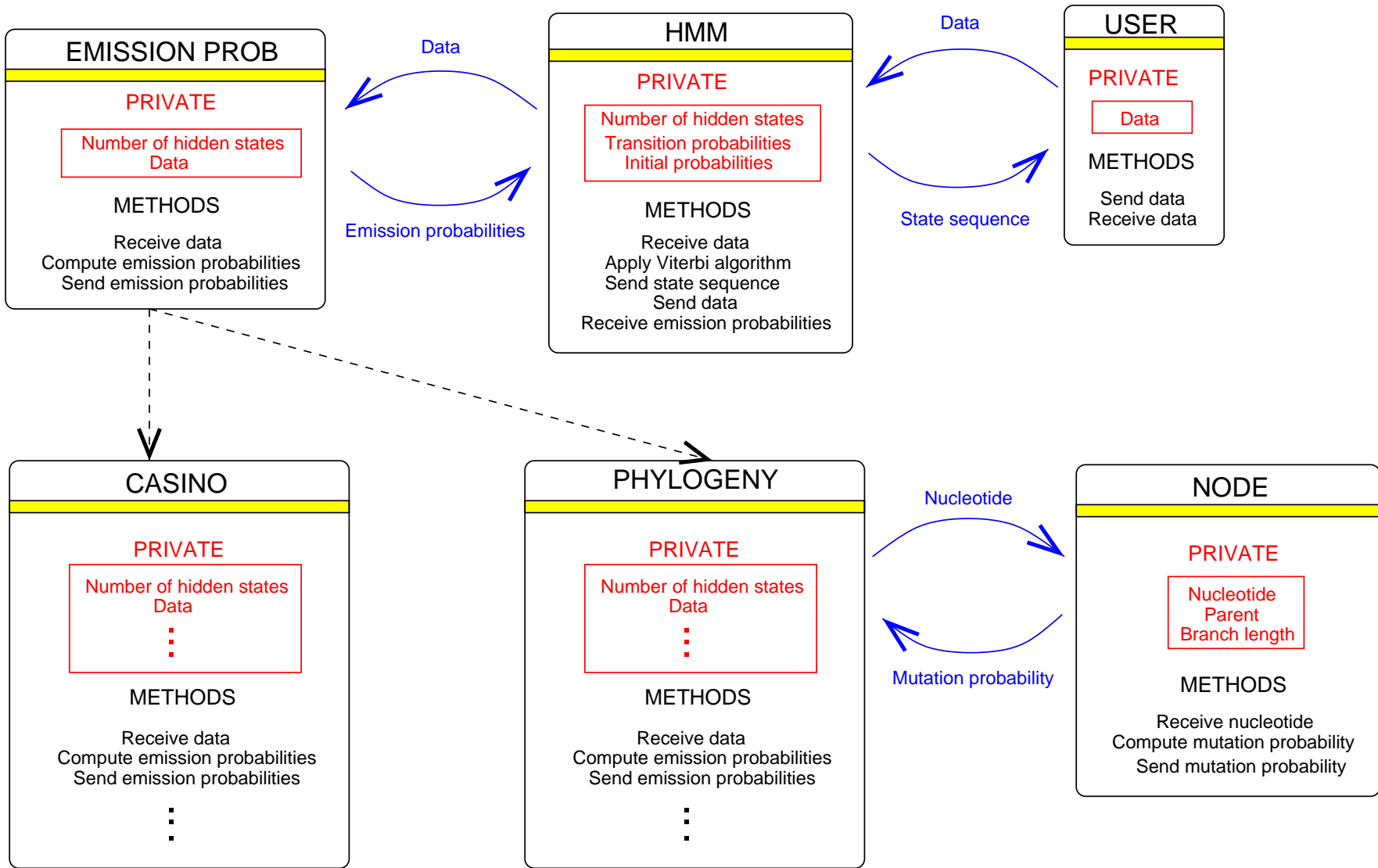
$$= P(y_1 | z_1, w_1) P(y_2 | z_1, w_2) P(z_2 | z_1, w_5) P(y_3 | z_2, w_3) P(y_4 | z_2, w_4) P(z_1)$$

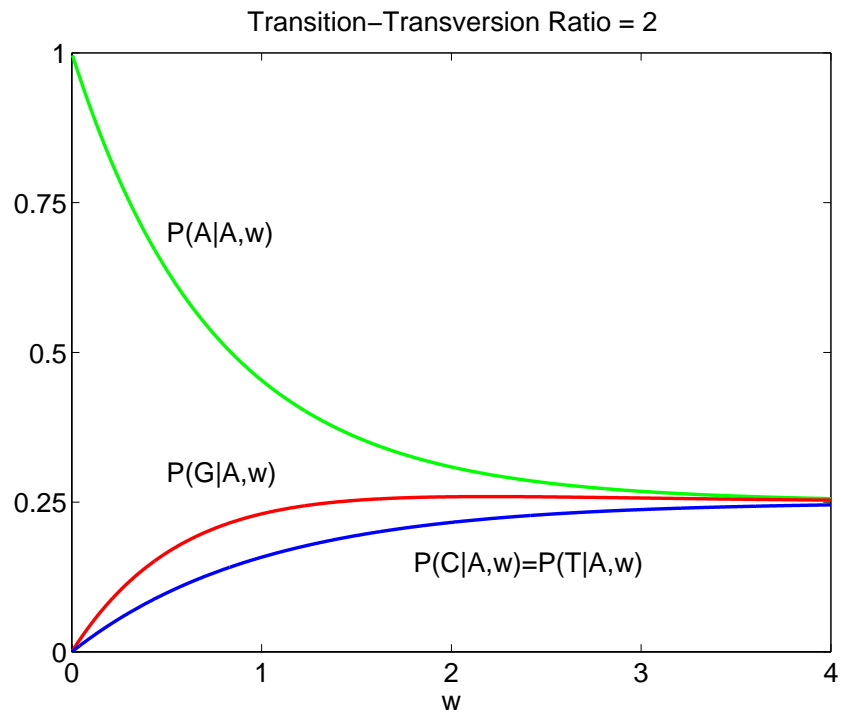
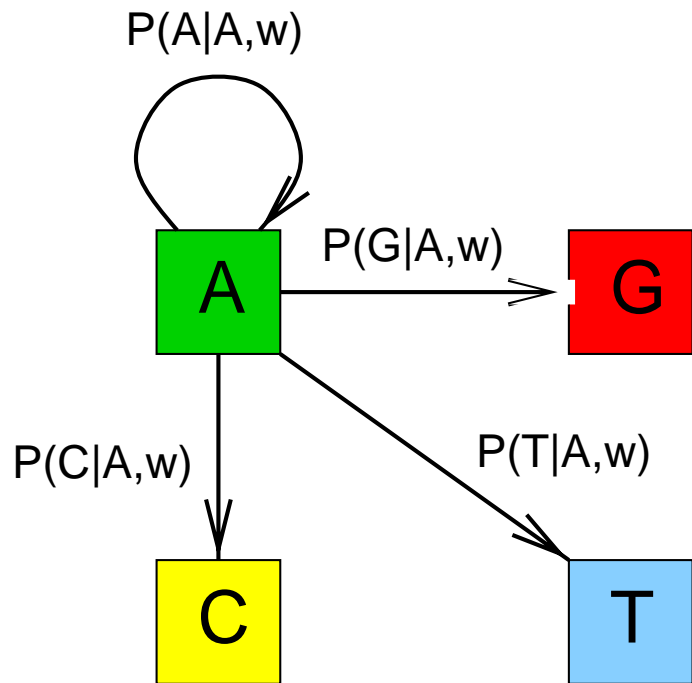


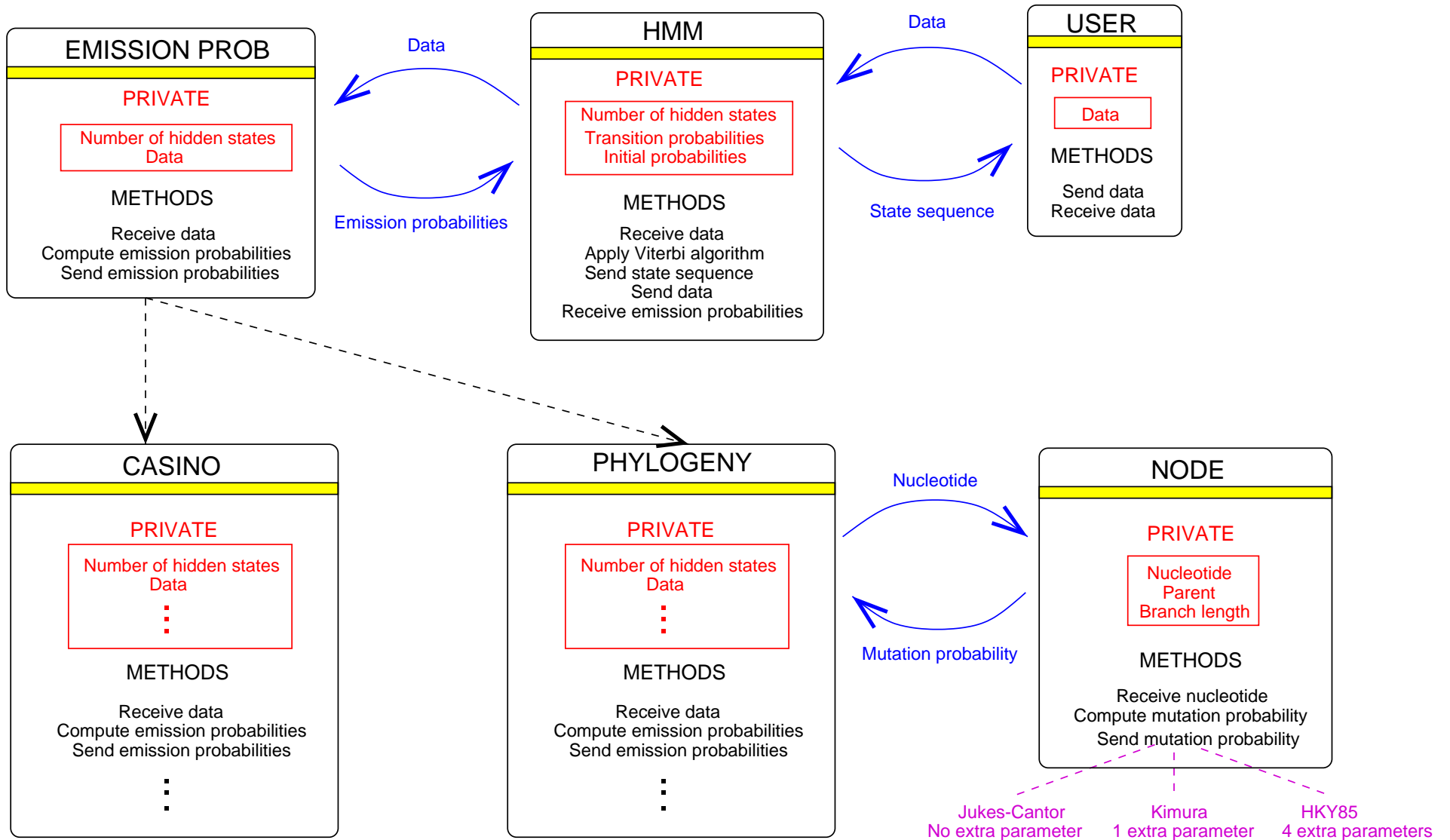


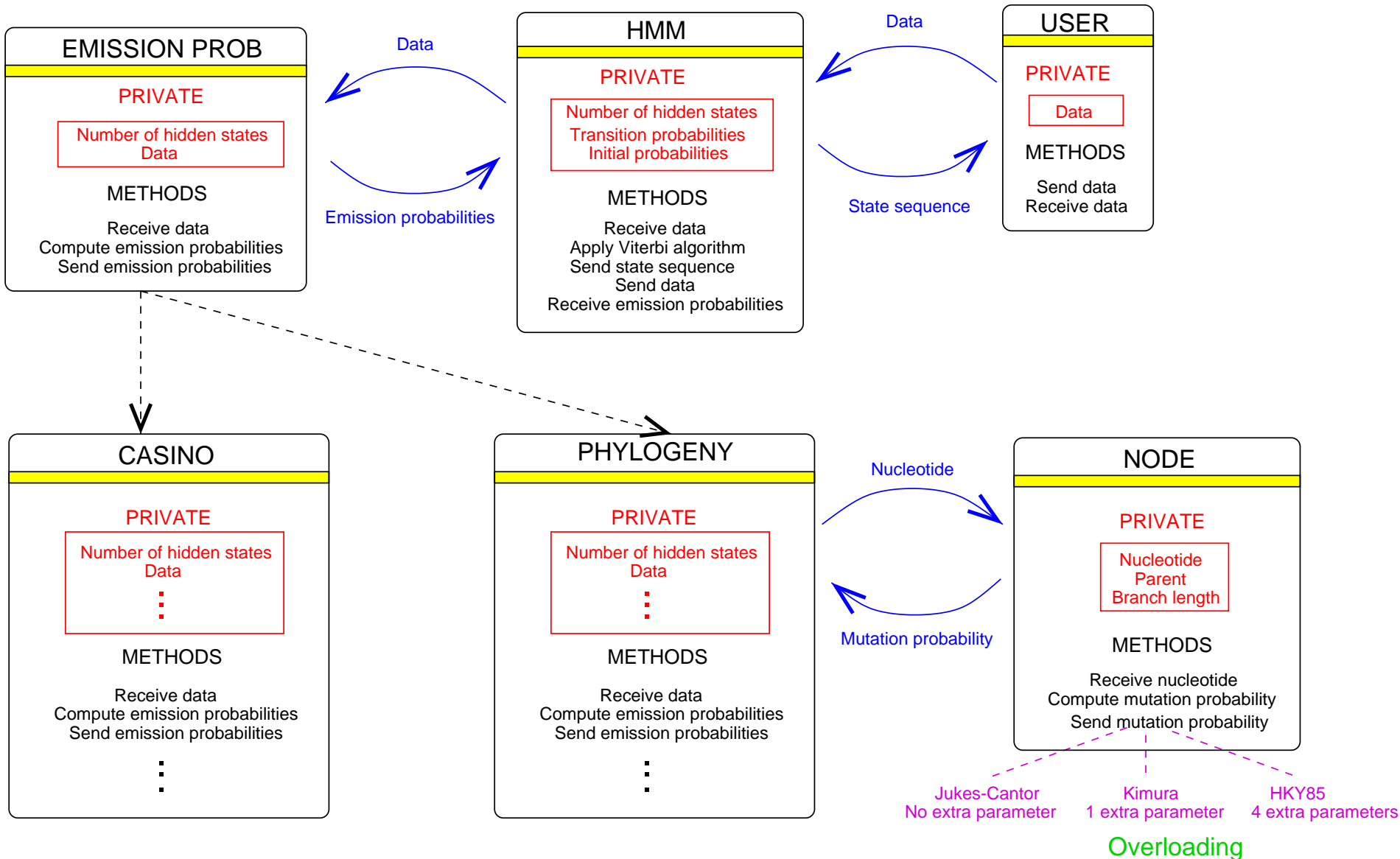


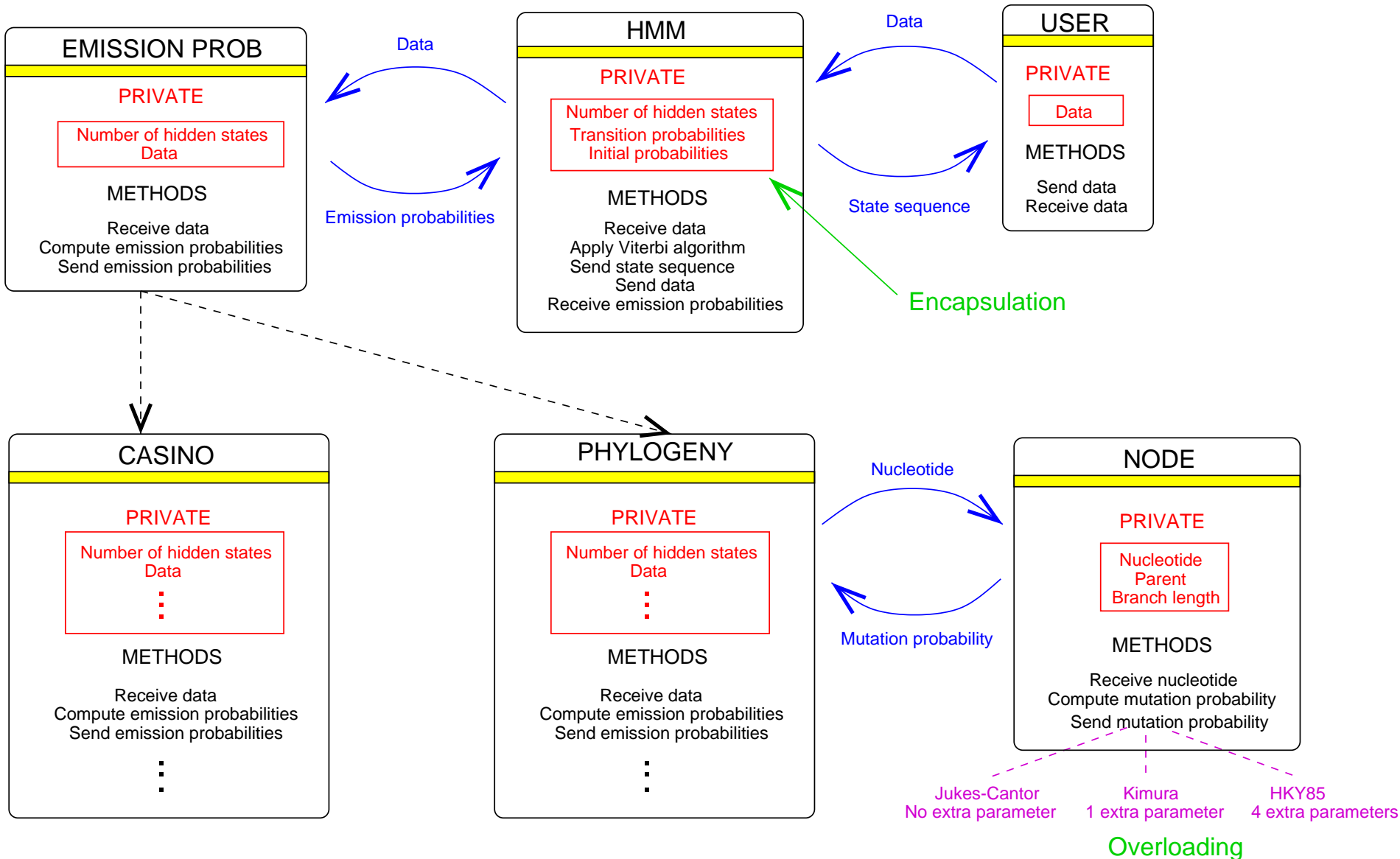


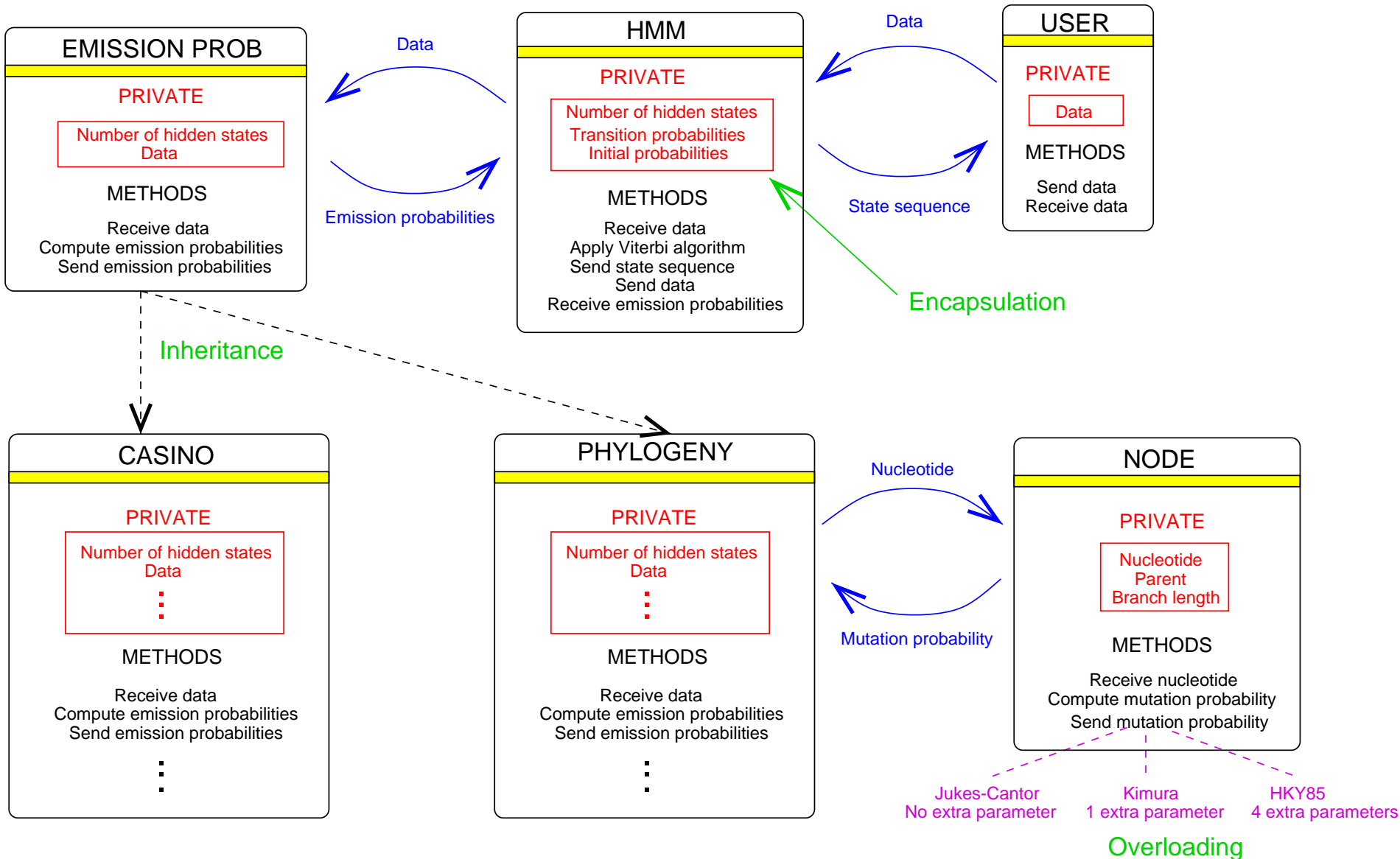


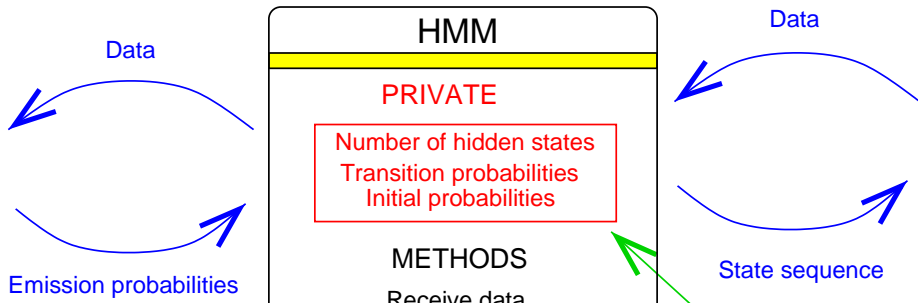
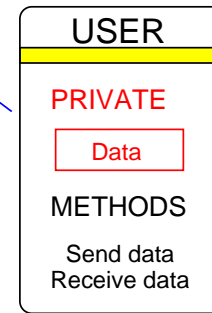
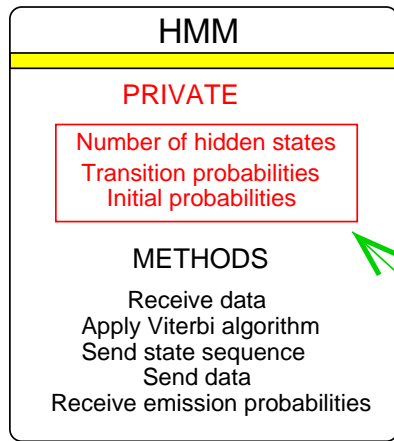
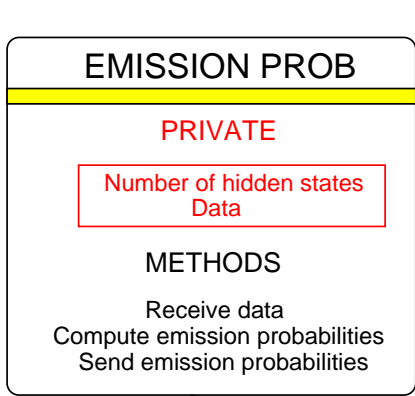








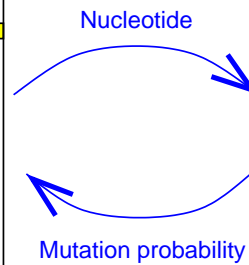
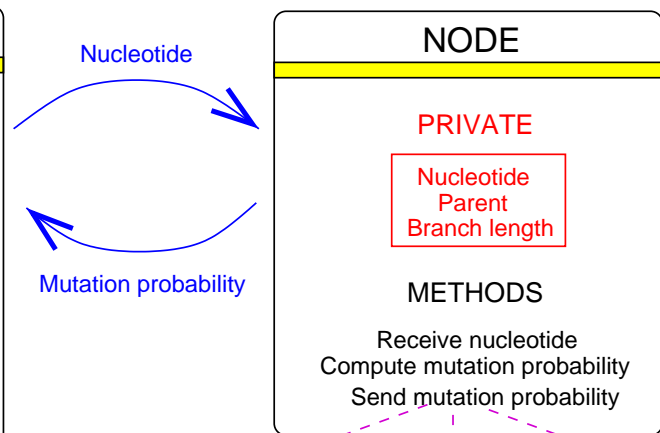
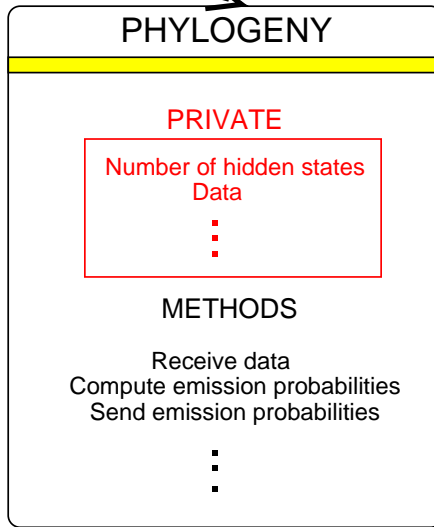
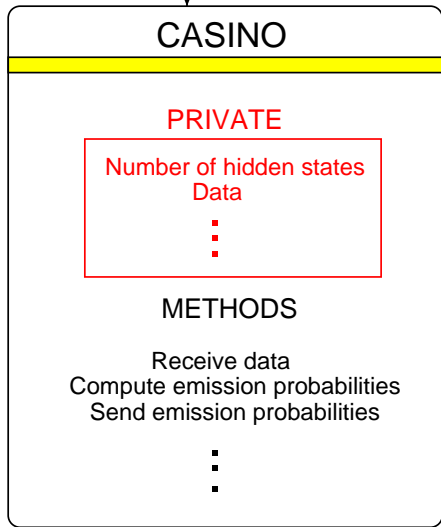




Encapsulation

Inheritance

Polymorphism



Jukes-Cantor
No extra parameter

Kimura
1 extra parameter

HKY85
4 extra parameters

Overloading