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

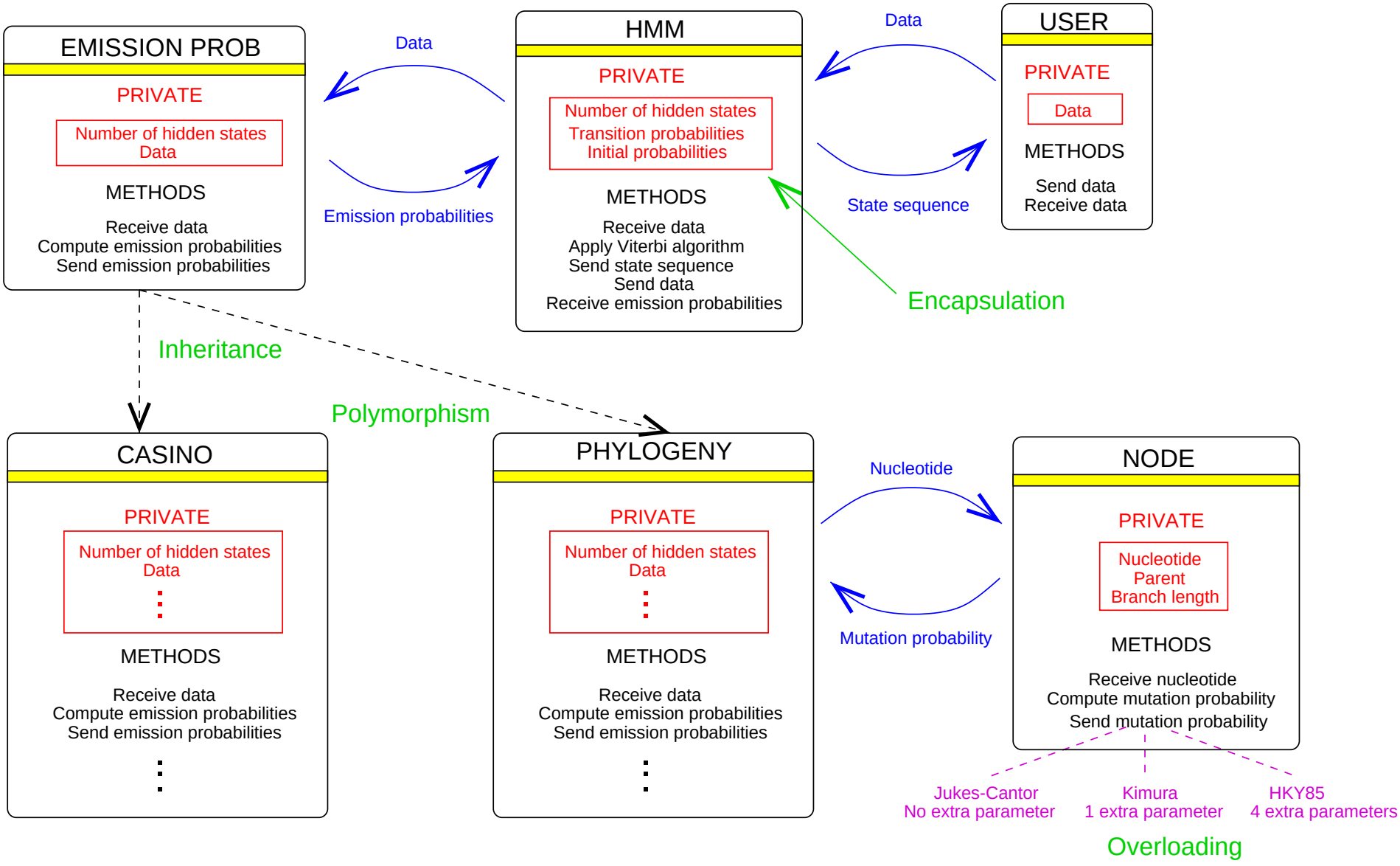
Husmeier, Wright (2001)
Journal of Computational Biology 8, 401-427.

SERAD

Searching for Evidence of Recombination in Alignments of DNA

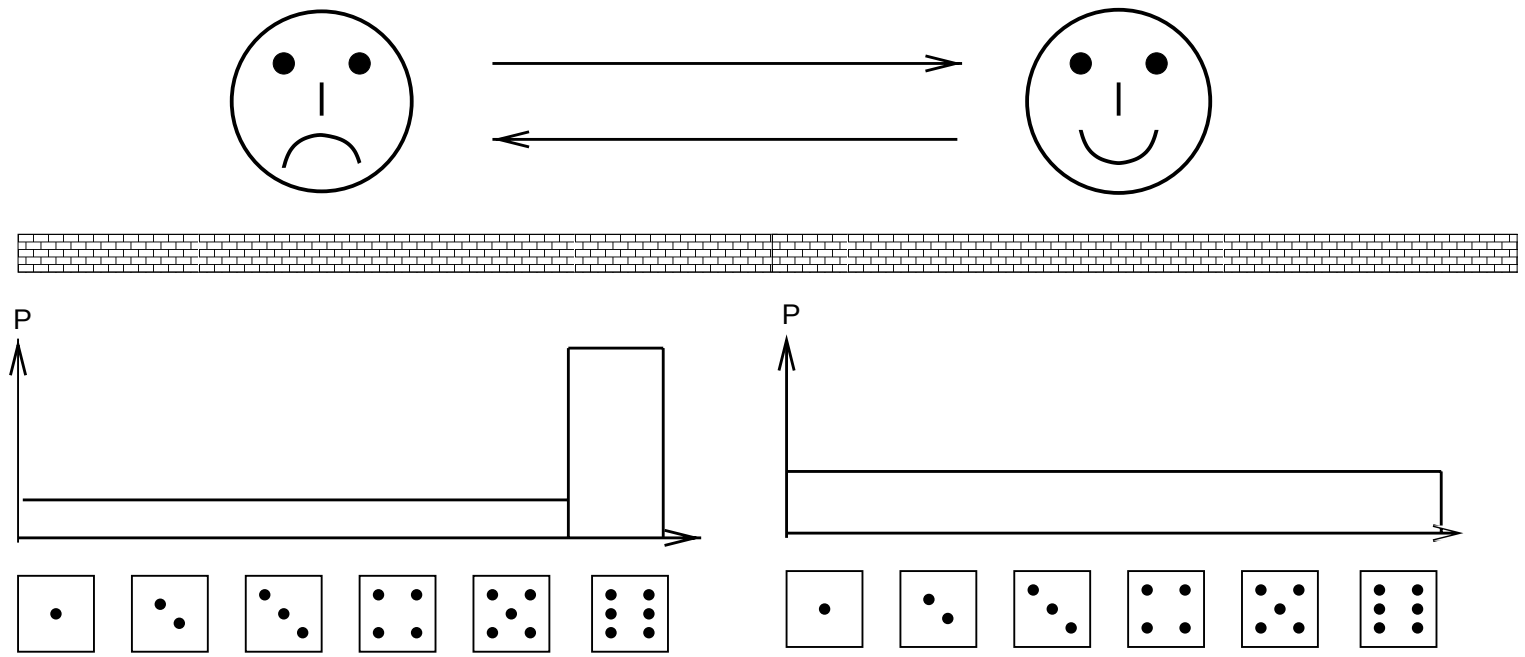
Husmeier, Wright (2001)
Journal of Computational Biology 8, 401-427.

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

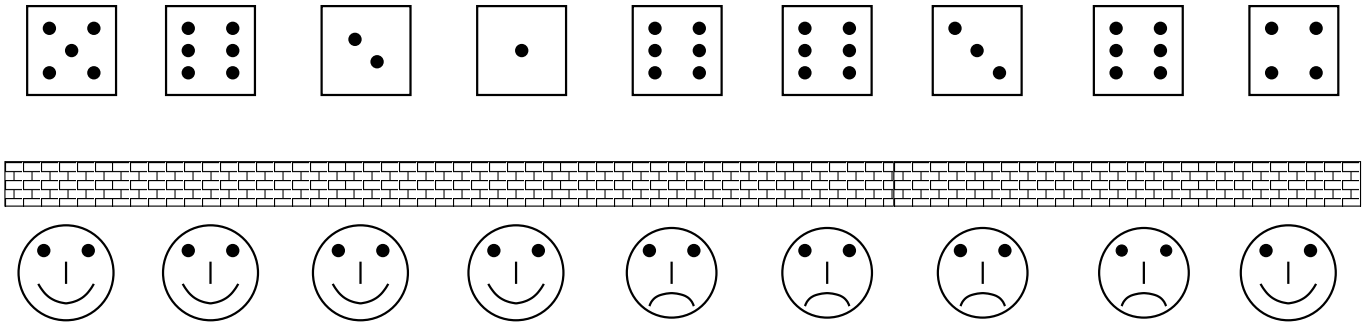

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
Encapsulation
Inheritance
Polymorphism
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters
Overloading

Hidden Markov Models

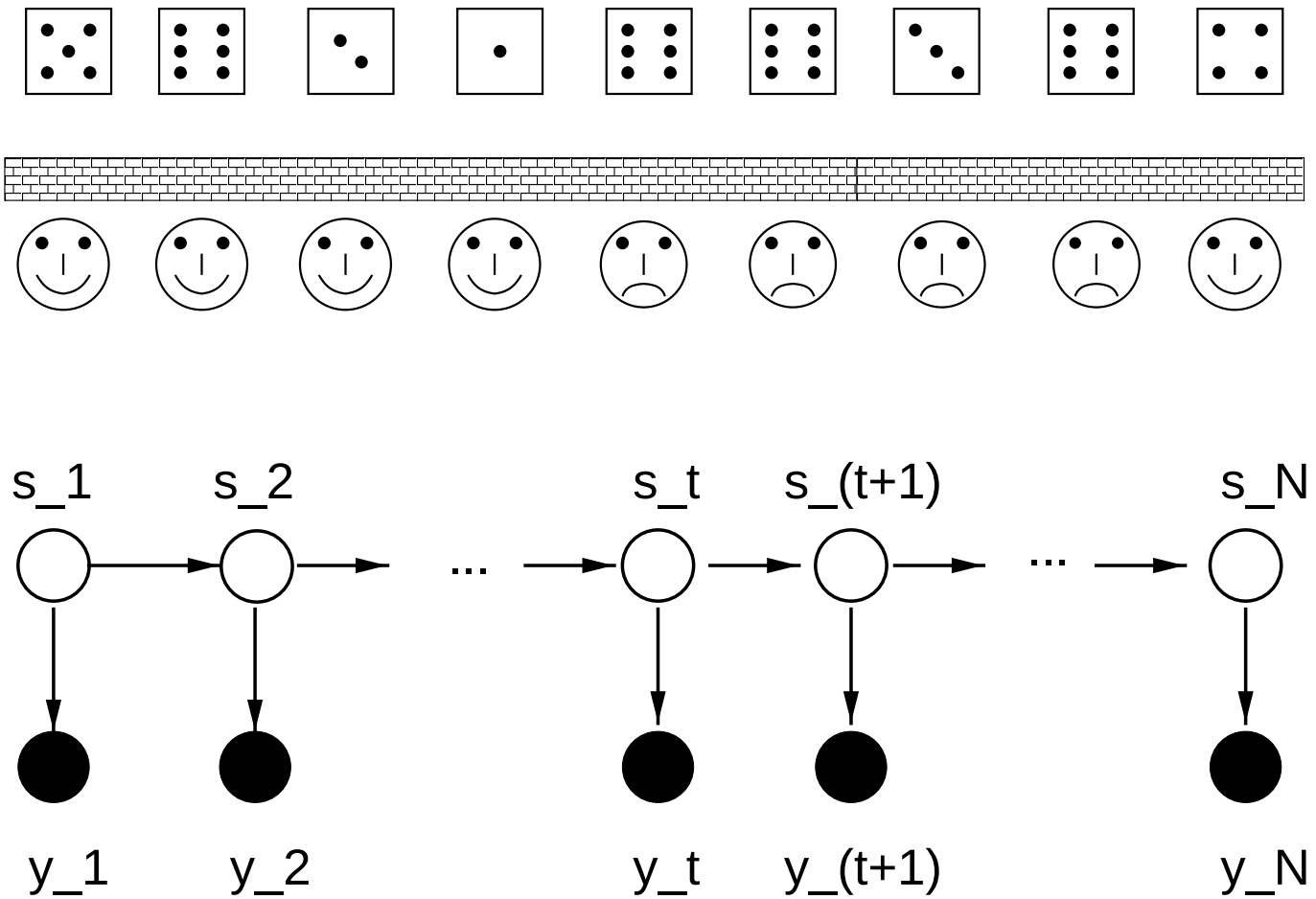
Example: The occasionally corrupt casino



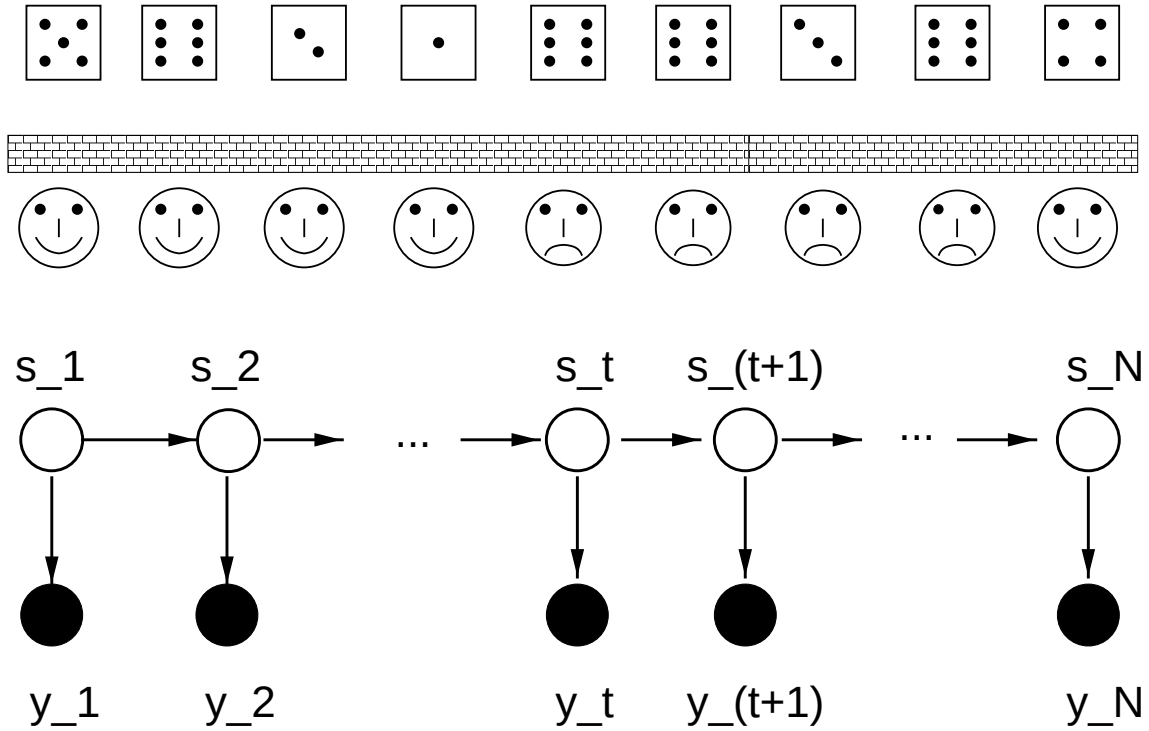
Example: HMM



Example: HMM



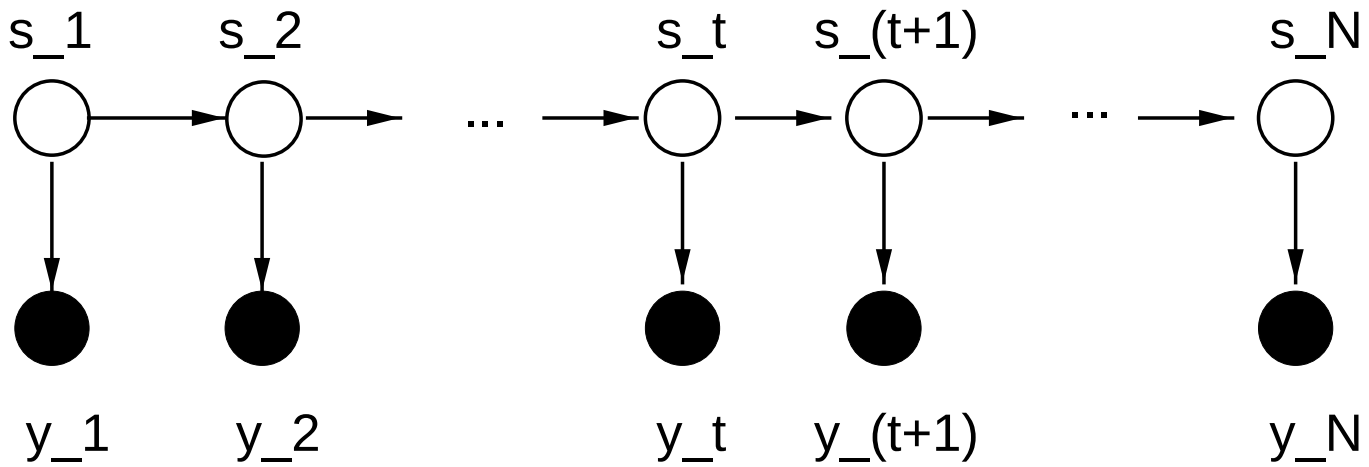
The most likely state sequence



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

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

Factorisation in HMMs

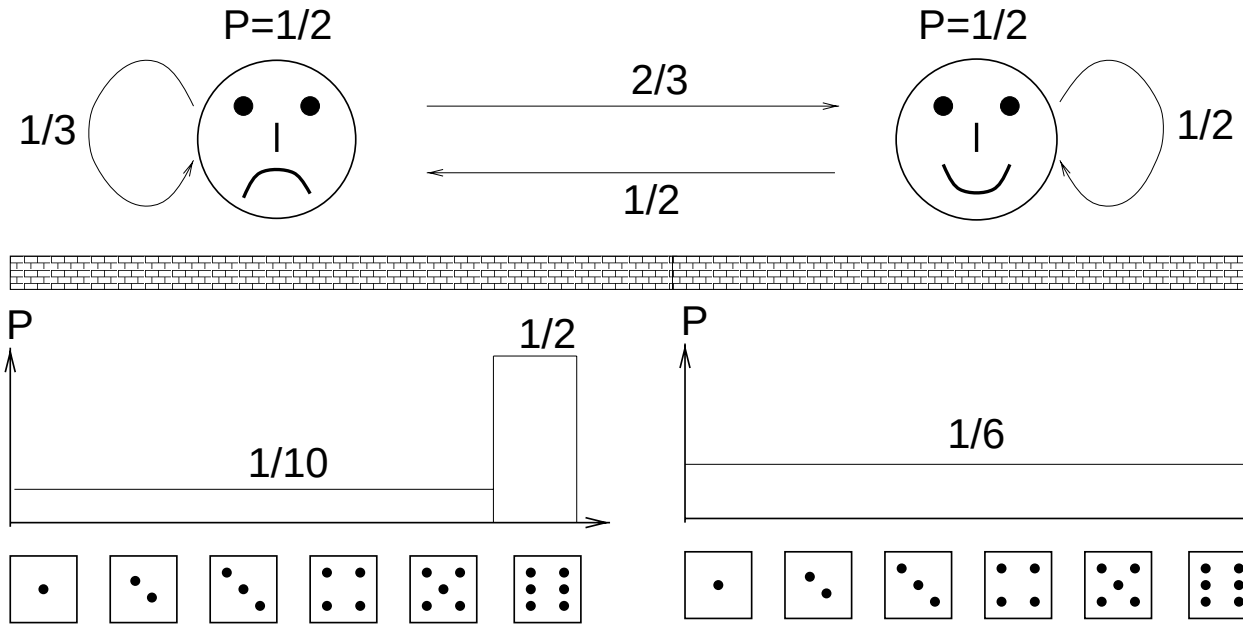


$$P(y_1, \ldots, y_N, S_1, \ldots, S_N) = \prod_{t=1}^{N} P(y_t|S_t) \prod_{t=2}^{N} P(S_t|S_{t-1}) P(S_1)$$

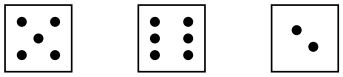
Viterbi algorithm $\longrightarrow P(S_1, \ldots, S_N|y_1, \ldots, y_N)$

Computational complexity and example

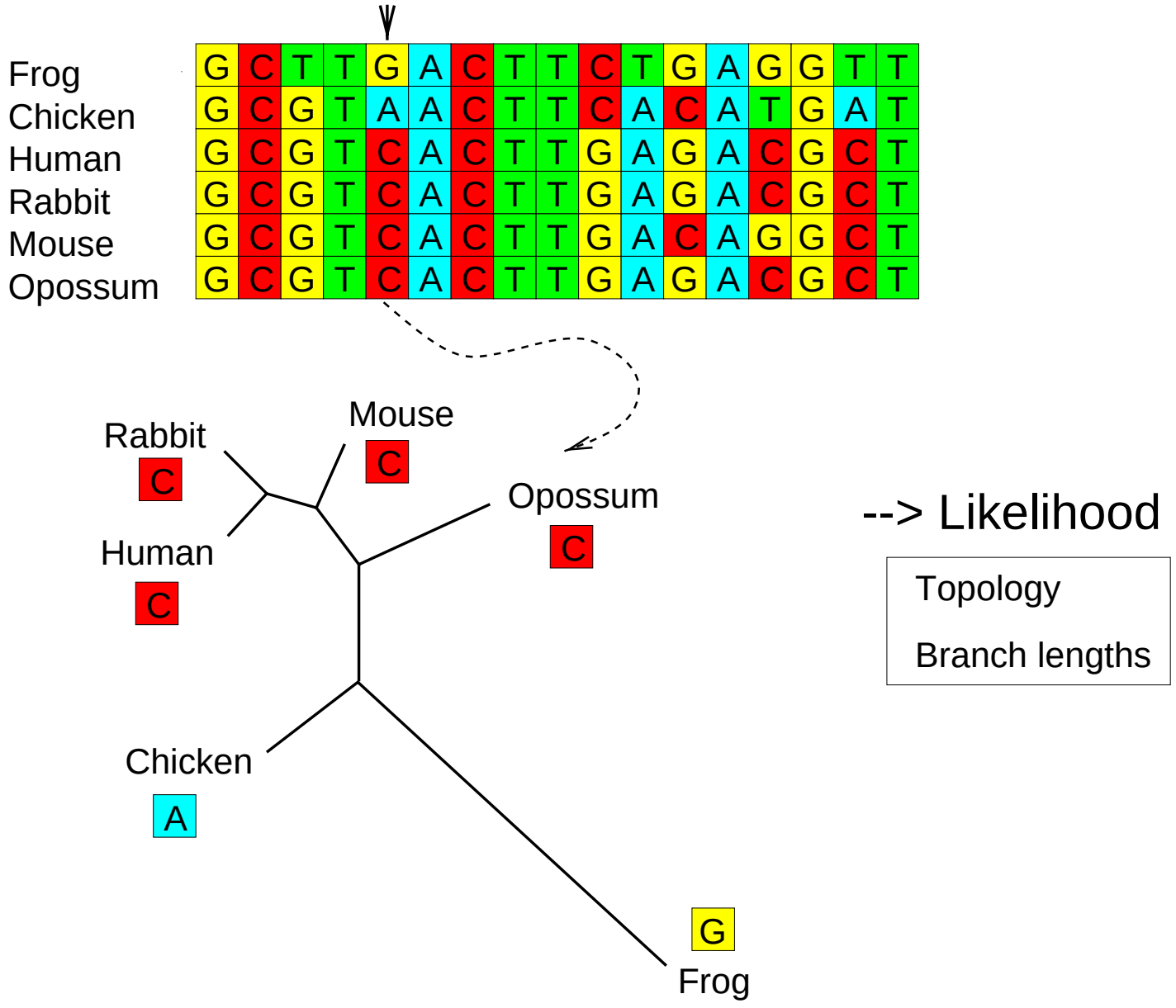
Computation complexity: $|\mathcal{H}|^N \longrightarrow N \times |\mathcal{H}|^2$



Observed sequence:

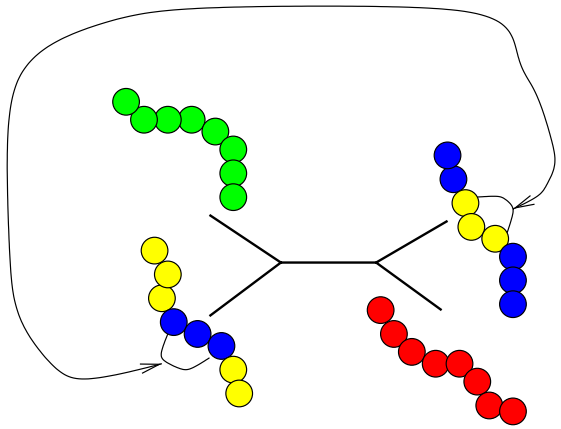


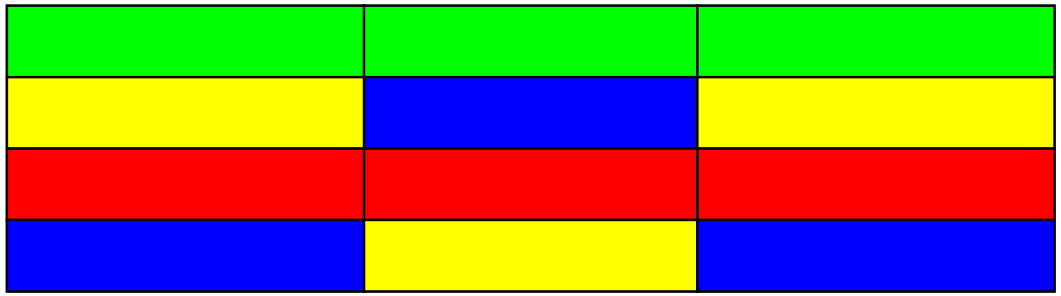
Phylogenetic trees

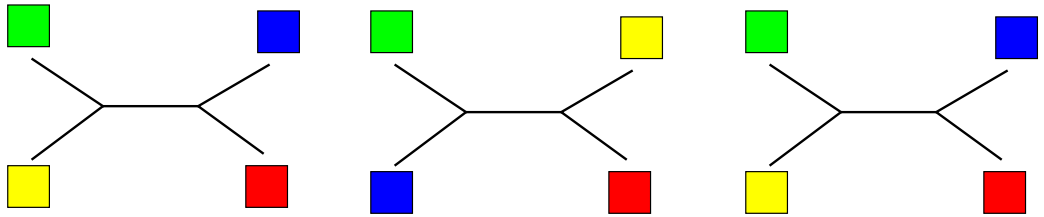

Frog
Chicken
Human
Rabbit
Mouse
Opossum
G C T T G A C T T C T G A G G T T
G C G T A A C T T C A C A T G A T
G C G T C A C T T G A G A C G C T
G C G T C A C T T G A G A C G C T
G C G T C A C T T G A C A G G C T
G C G T C A C T T G A G A C G C T
Rabbit
C
Mouse
C
Opossum
C
Human
C
Chicken
A
G
Frog
--> Likelihood
Topology
Branch lengths

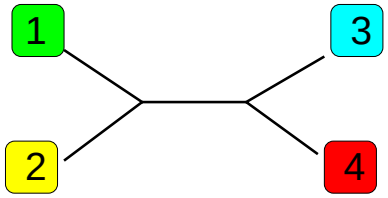
Recombination

Recombination

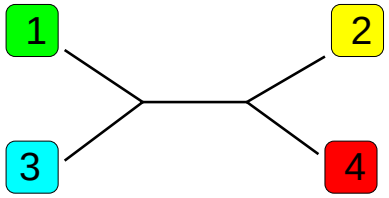




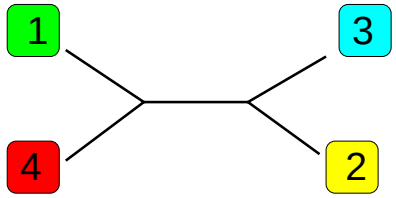




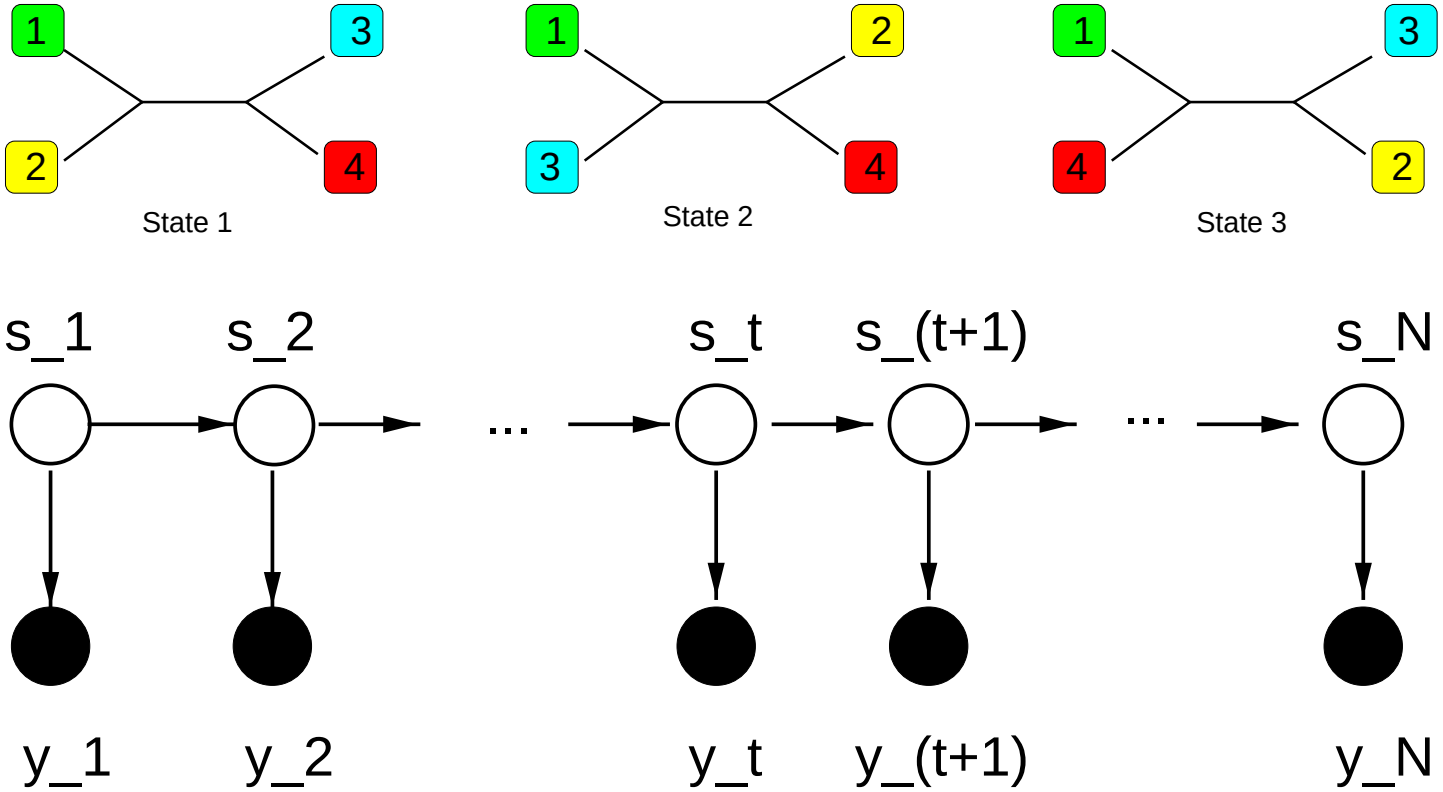
State 1



State 2

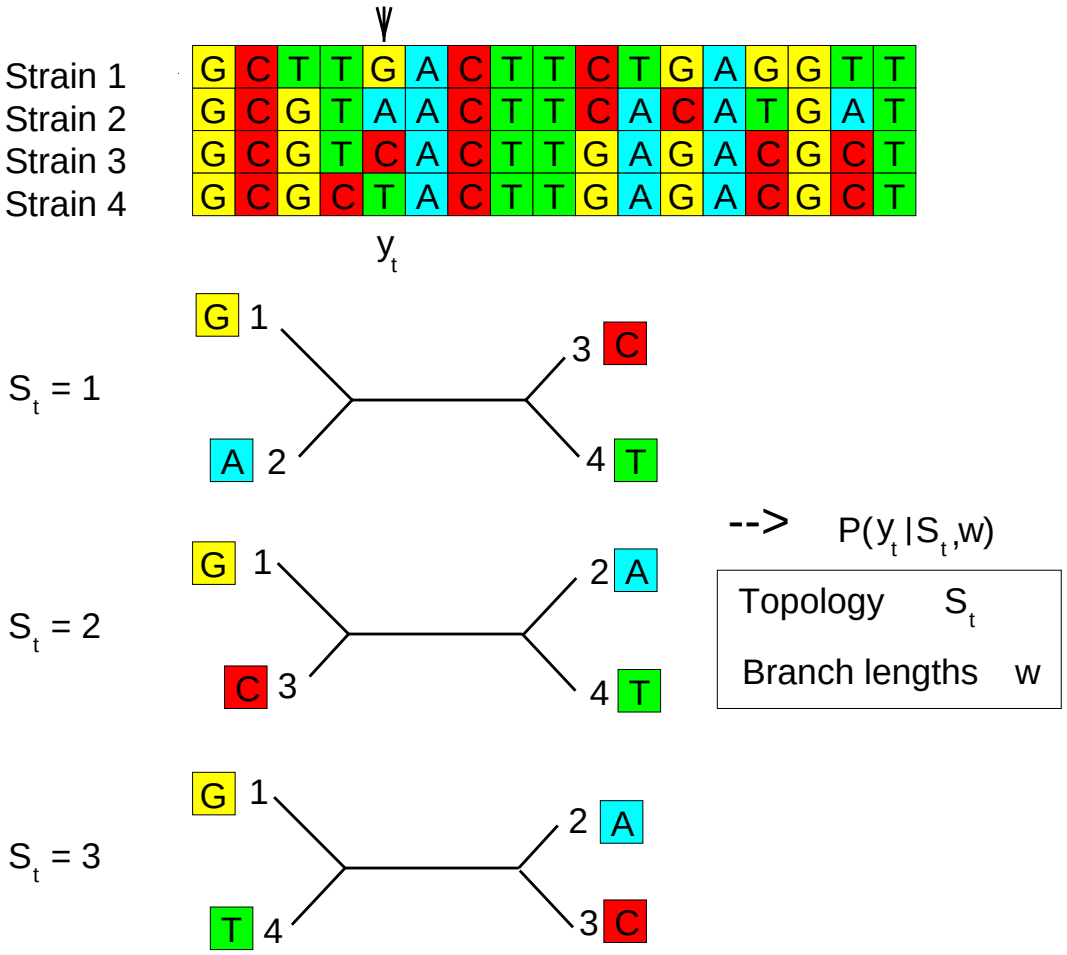


State 3

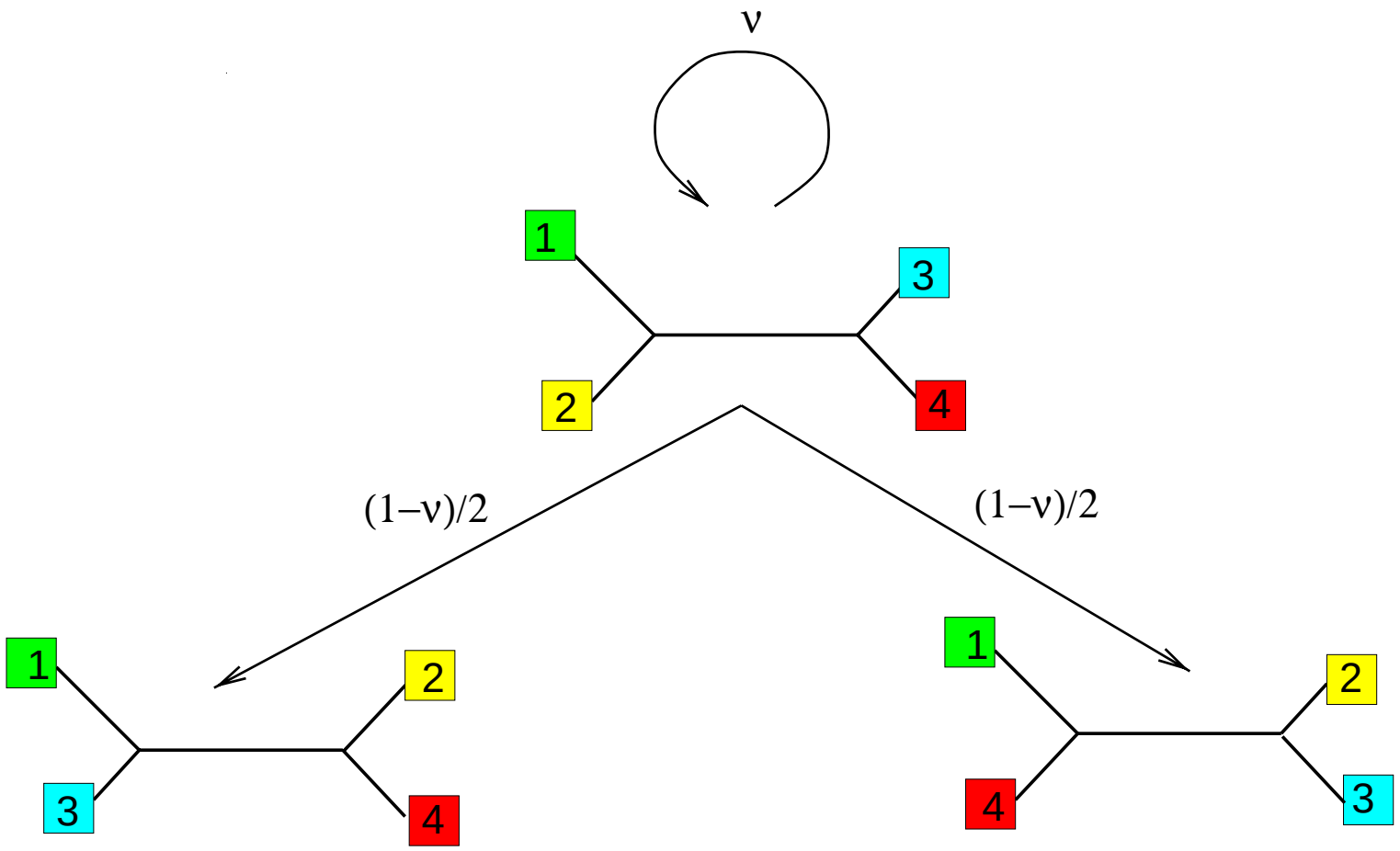


AGCATCGTTCTATTTTACCGGCTCCCG
TGTGTCGCTCAAGATTGCCATCGCGCG
TGTCGTGGTCTAGATTGCCATCGCGCG
TGTATCGCTCTAGTTTGCCAGCTCCCG

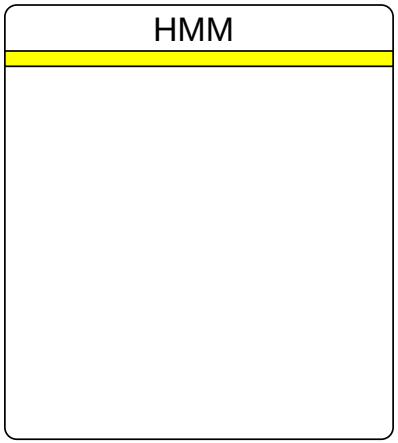
Emission probabilities

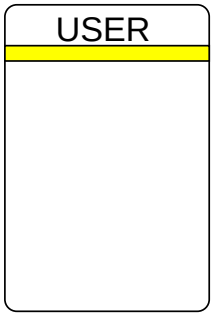


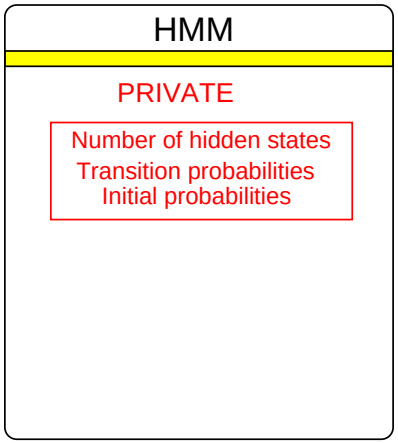
Transition probabilities

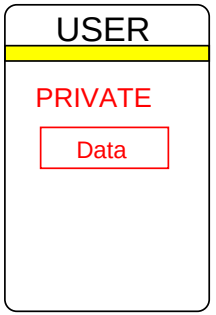


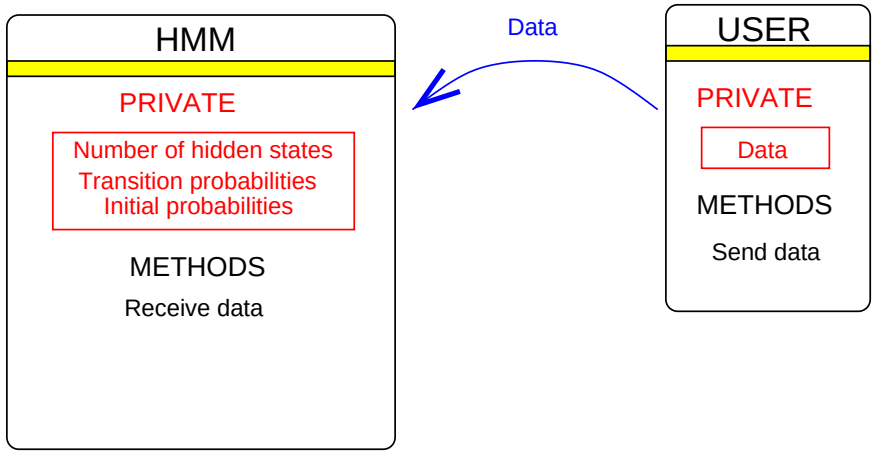
Object-oriented programming

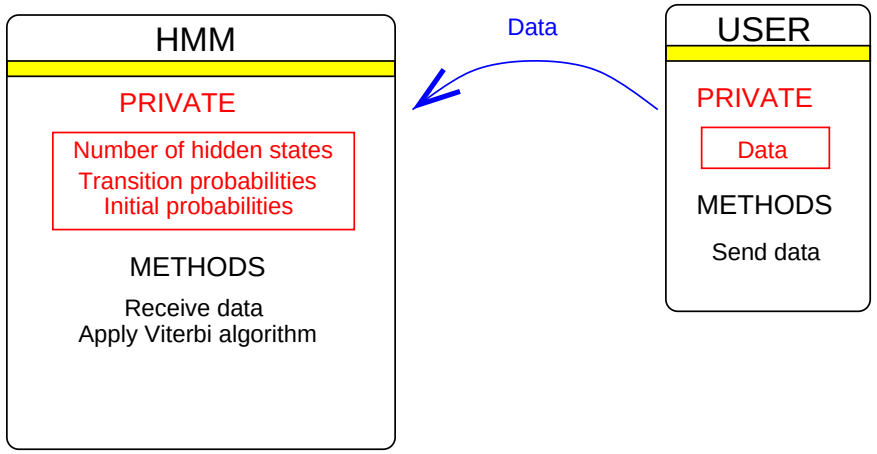

HMM

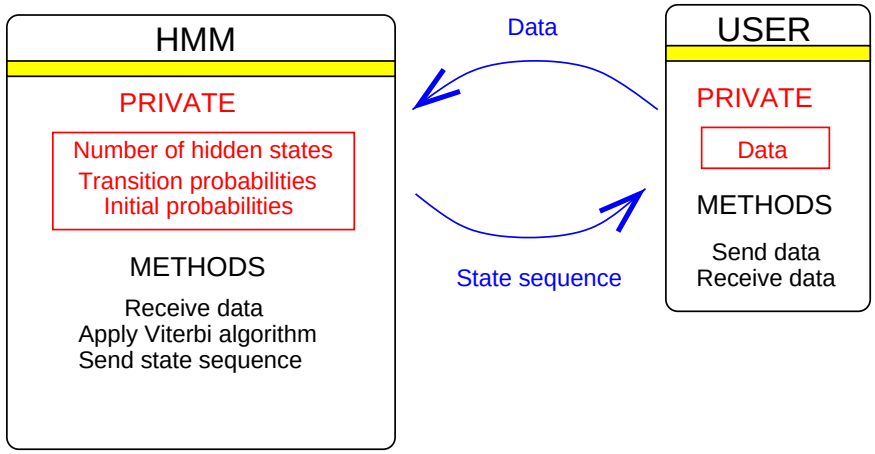

USER

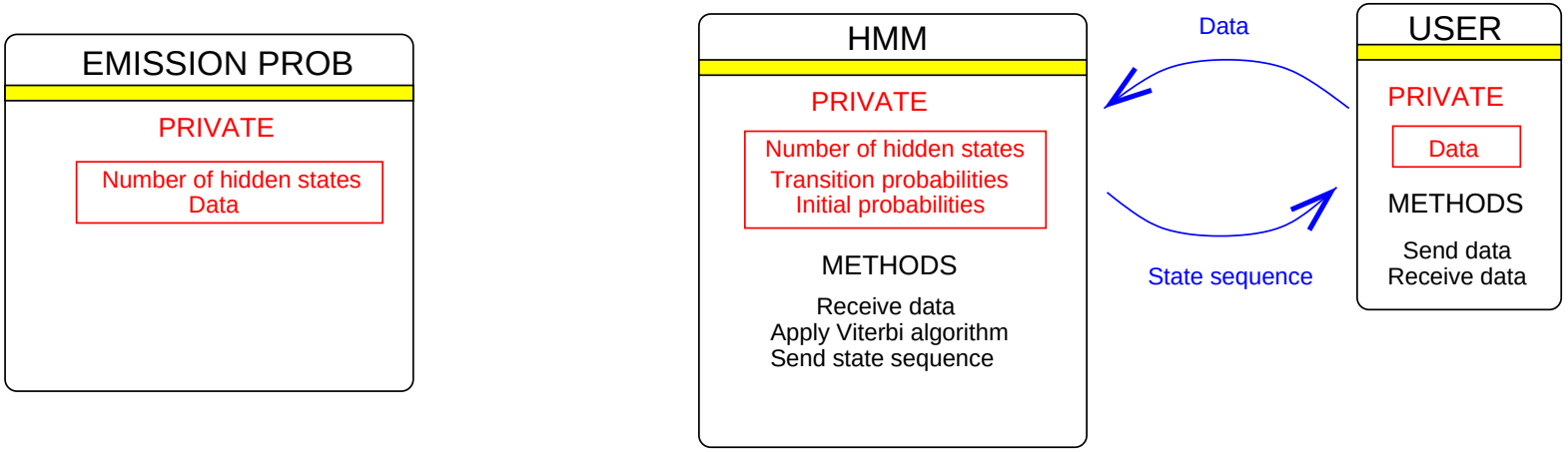

HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities

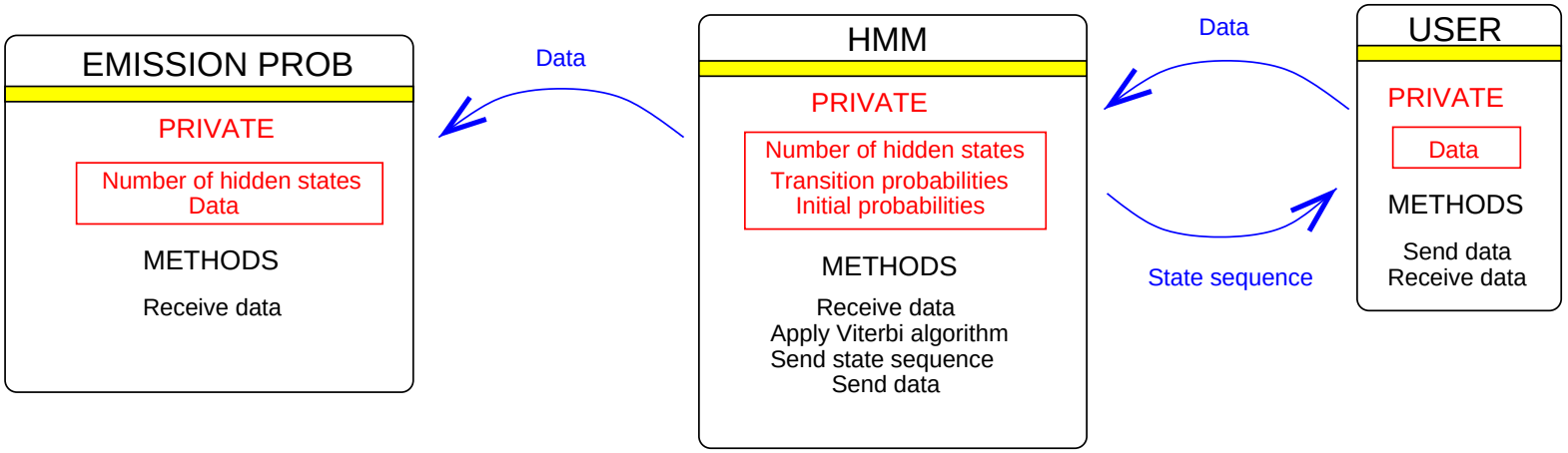

USER
PRIVATE
Data

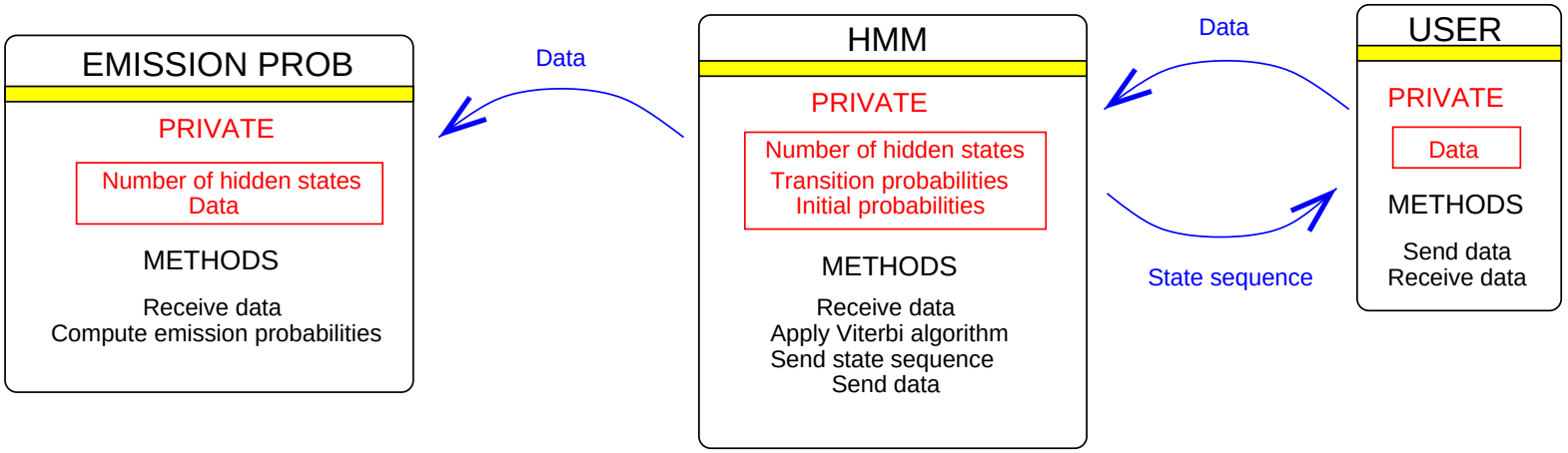

HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Data
USER
PRIVATE
Data
METHODS
Send data

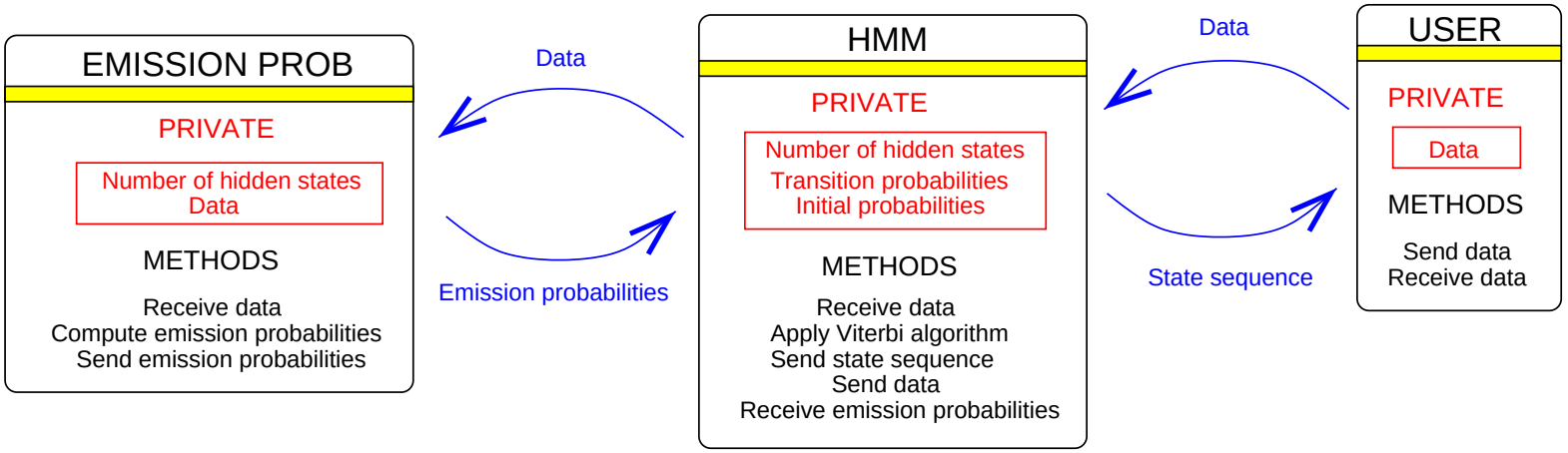

HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Data
USER
PRIVATE
Data
METHODS
Send data

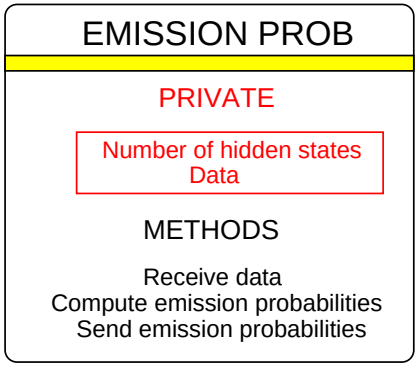

HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data

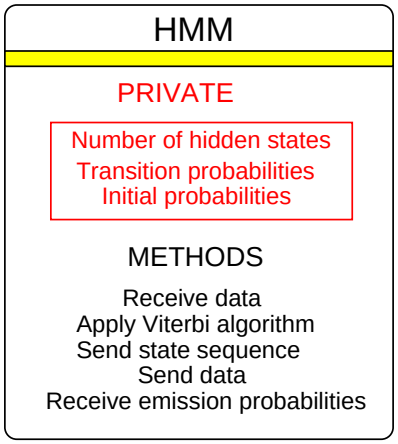

EMISSION PROB
PRIVATE
Number of hidden states
Data
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data

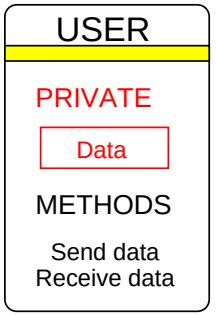

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Data
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data

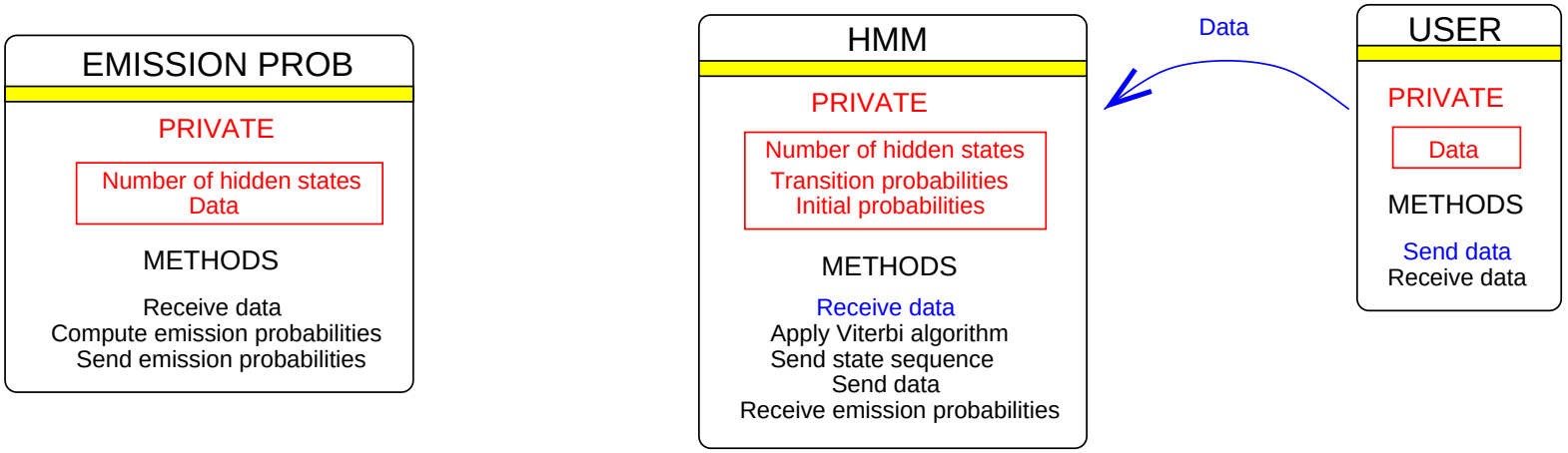

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Data
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data

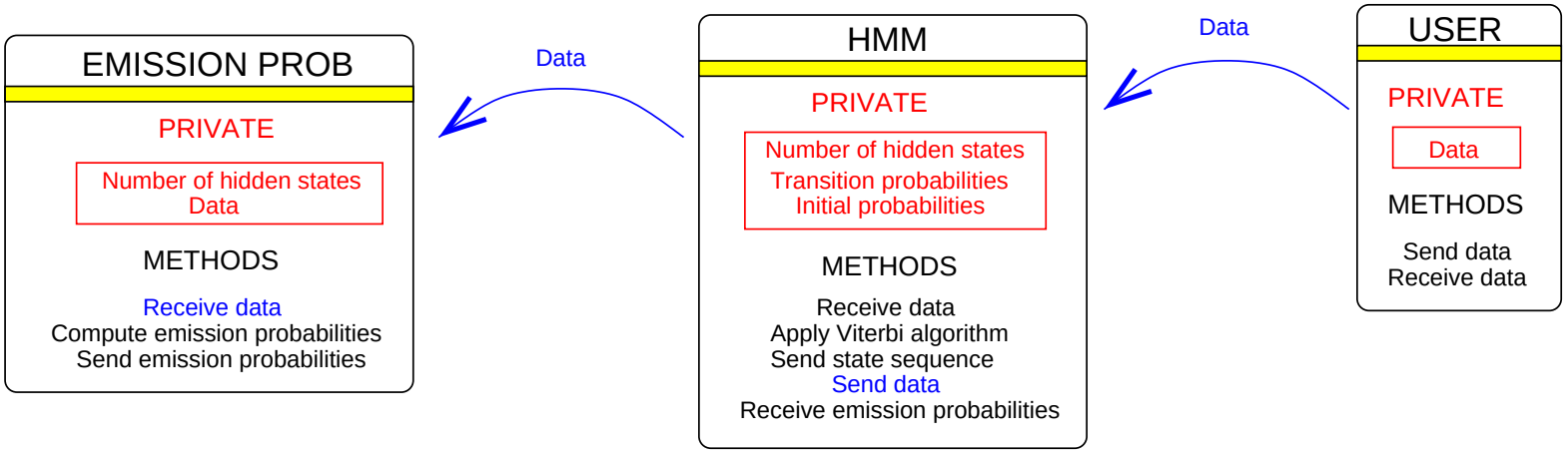

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data

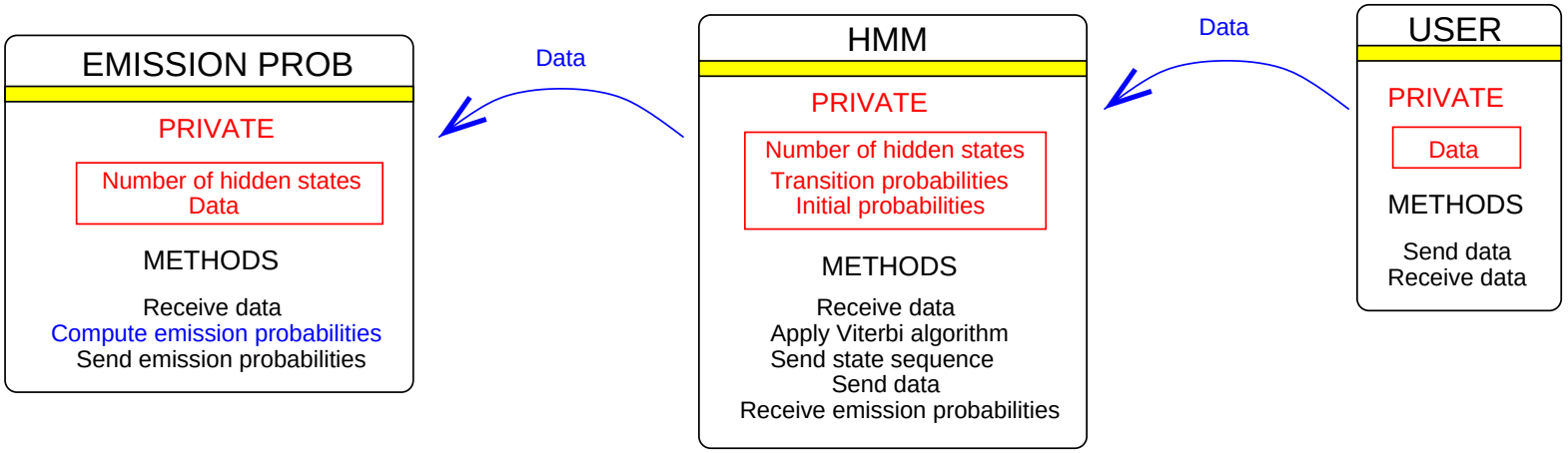

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities

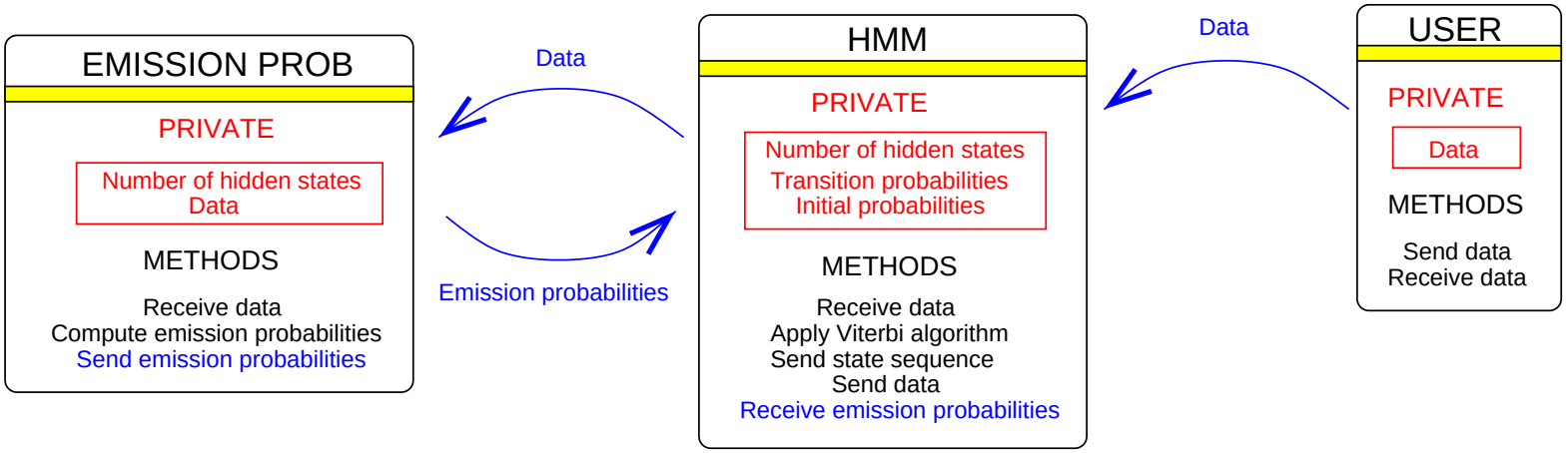

HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities

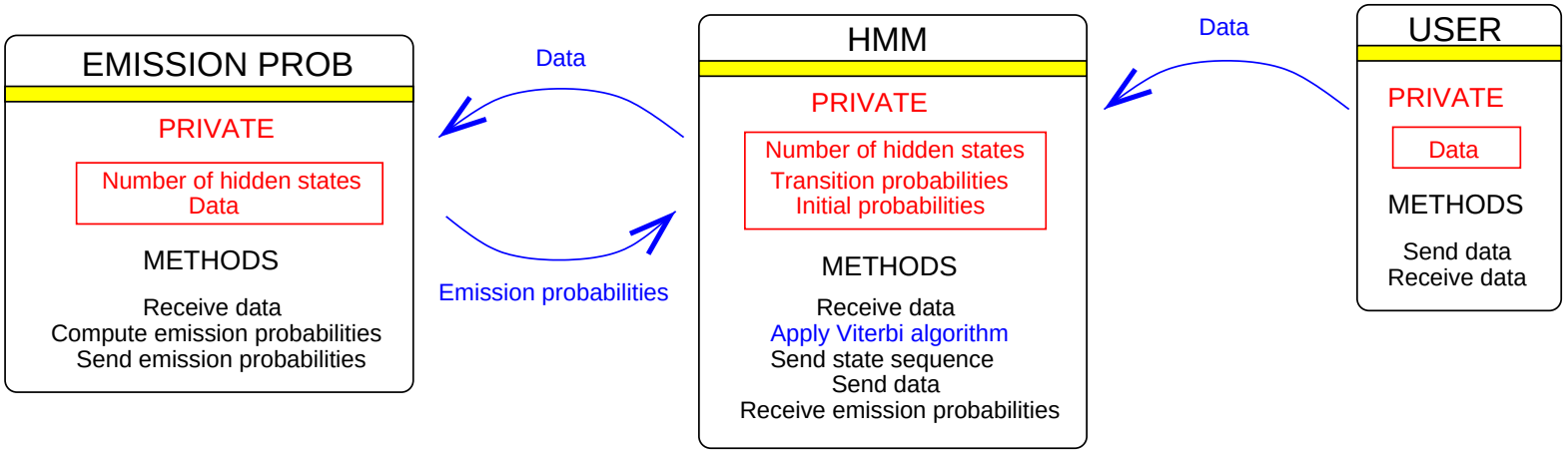

USER
PRIVATE
Data
METHODS
Send data
Receive data

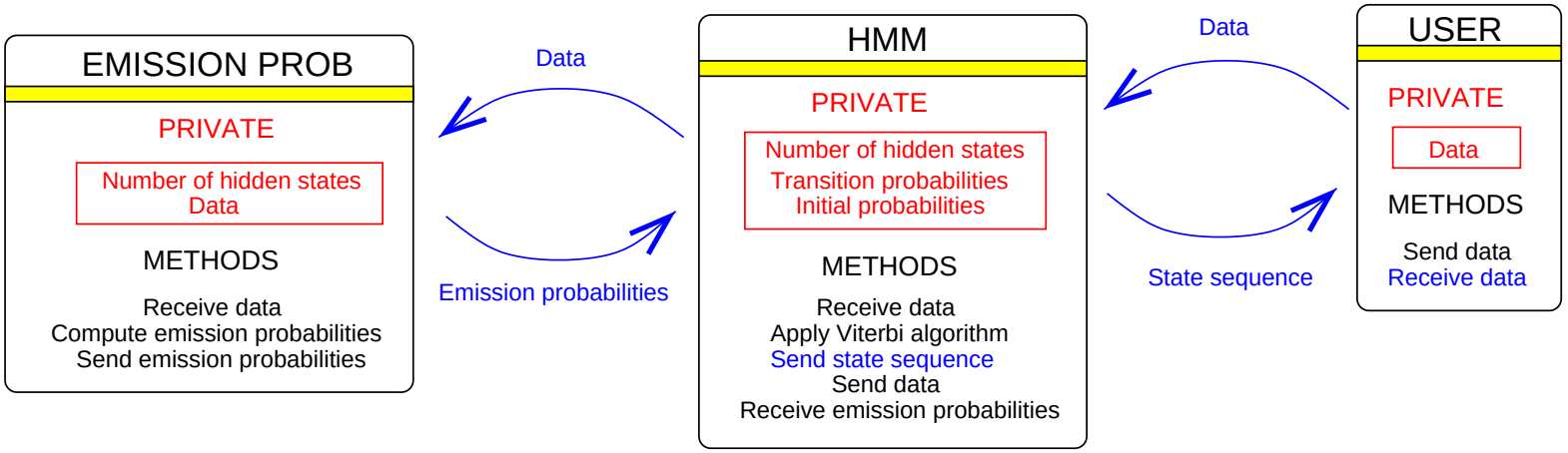

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
USER
PRIVATE
Data
METHODS
Send data
Receive data

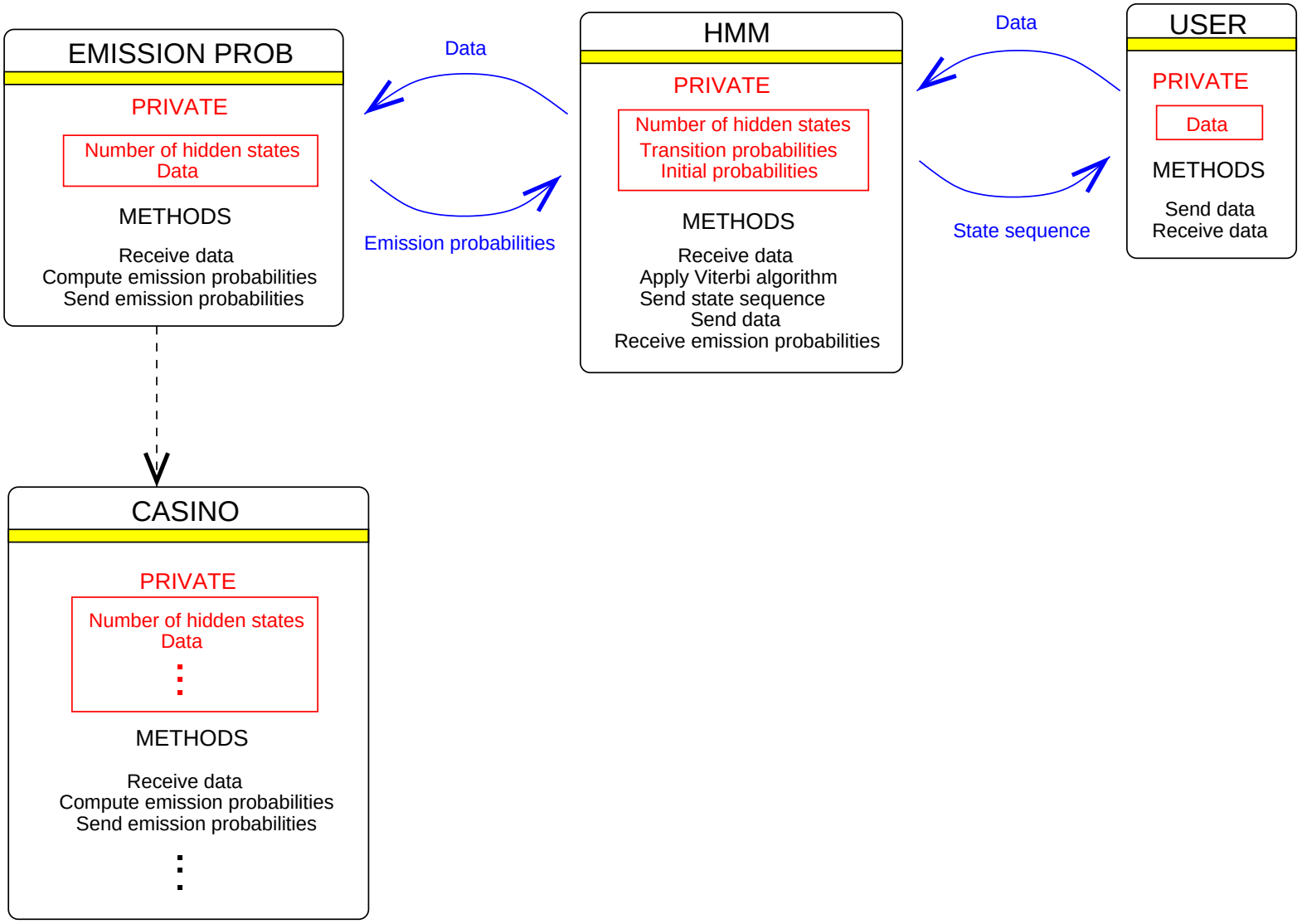

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
USER
PRIVATE
Data
METHODS
Send data
Receive data

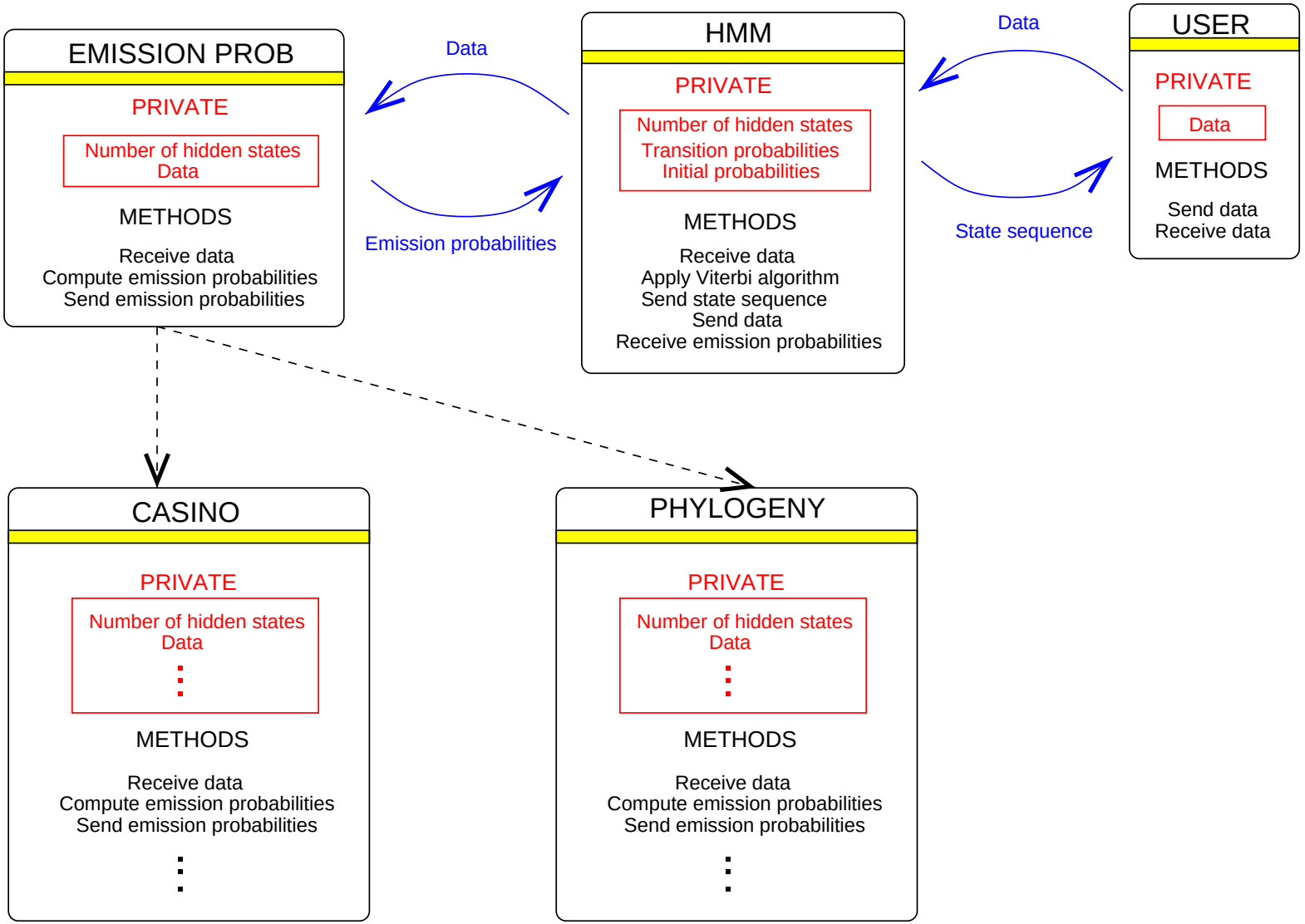

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
USER
PRIVATE
Data
METHODS
Send data
Receive data

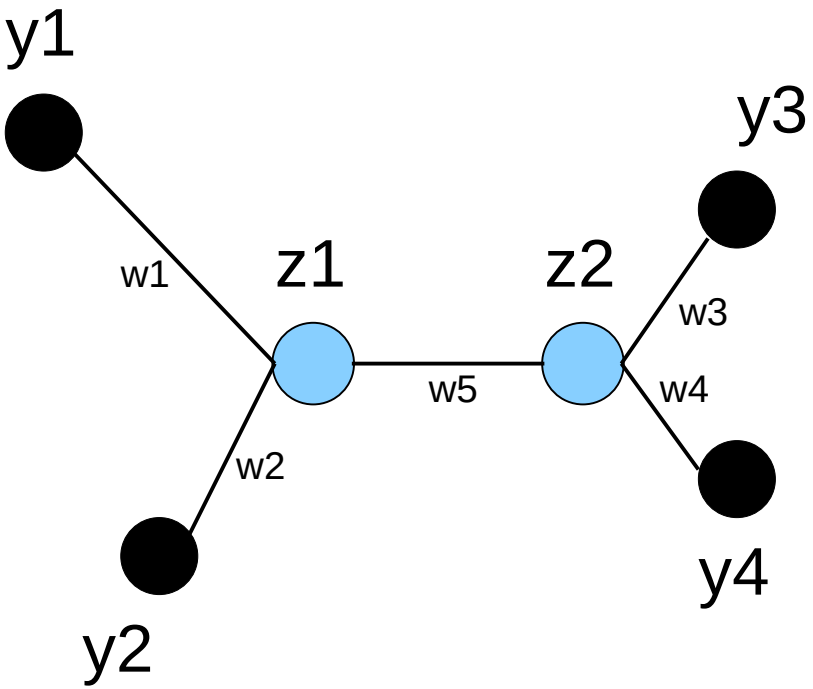

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
USER
PRIVATE
Data
METHODS
Send data
Receive data


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
USER
PRIVATE
Data
METHODS
Send data
Receive data

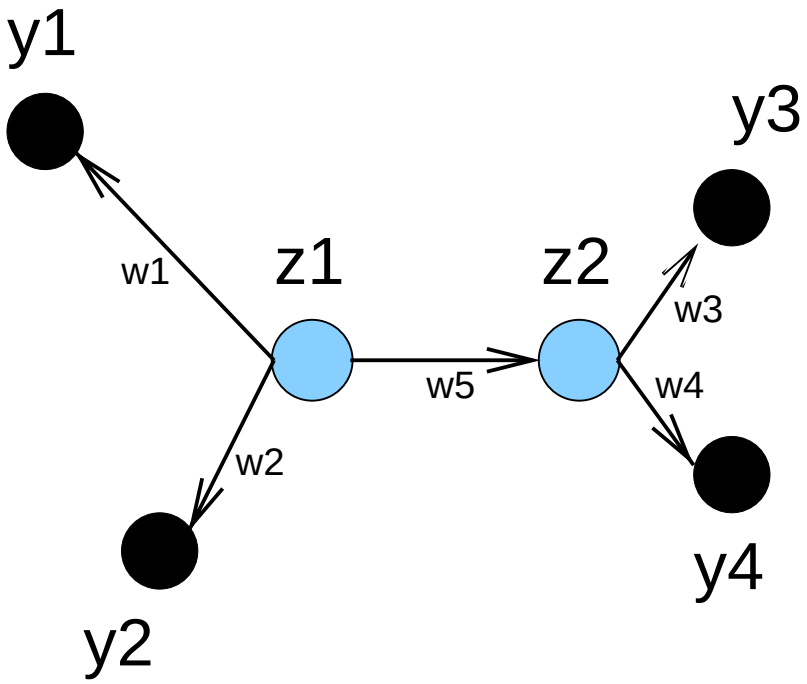

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities

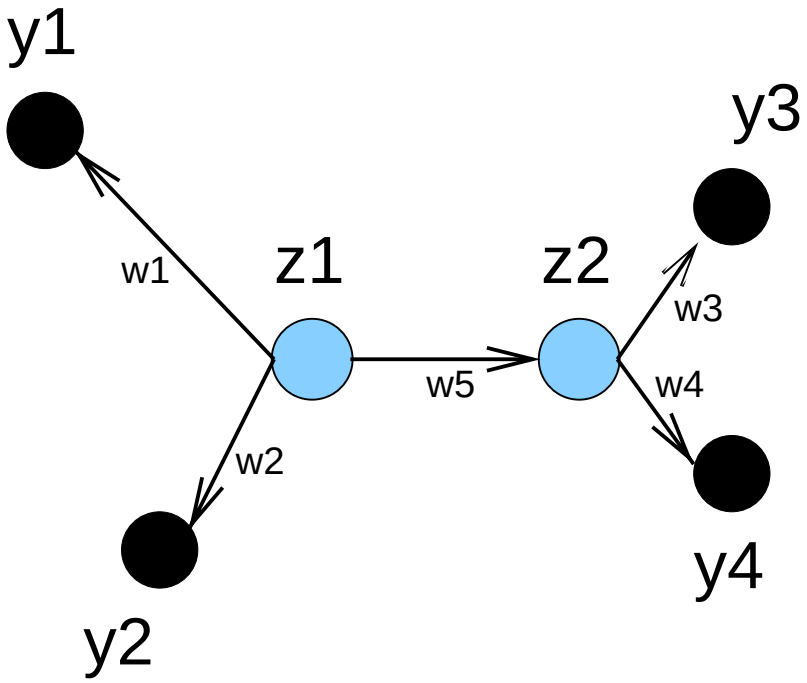

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities



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

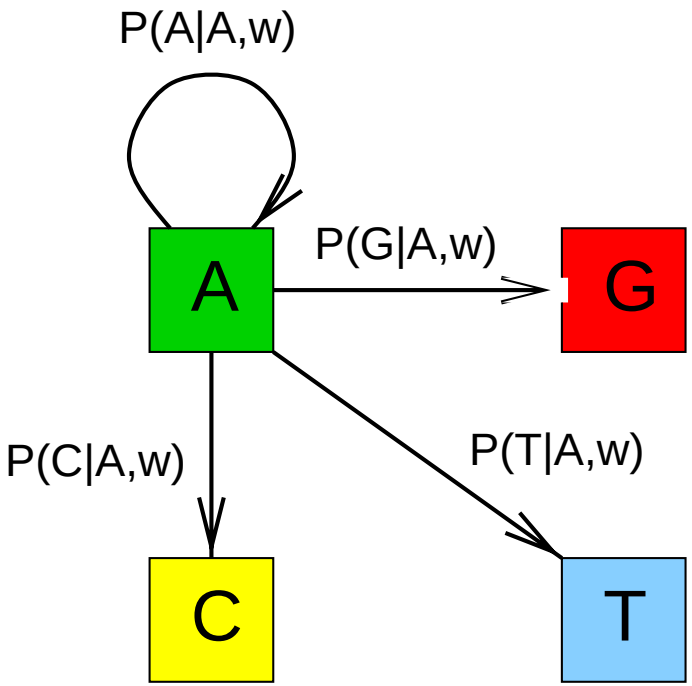


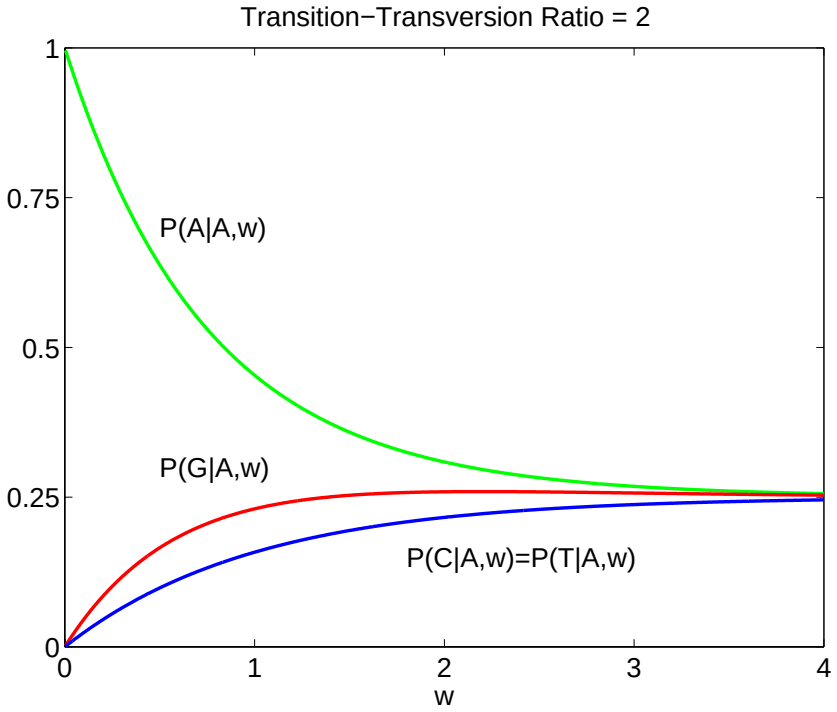
$$P(y_1, y_2, y_3, y_4, z_1, z_2 | \mathbf{w})$$

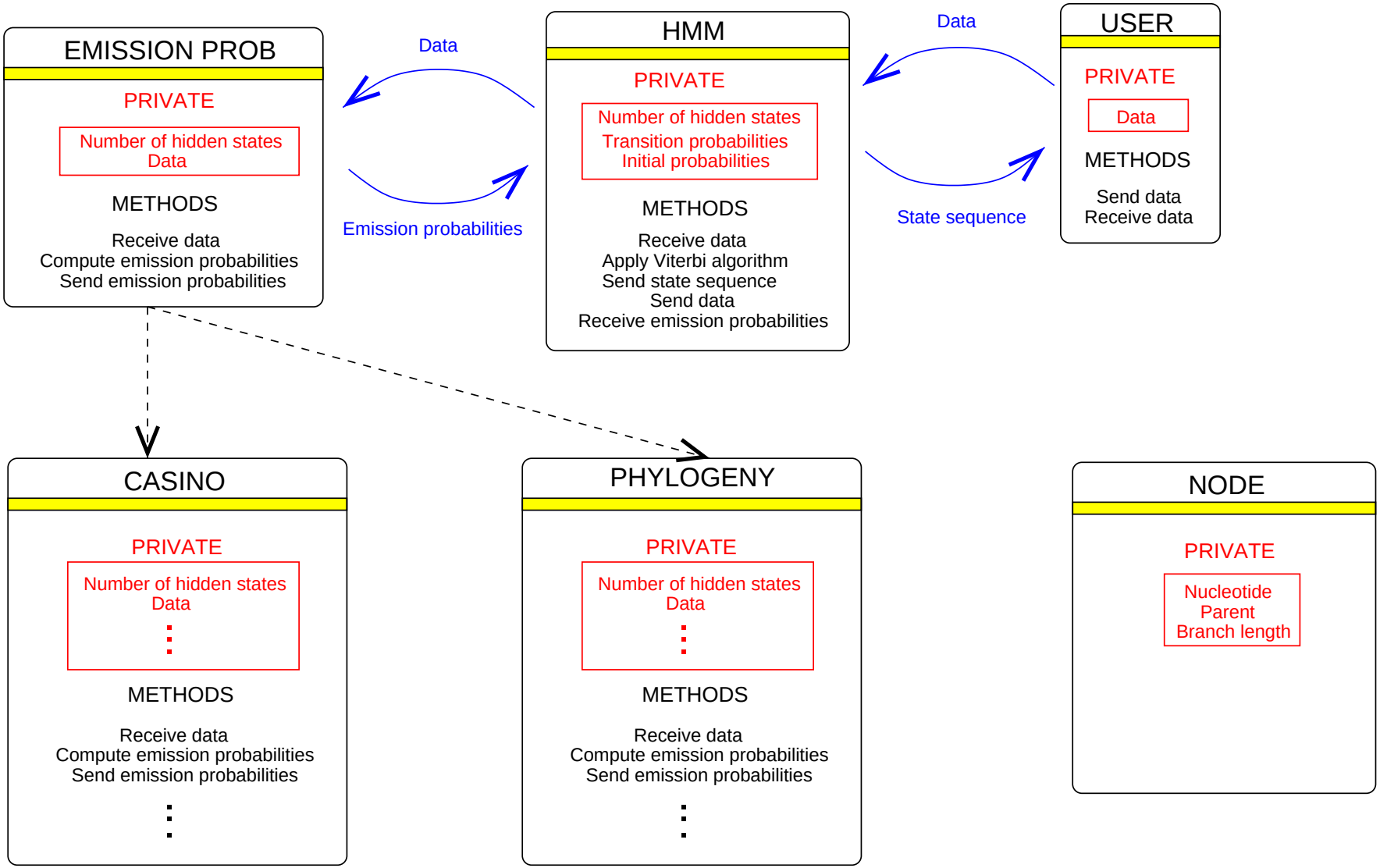


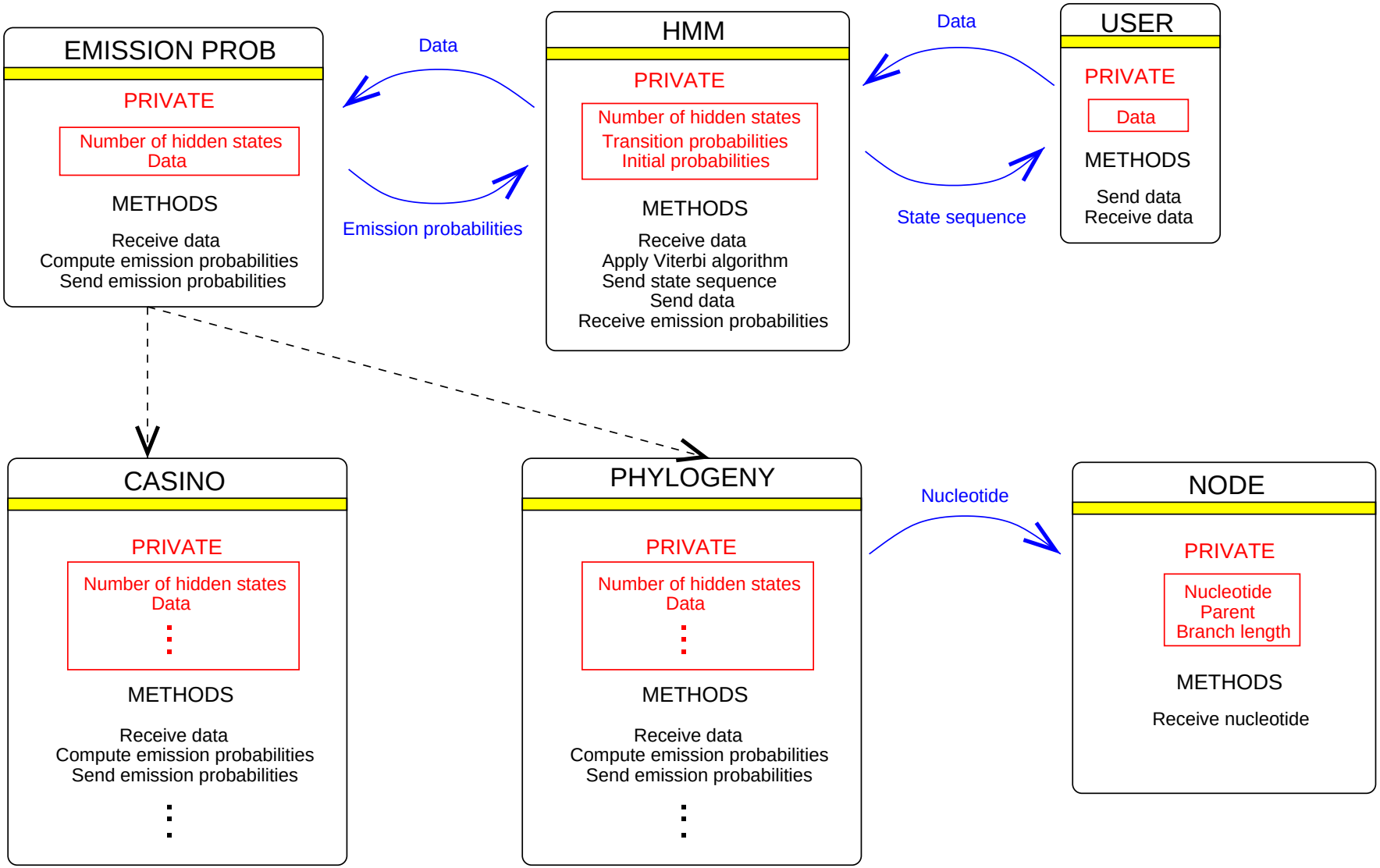
$$P(y_1, y_2, y_3, y_4, z_1, z_2 | \mathbf{w})$$

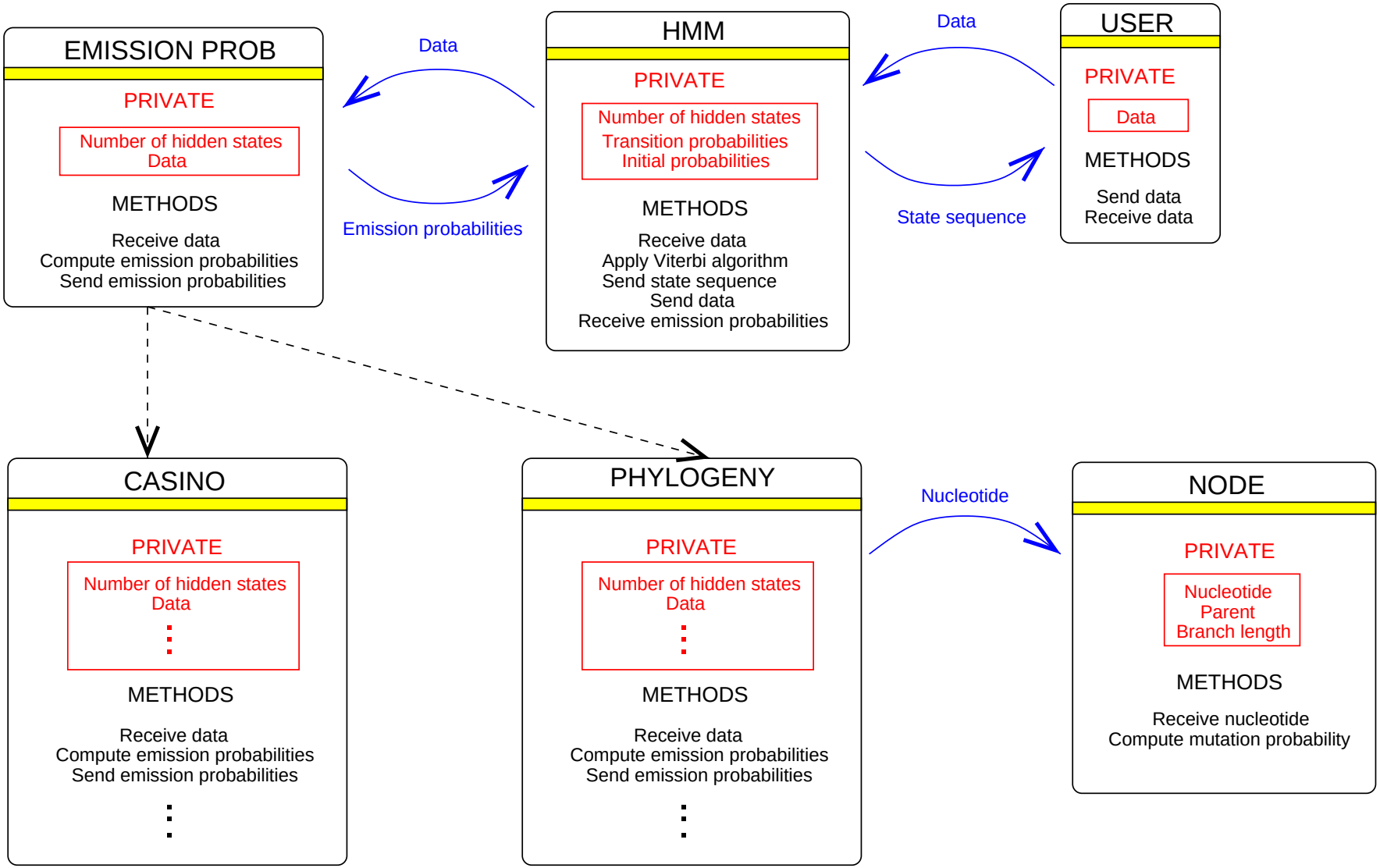
$$= P(y_1|z_1, w_1)P(y_2|z_1, w_2)P(z_2|z_1, w_5)P(y_3|z_2, w_3)P(y_4|z_2, w_4)P(z_1)$$

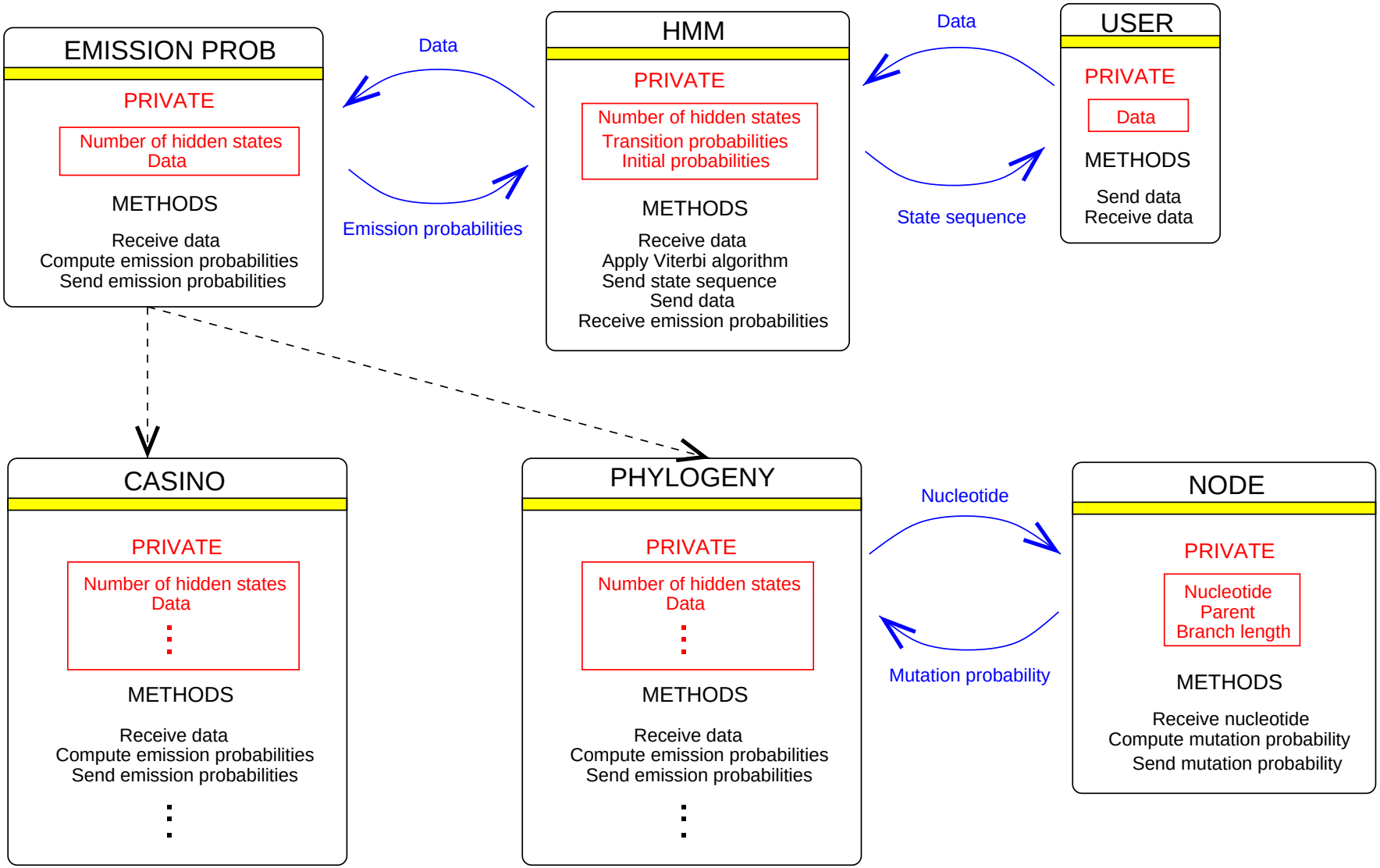

P(A|A,w)
A
P(G|A,w)
G
P(C|A,w)
P(T|A,w)
C
T

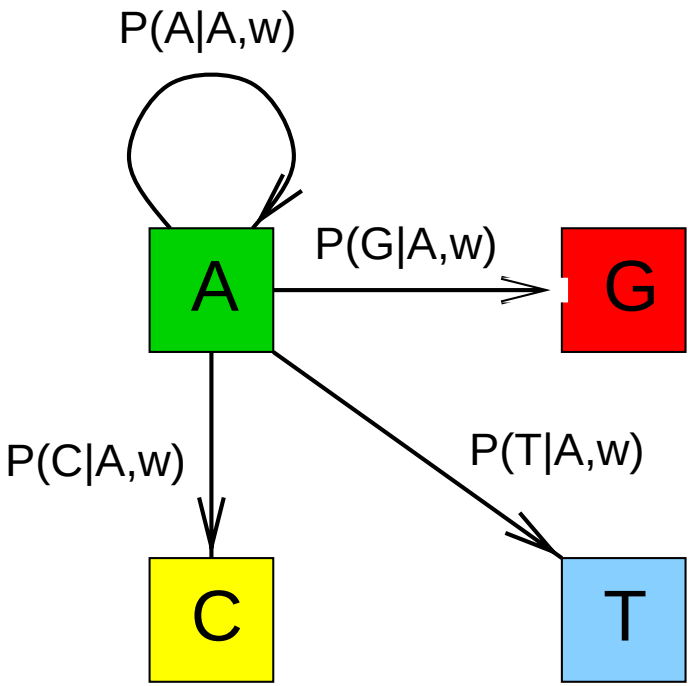

Transition−Transversion Ratio = 2
1
0.75
0.5
0.25
0
P(A|A,w)
P(G|A,w)
P(C|A,w)=P(T|A,w)
0
1
2
3
4
w

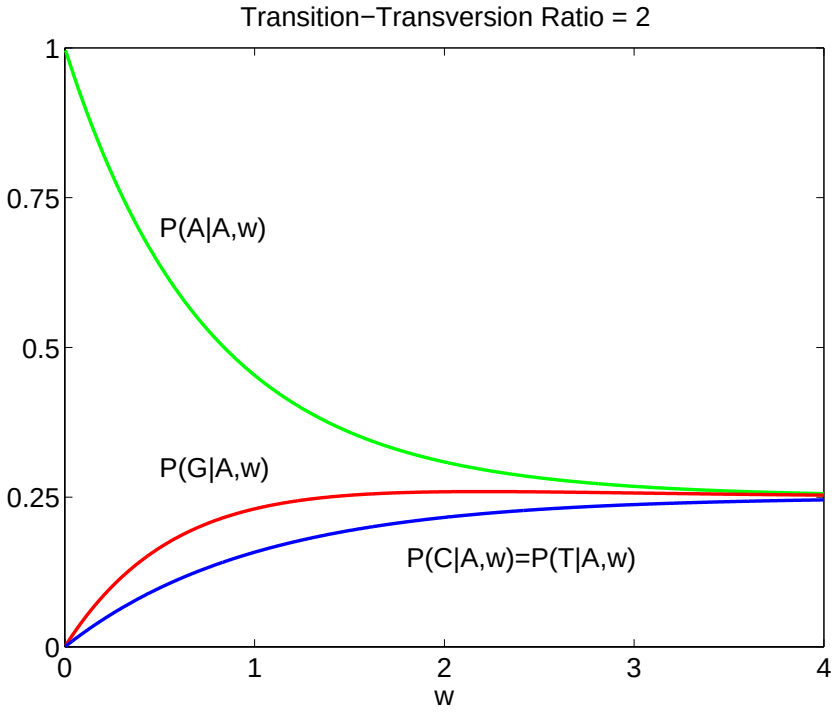

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
NODE
PRIVATE
Nucleotide
Parent
Branch length

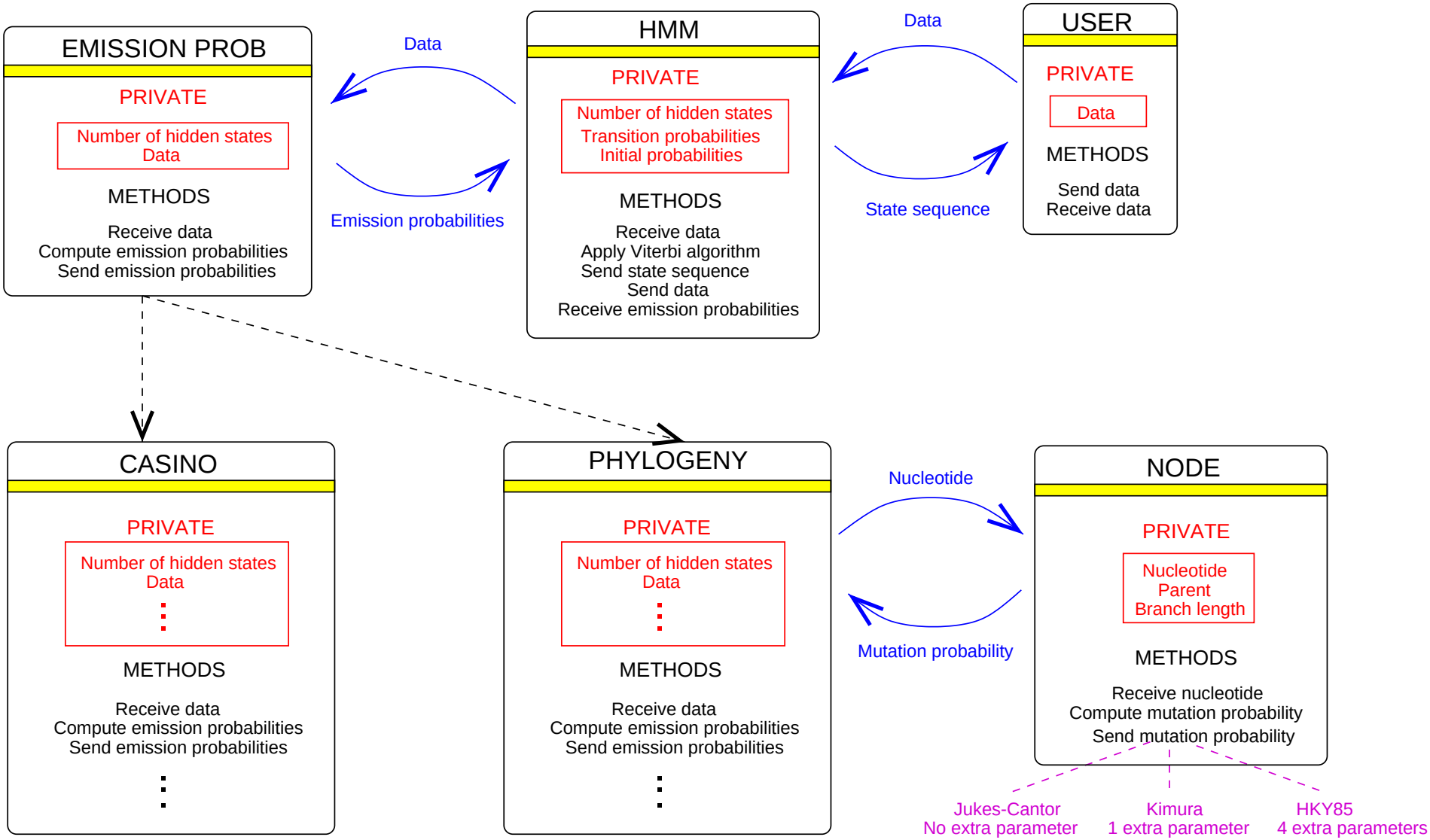

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide

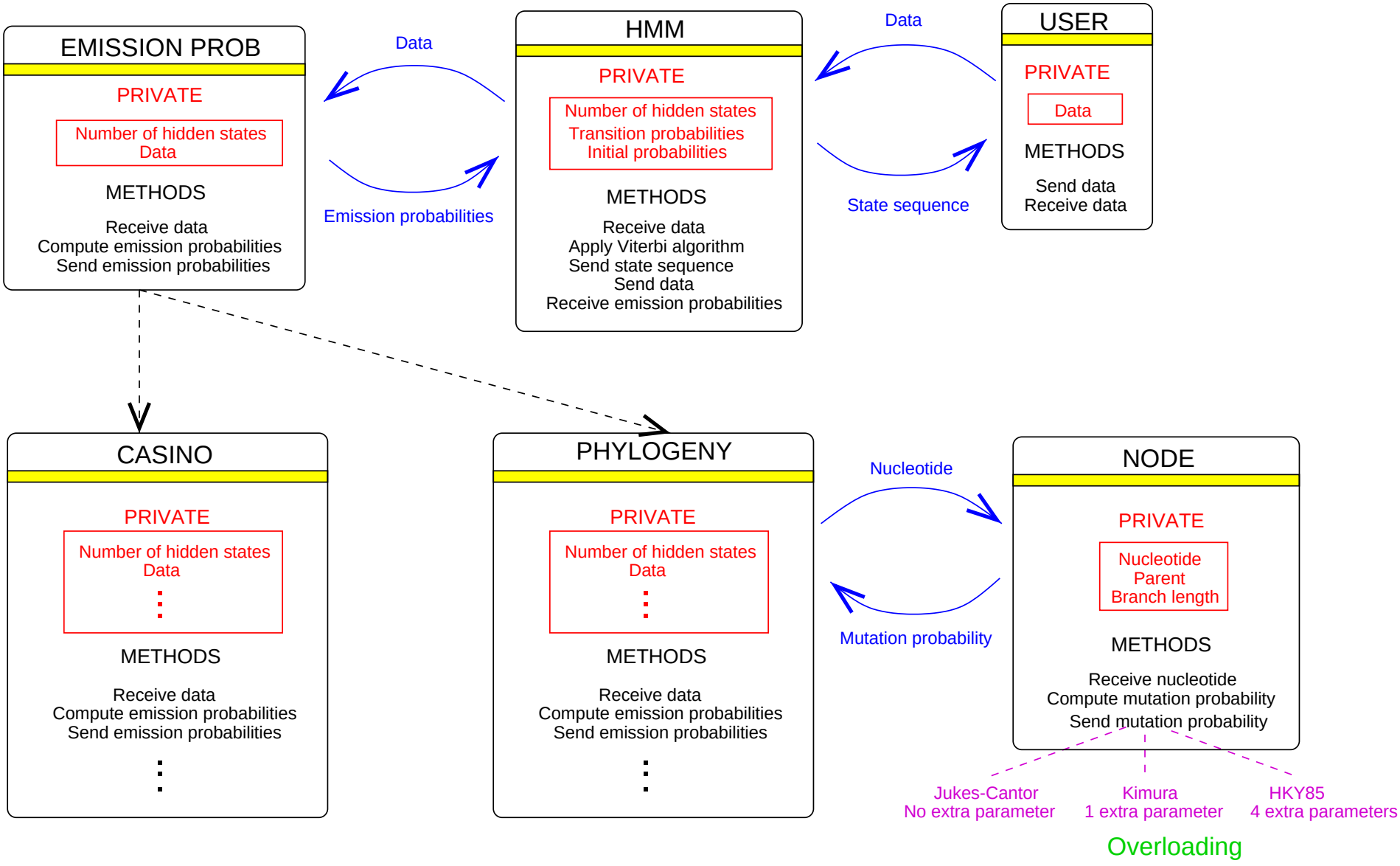

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability

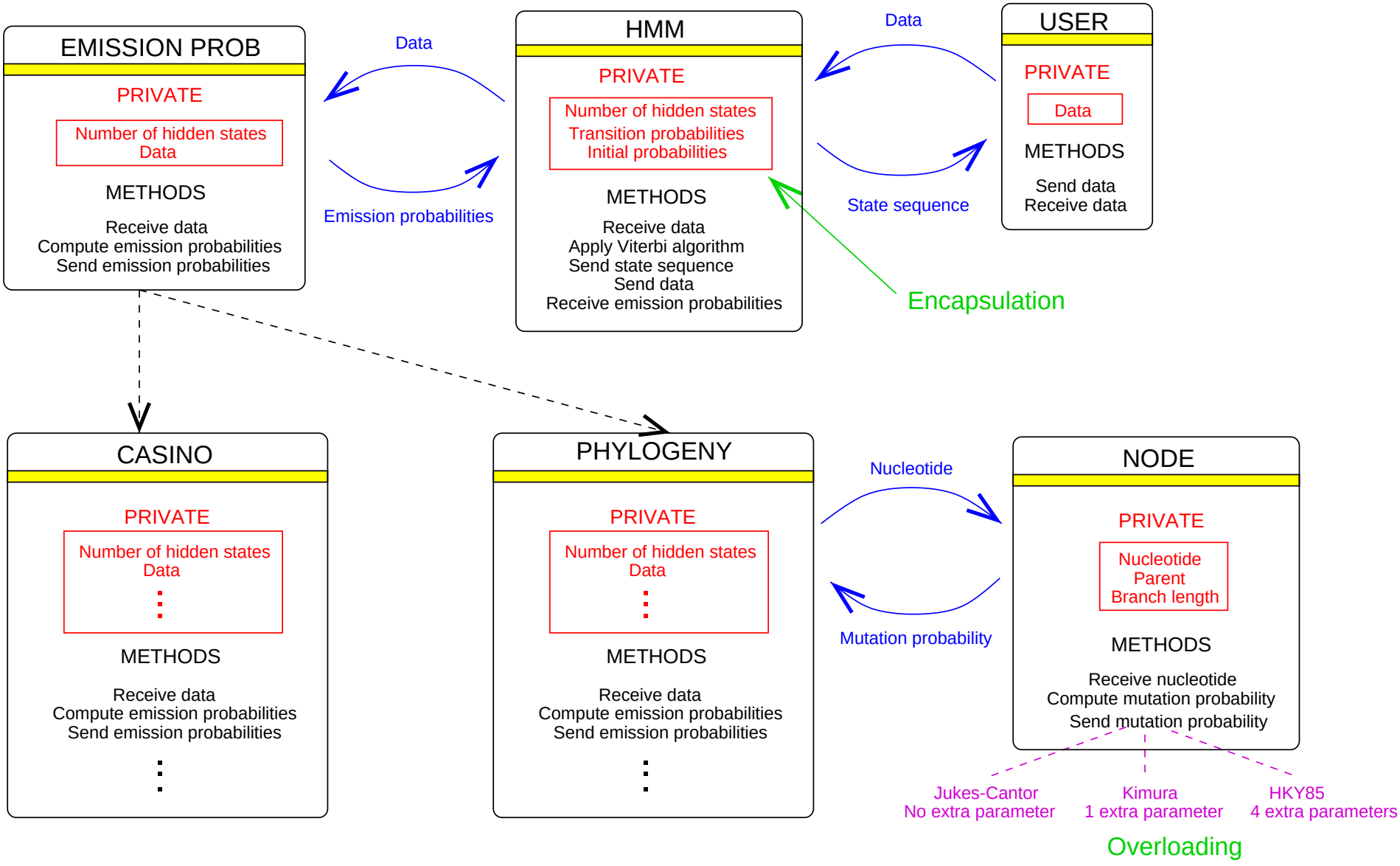

EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability

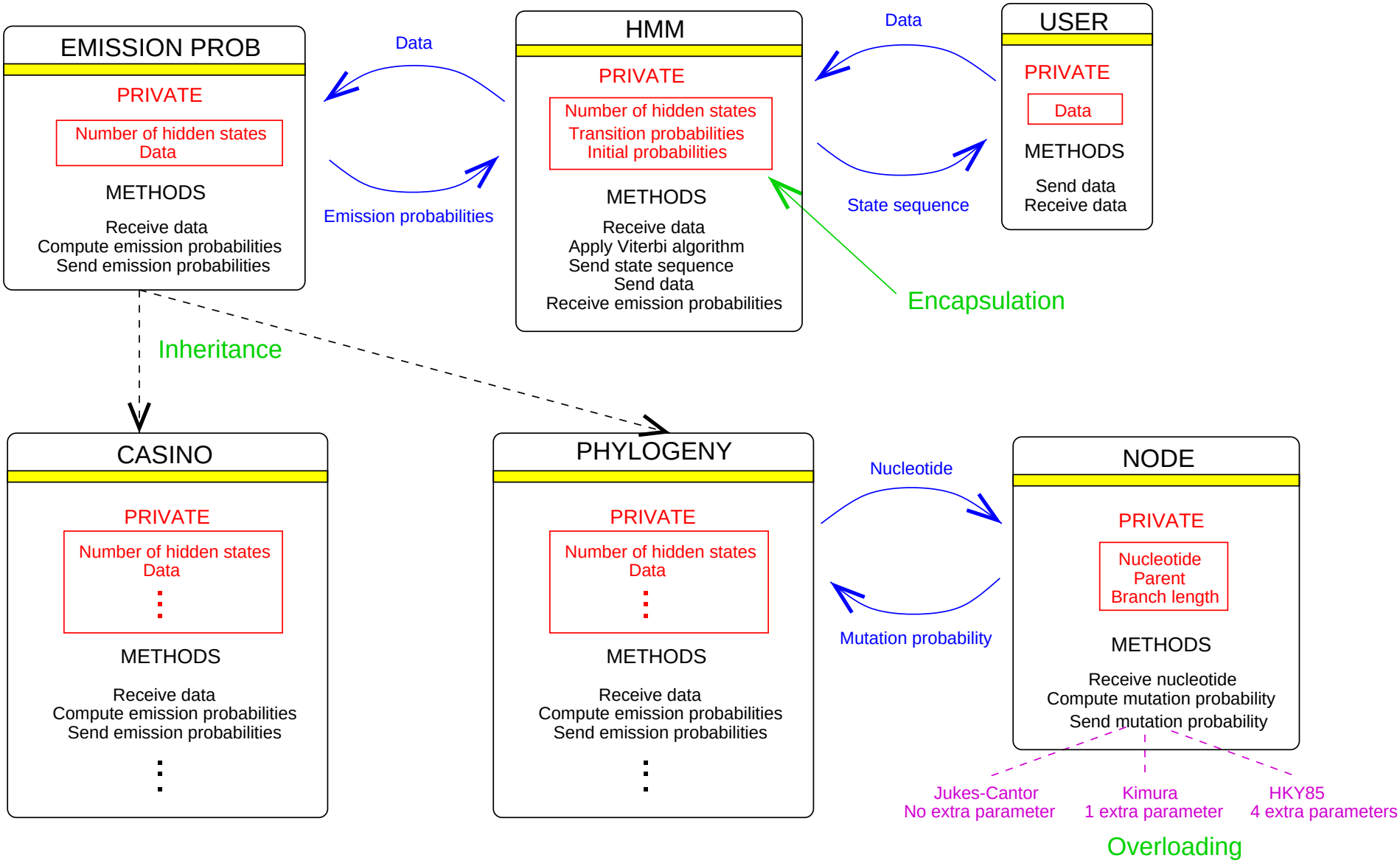

P(A|A,w)
A
P(G|A,w)
G
P(C|A,w)
P(T|A,w)
C
T

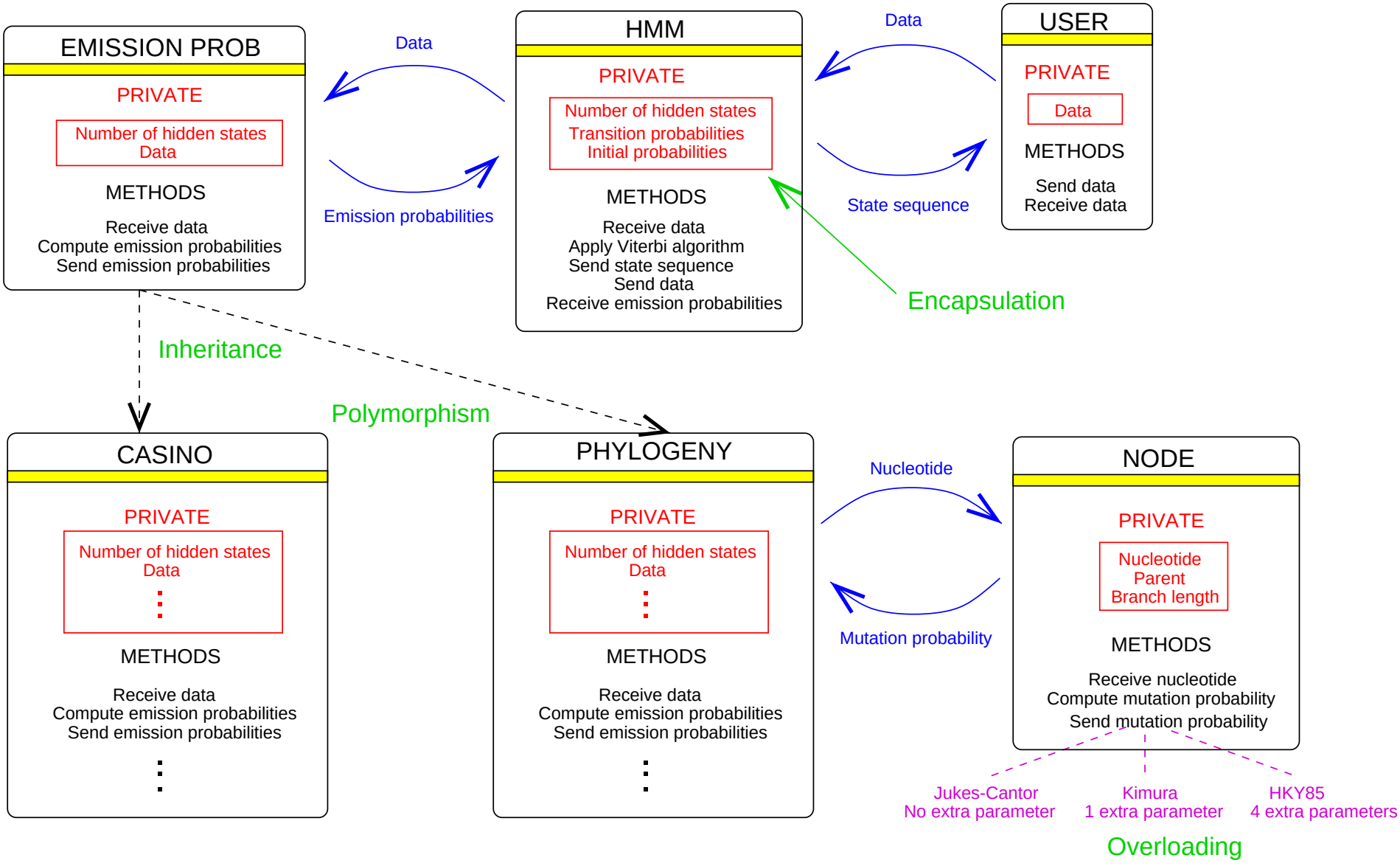

Transition−Transversion Ratio = 2
1
0.75
0.5
0.25
0
P(A|A,w)
P(G|A,w)
P(C|A,w)=P(T|A,w)
0
1
2
3
4
w


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters
Overloading


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
Encapsulation
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters
Overloading


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
Encapsulation
Inheritance
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters
Overloading


EMISSION PROB
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Data
Emission probabilities
HMM
PRIVATE
Number of hidden states
Transition probabilities
Initial probabilities
METHODS
Receive data
Apply Viterbi algorithm
Send state sequence
Send data
Receive emission probabilities
Data
State sequence
USER
PRIVATE
Data
METHODS
Send data
Receive data
Encapsulation
Inheritance
Polymorphism
CASINO
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
PHYLOGENY
PRIVATE
Number of hidden states
Data
METHODS
Receive data
Compute emission probabilities
Send emission probabilities
Nucleotide
Mutation probability
NODE
PRIVATE
Nucleotide
Parent
Branch length
METHODS
Receive nucleotide
Compute mutation probability
Send mutation probability
Jukes-Cantor
No extra parameter
Kimura
1 extra parameter
HKY85
4 extra parameters
Overloading