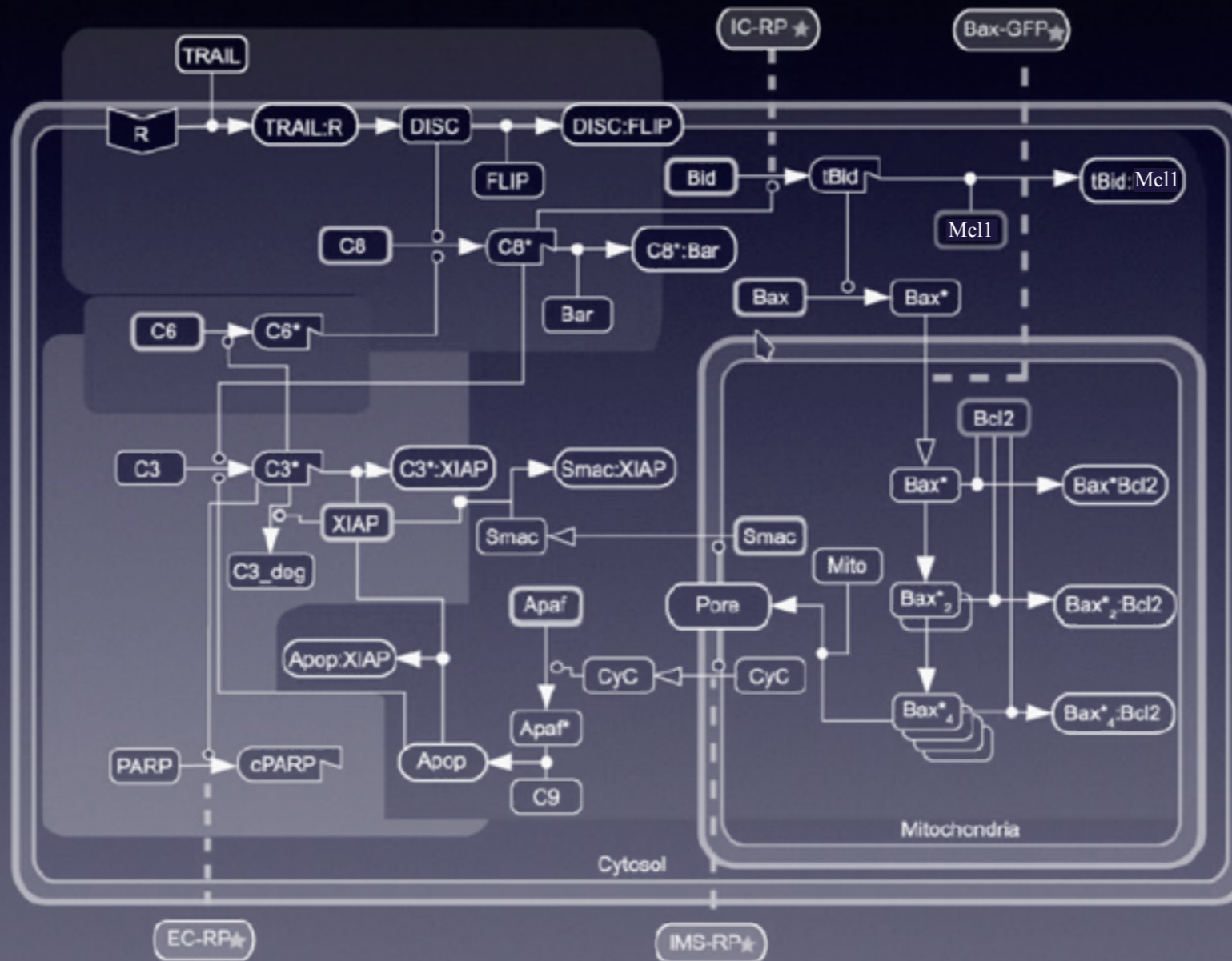


Inference in DBNs with non-disjoint clusters

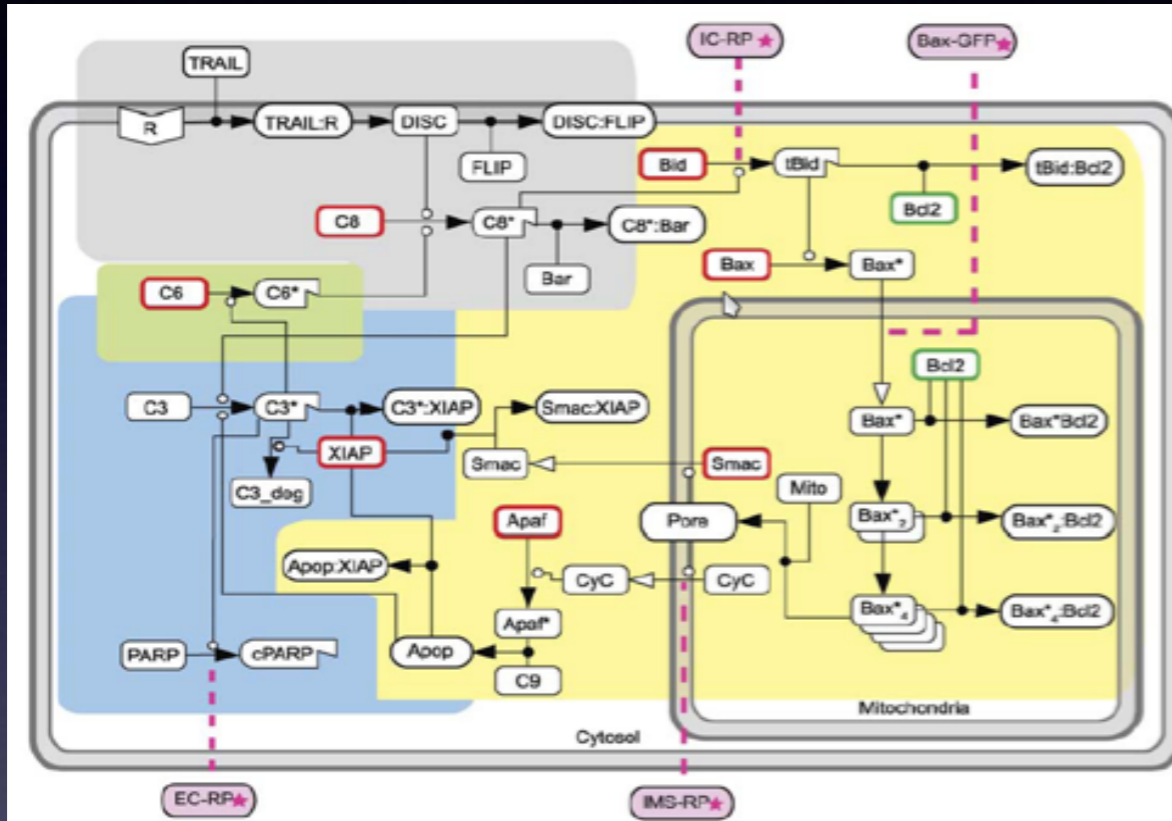
Matthieu Pichéné

Introduction

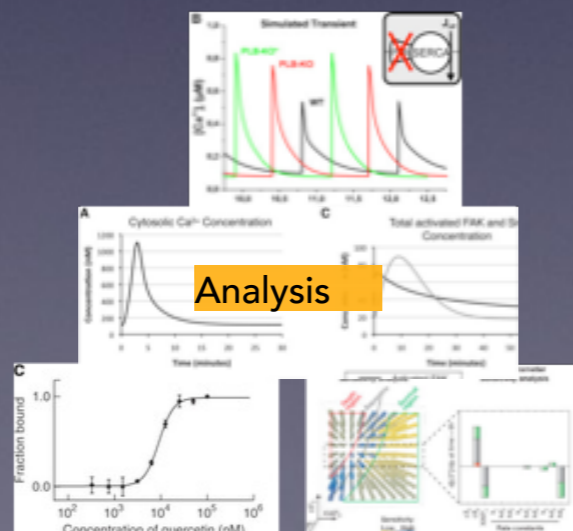
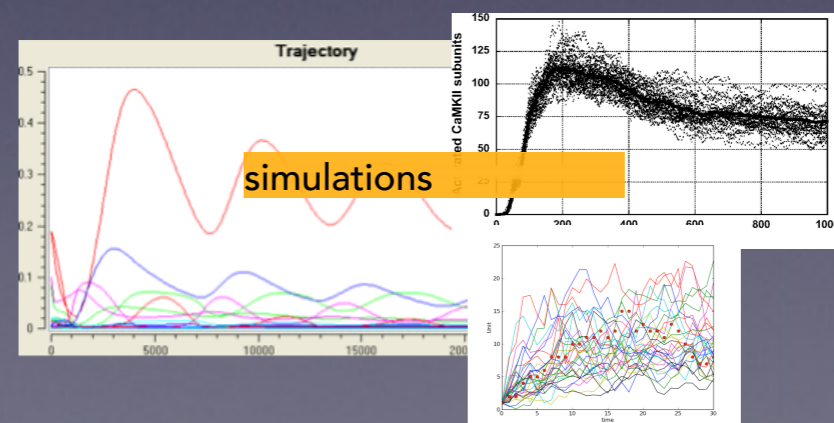
Apoptosis pathway



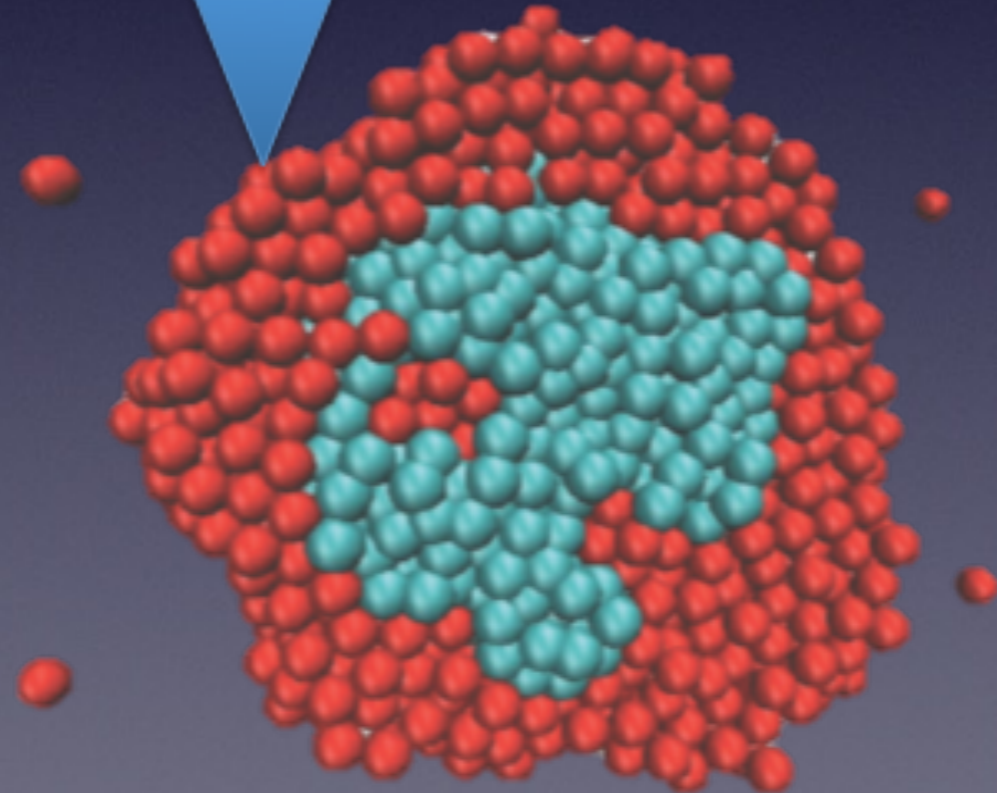
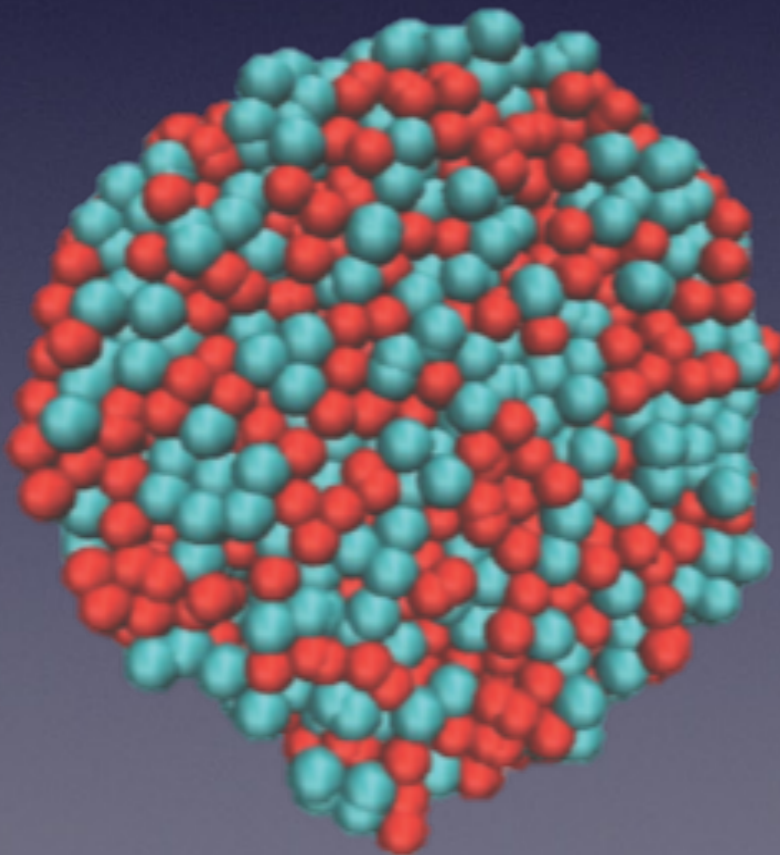
Method




$$\begin{aligned} \frac{dc_1}{dt} &= -k_1 \cdot c_1 \cdot c_2 + k_2 \cdot c_3 \\ \frac{dc_2}{dt} &= -k_1 \cdot c_1 \cdot c_2 + k_2 \cdot c_3 + k_{17} \cdot c_{18} + k_{11} \cdot c_{11} \\ \frac{dc_3}{dt} &= k_1 \cdot c_1 \cdot c_2 - k_2 \cdot c_3 - k_3 \cdot c_3 \cdot c_4 + k_4 \cdot c_5 \\ \frac{dc_4}{dt} &= -k_3 \cdot c_3 \cdot c_4 + k_4 \cdot c_5 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_5}{dt} &= \text{MATHEMATICAL FORMALISM} \cdot k_6 \cdot c_7 \\ \frac{dc_6}{dt} &= -k_5 \cdot c_5 \cdot c_6 + k_6 \cdot c_7 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_7}{dt} &= k_5 \cdot c_5 \cdot c_6 - k_6 \cdot c_7 - k_7 \cdot c_7 \cdot c_8 + k_8 \cdot c_9 \\ \frac{dc_8}{dt} &= -k_7 \cdot c_7 \cdot c_8 + k_8 \cdot c_9 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_9}{dt} &= k_7 \cdot c_7 \cdot c_8 - k_8 \cdot c_9 - k_9 \cdot c_9 \cdot c_{10} + k_{10} \cdot c_{11} \\ &\quad - k_{15} \cdot c_9 \cdot c_{17} + k_{16} \cdot c_{18} \end{aligned}$$



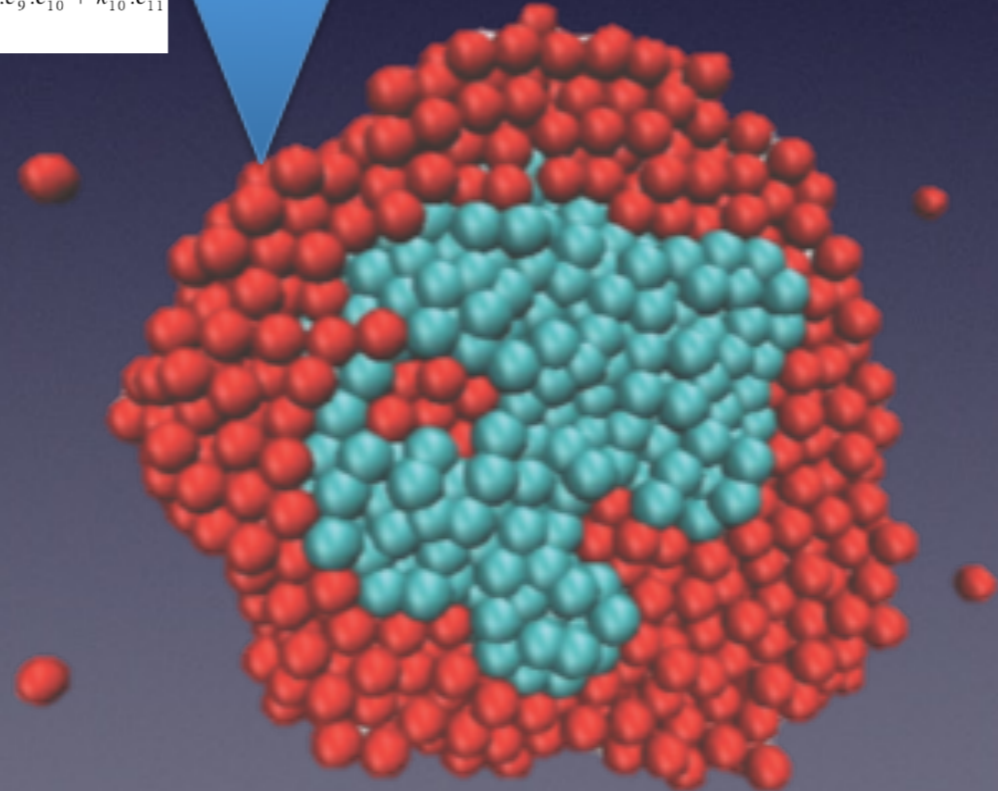
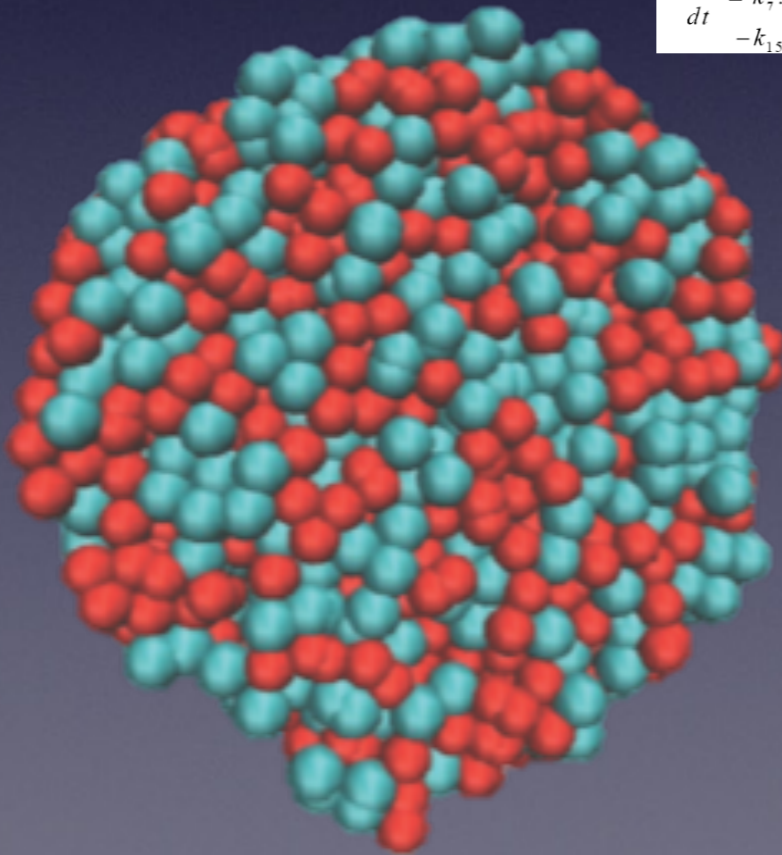
Method



Method

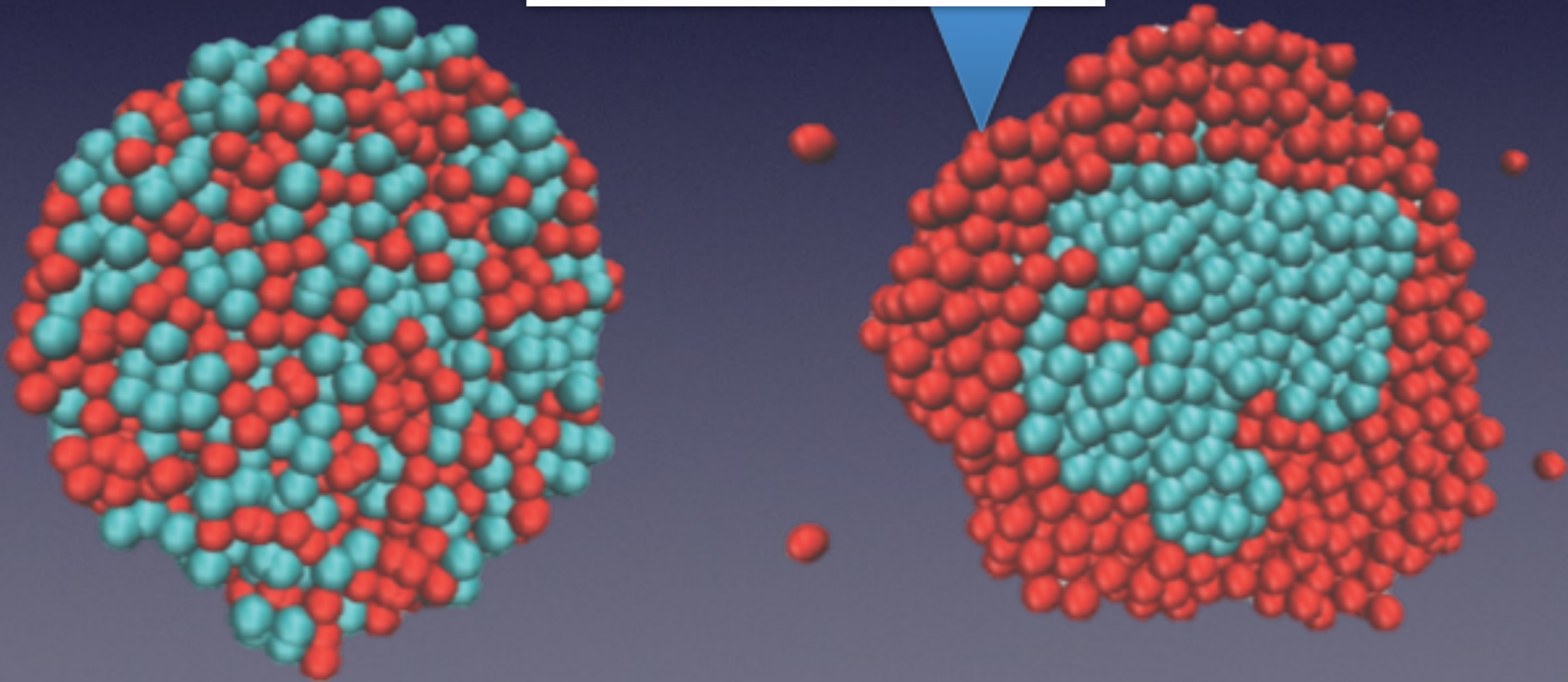


A diagram of a multi-compartment system with nodes and arrows, representing a reaction network. The diagram shows a central compartment with several smaller compartments attached to its top and bottom. Arrows indicate the flow of particles between these compartments.

$$\begin{aligned}\frac{dc_1}{dt} &= -k_1 \cdot c_1 \cdot c_2 + k_2 \cdot c_3 \\ \frac{dc_2}{dt} &= -k_1 \cdot c_1 \cdot c_2 + k_2 \cdot c_3 + k_{17} \cdot c_{18} + k_{11} \cdot c_{11} \\ \frac{dc_3}{dt} &= k_1 \cdot c_1 \cdot c_2 - k_2 \cdot c_3 - k_3 \cdot c_3 \cdot c_4 + k_4 \cdot c_5 \\ \frac{dc_4}{dt} &= -k_3 \cdot c_3 \cdot c_4 + k_4 \cdot c_5 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_5}{dt} &= k_3 \cdot c_3 \cdot c_4 - k_4 \cdot c_5 - k_5 \cdot c_5 \cdot c_6 + k_6 \cdot c_7 \\ \frac{dc_6}{dt} &= -k_5 \cdot c_5 \cdot c_6 + k_6 \cdot c_7 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_7}{dt} &= k_5 \cdot c_5 \cdot c_6 - k_6 \cdot c_7 - k_7 \cdot c_7 \cdot c_8 + k_8 \cdot c_9 \\ \frac{dc_8}{dt} &= -k_7 \cdot c_7 \cdot c_8 + k_8 \cdot c_9 + k_{11} \cdot c_{11} + k_{20} \cdot c_{21} \\ \frac{dc_9}{dt} &= k_7 \cdot c_7 \cdot c_8 - k_8 \cdot c_9 - k_9 \cdot c_9 \cdot c_{10} + k_{10} \cdot c_{11} \\ &\quad - k_{15} \cdot c_9 \cdot c_{17} + k_{16} \cdot c_{18}\end{aligned}$$


Method

(Approximate)
abstraction
of the low level
biochemical model



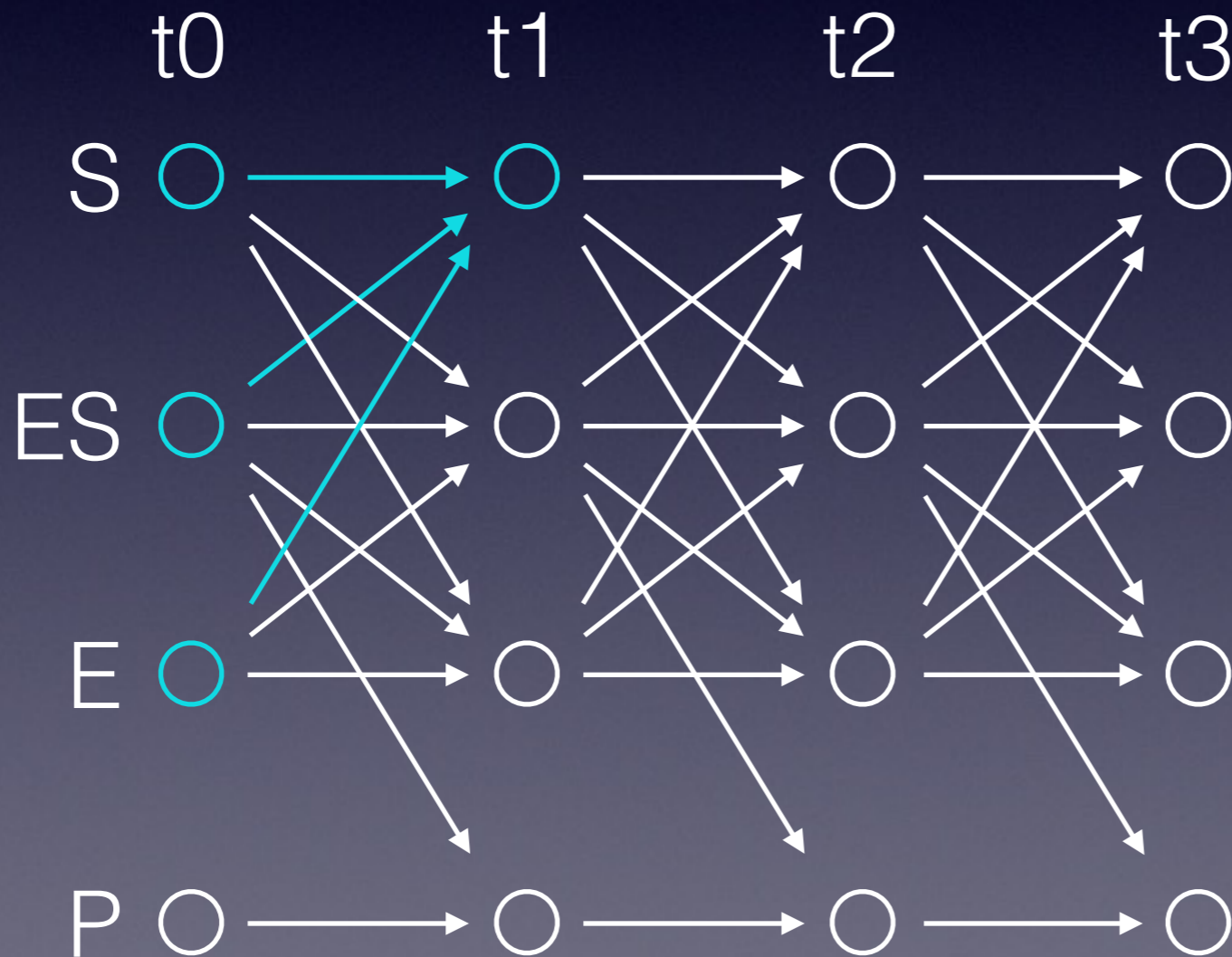
DBNs



Every specie at time point t is a random variable over a discrete number of values.

Number of configurations at each time point: Values Species

DBNs



+



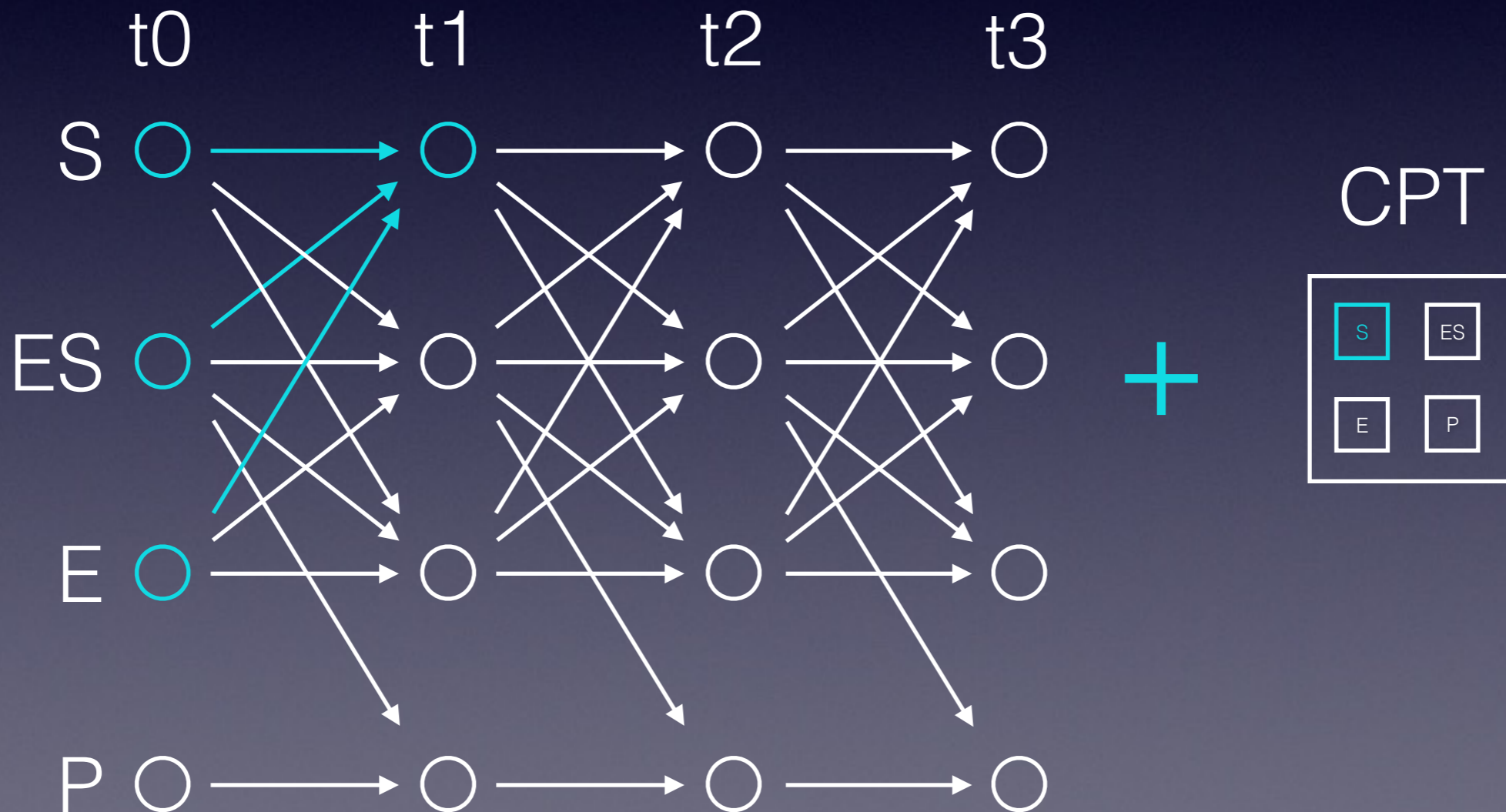
CPT



S					ES					
S	S	E	ES	Pr	ES	S	E	ES	P	Pr
1	1	1	1	0.1	1					
1	2	1	2	0.2	1					
2	2	3	3	0.1	2					
...					...					

E					P			
E	S	E	ES	Pr	P	ES	P	Pr
1					1			
1					1			
2					2			
...					...			

DBNs

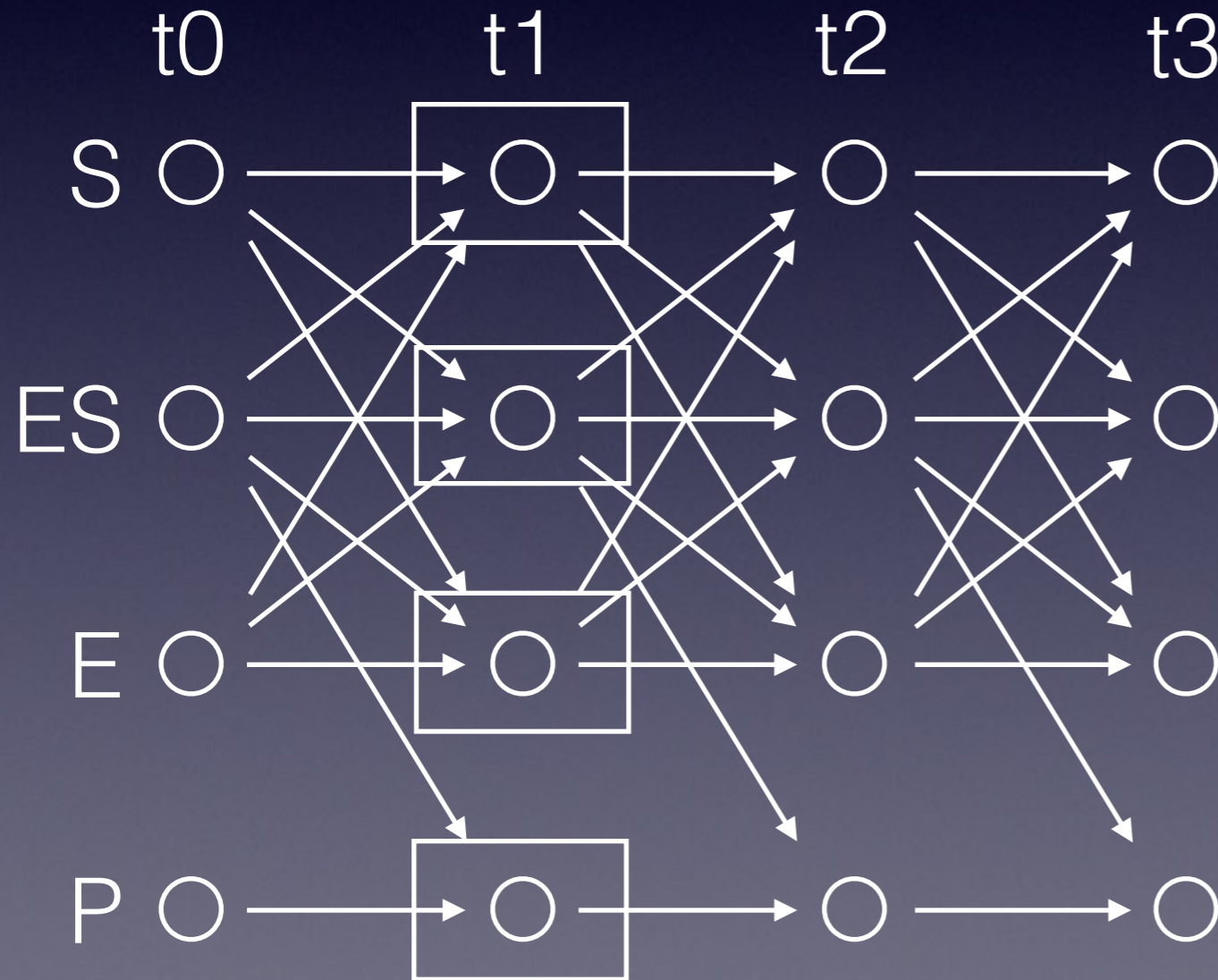


Complexity of exact inference: at least Values **Species**

DBNs

- We need an approximation. Express configurations as product of probabilities
- Simplest idea : Consider all species independent (Factored Frontier)

Factored Frontier



$$P^{t_2}(P=h) = f(P^{t_1}(P), P^{t_1}(ES), CPT)$$

complexity of FF inference:
Species x Values^{NbPar+1}

Low accuracy

Hypothesis : Independent

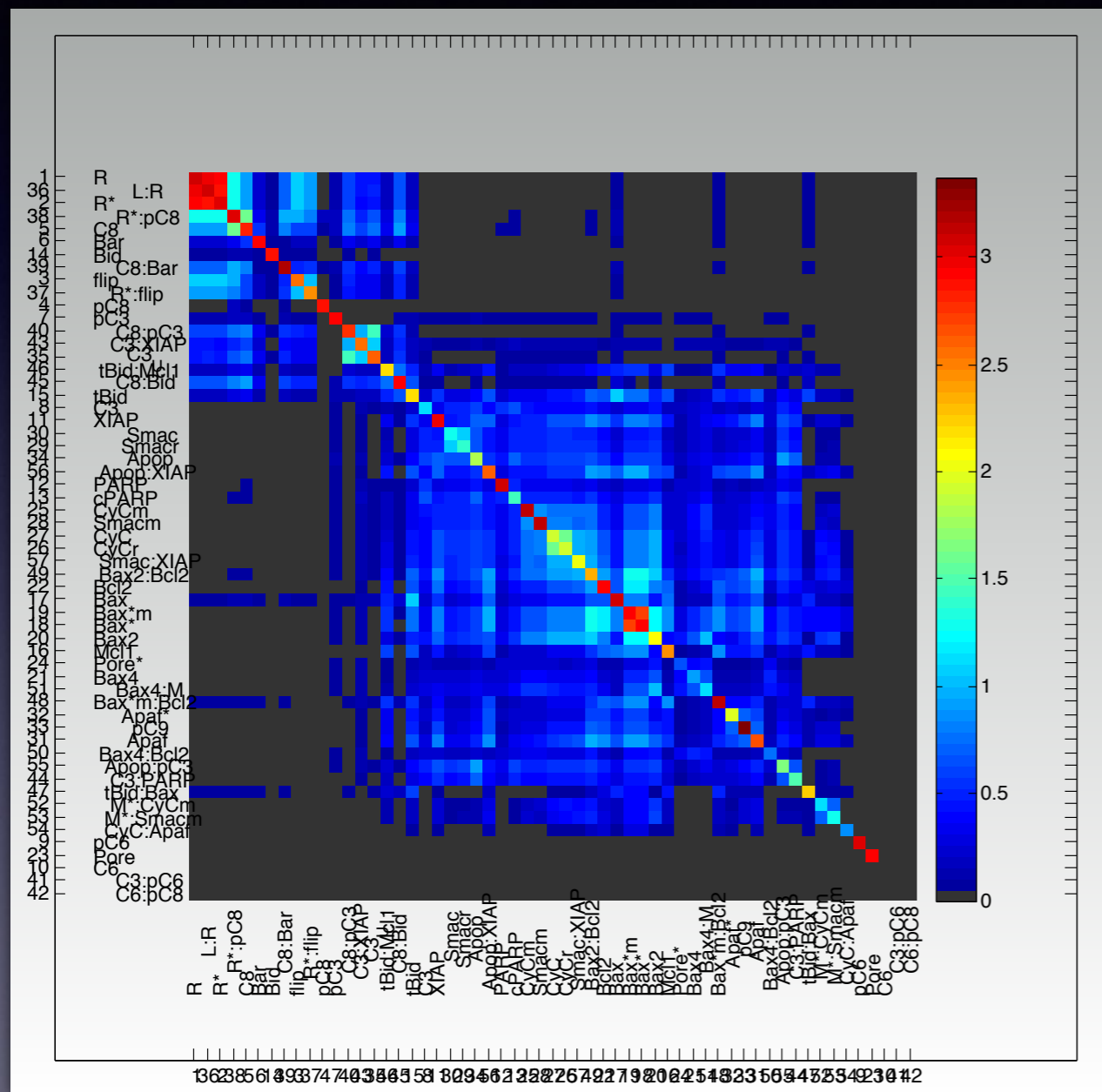
Clustered Factored Frontier

- Use of clusters containing the species that have the most mutual information
- Clusters may vary over time
- All sets of states for species in a clusters are calculated (that limits the length of clusters)

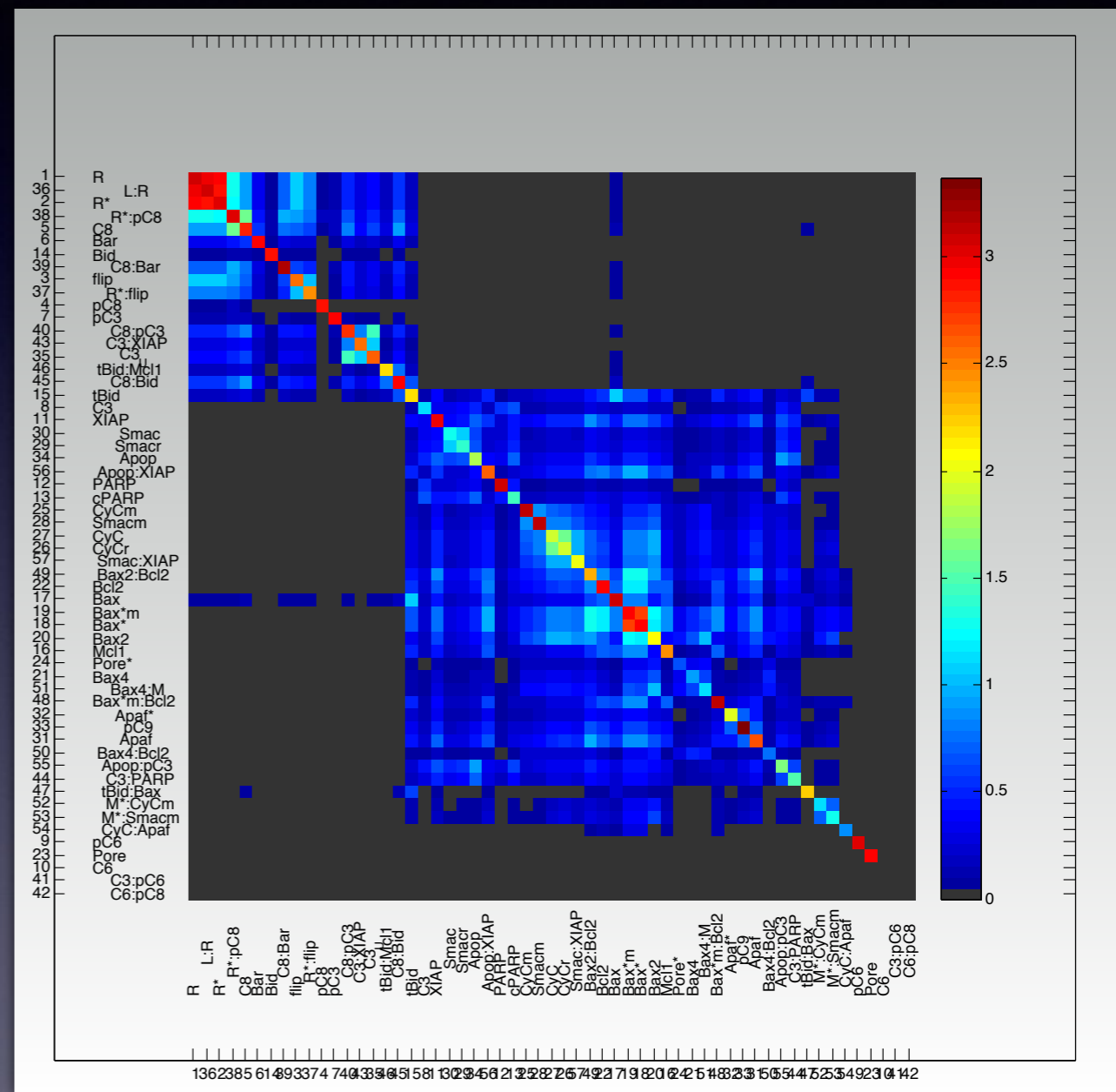
Clustered Factored Frontier

- Use information theory (Eric) to obtain the important relations
- We (Eric) chose the tree to minimize distance $(P(X_1 = v_1, \dots, X_n = v_n); \prod_{\ell=1}^k \frac{P((X_i)_{i \in C_\ell} = (v_i)_{i \in C_\ell})}{P((X_i)_{i \in C_\ell \cup \bigcup_{i=1}^{\ell-1} C_i} = (v_i)_{i \in C_\ell \cup \bigcup_{i=1}^{\ell-1} C_i})})$
- Tree implies cluster of size 2

Species correlations (Eric)



Mutual information on the whole graph



Mutual Information on the Tree Approximation

Clustered Factored Frontier

we assume that relations not in tree are irrelevant

Hypothesis :

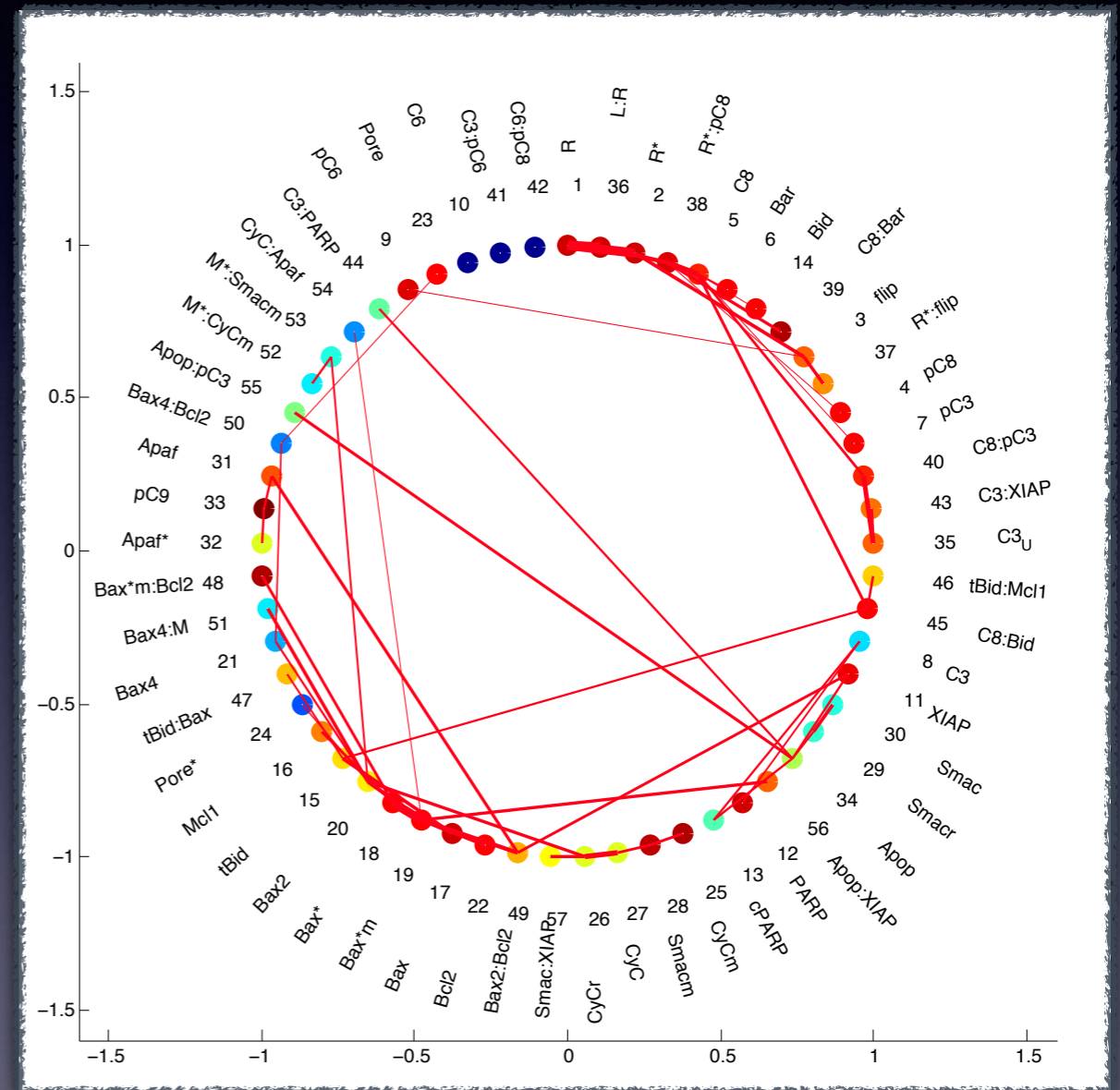
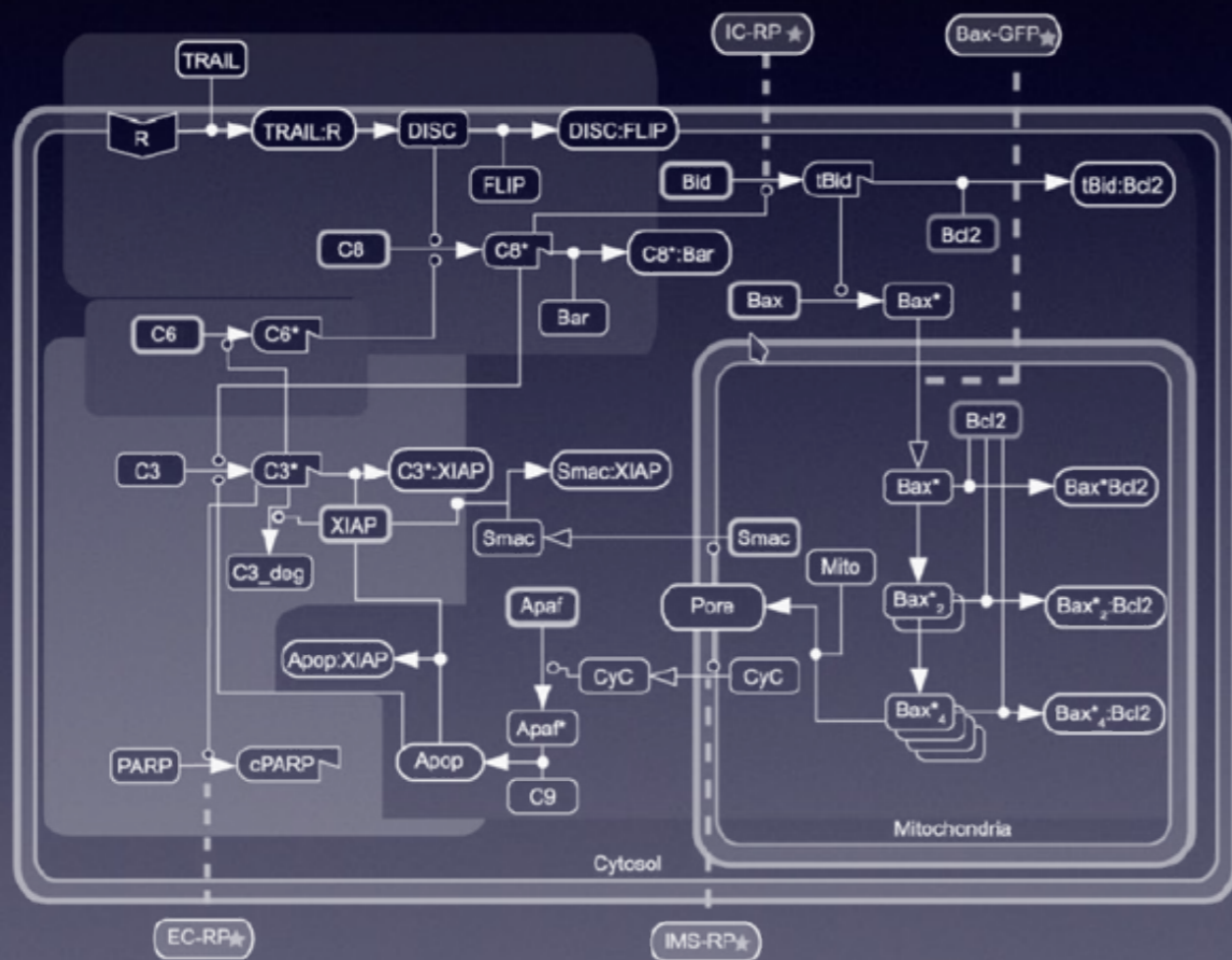


$$\Pr(S^t=h, ES^t=l, E^t=m, P^t=h) =$$

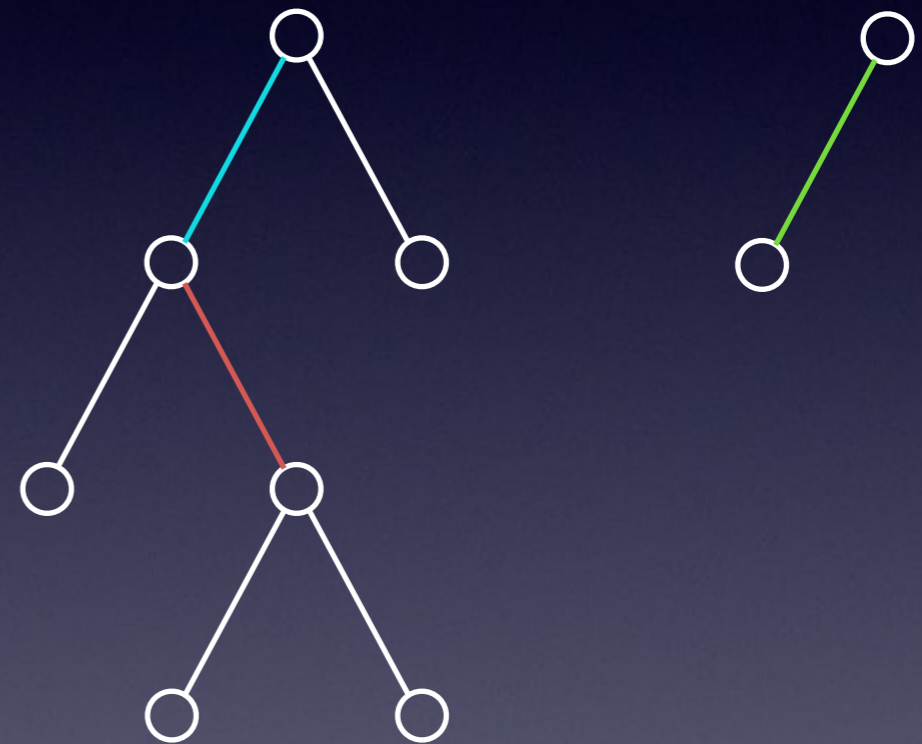
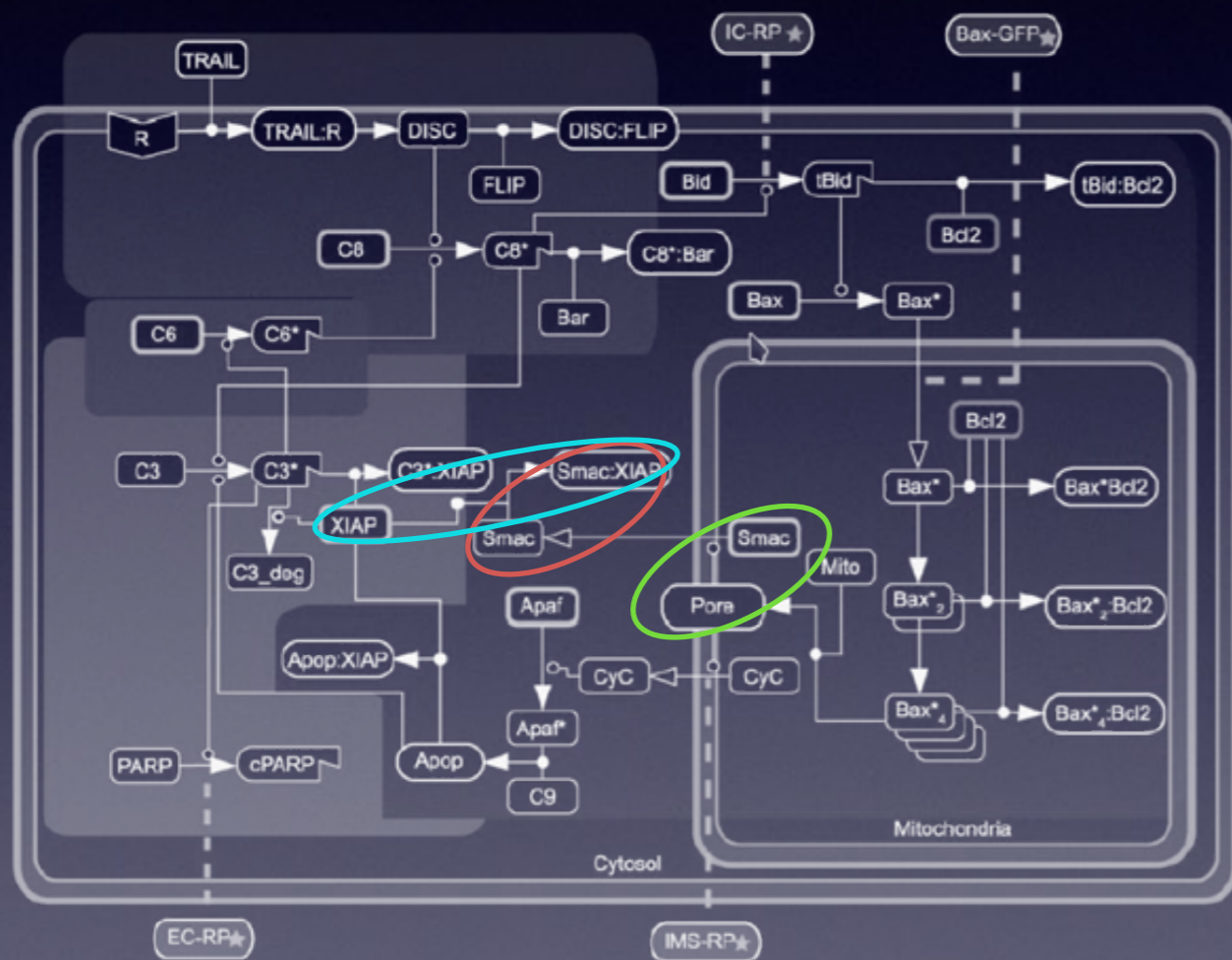
$$\Pr(S^t=h, ES^t=l) \Pr(ES^t=l, E^t=m) \Pr(ES^t=l, P^t=h)$$

$$\Pr^2(ES^t=l)$$

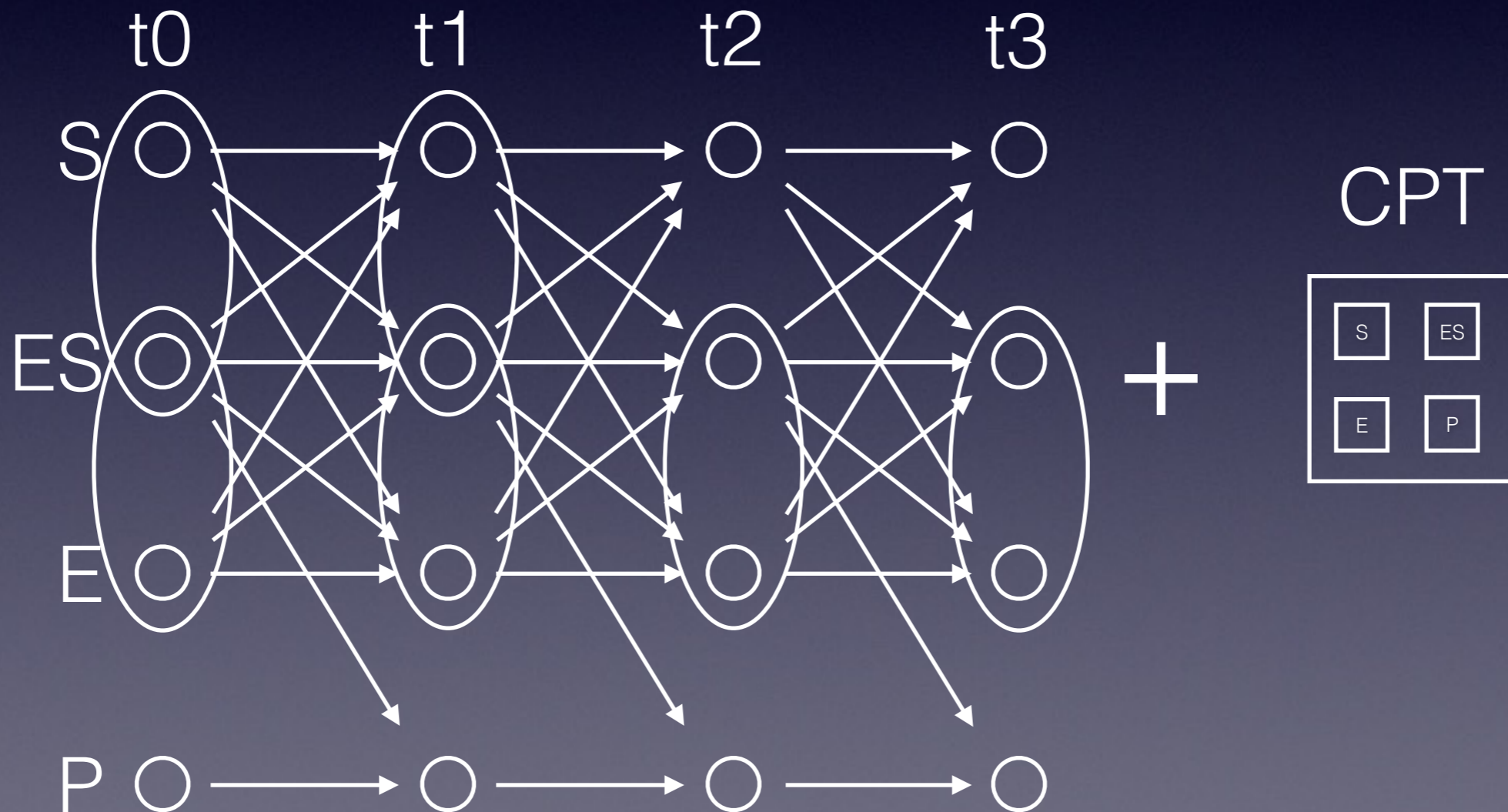
Apoptosis pathway



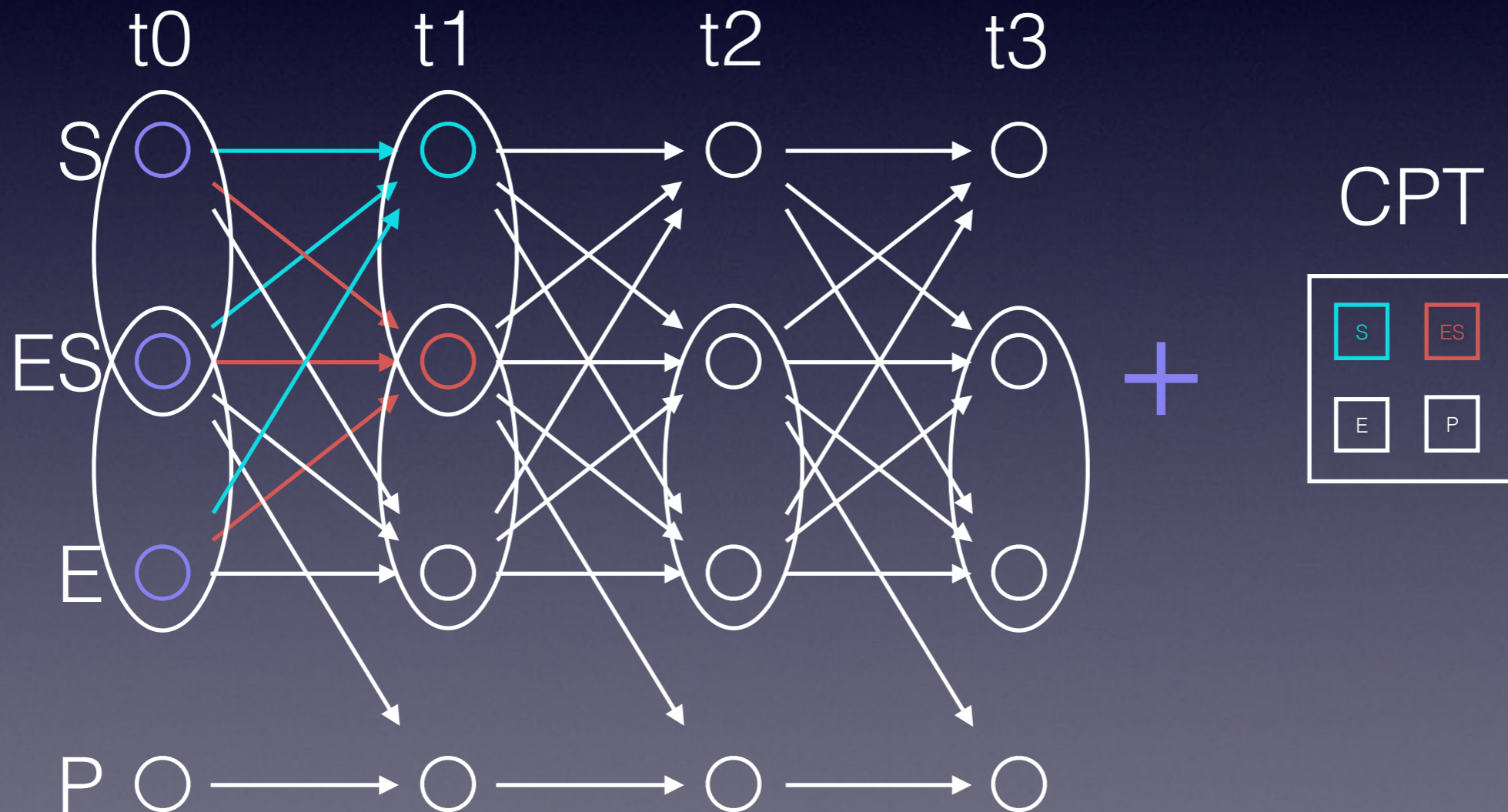
Apoptosis pathway



Clustered Factored Frontier



Clustered Factored Frontier



$$P^{t_1}(s', es') = \sum_{s, es, e} (P^{t_0}(s, es, e) \text{CPT}(s, es, e, s') \text{CPT}(s, es, e, es'))$$

How our algorithm work

Hypothesis :

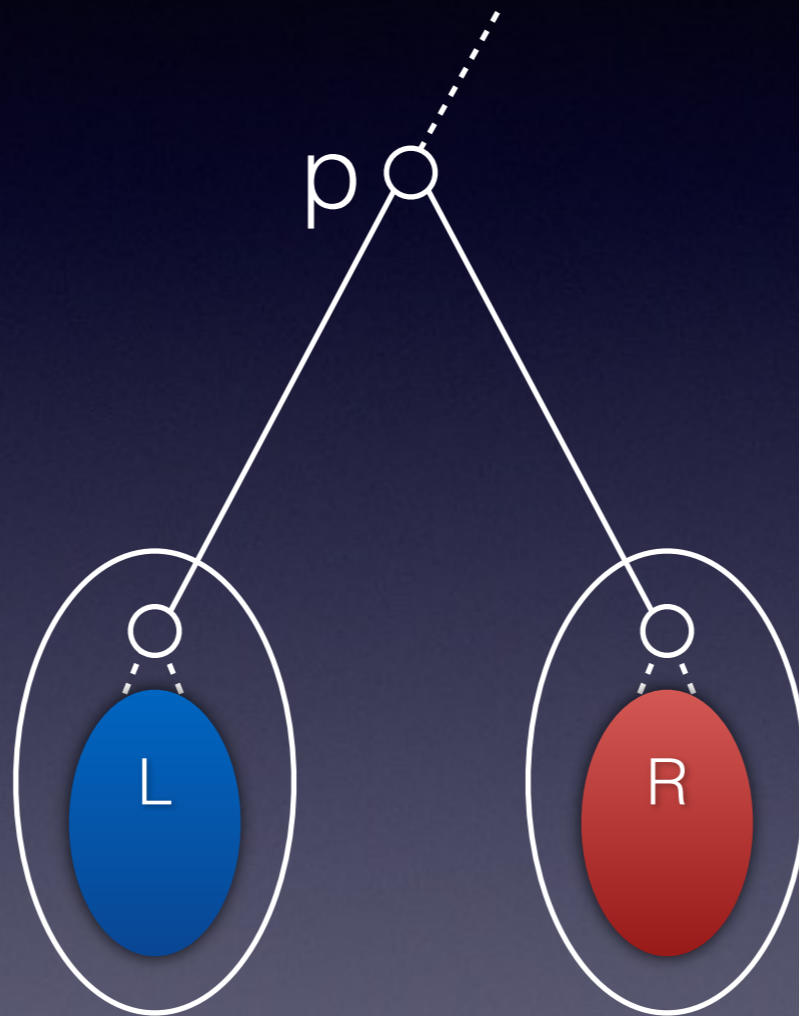


$$\Pr(S^t=h, ES^t=l, E^t=m, P^t=h) =$$

$$\Pr(S^t=h, ES^t=l) \Pr(ES^t=l, E^t=m) \Pr(ES^t=l, P^t=h)$$

$$\Pr^2(ES^t=l)$$

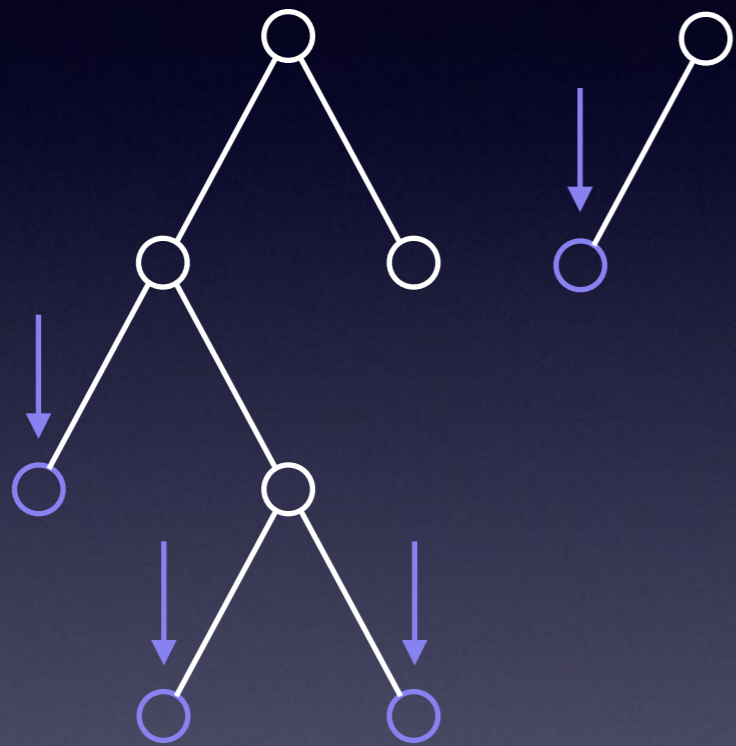
How to compute $P(\text{parents}(\text{Cluster}))$



Proposition :

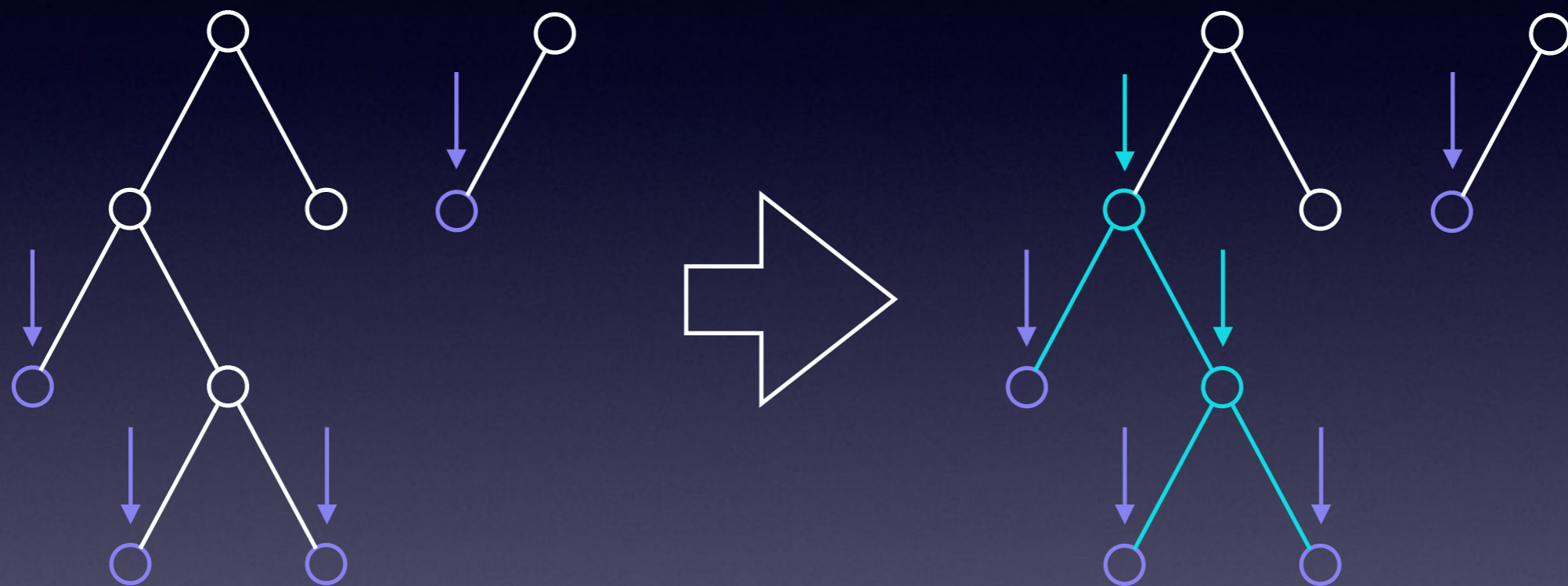
$$P(X_p = v_p, X_L = V_L, X_R = V_R) = \frac{P(X_p = v_p, X_L = V_L) \times P(X_p = v_p, X_R = V_R)}{P(X_p = v_p)}$$

How to compute $P(\text{parents}(\text{Cluster}))$

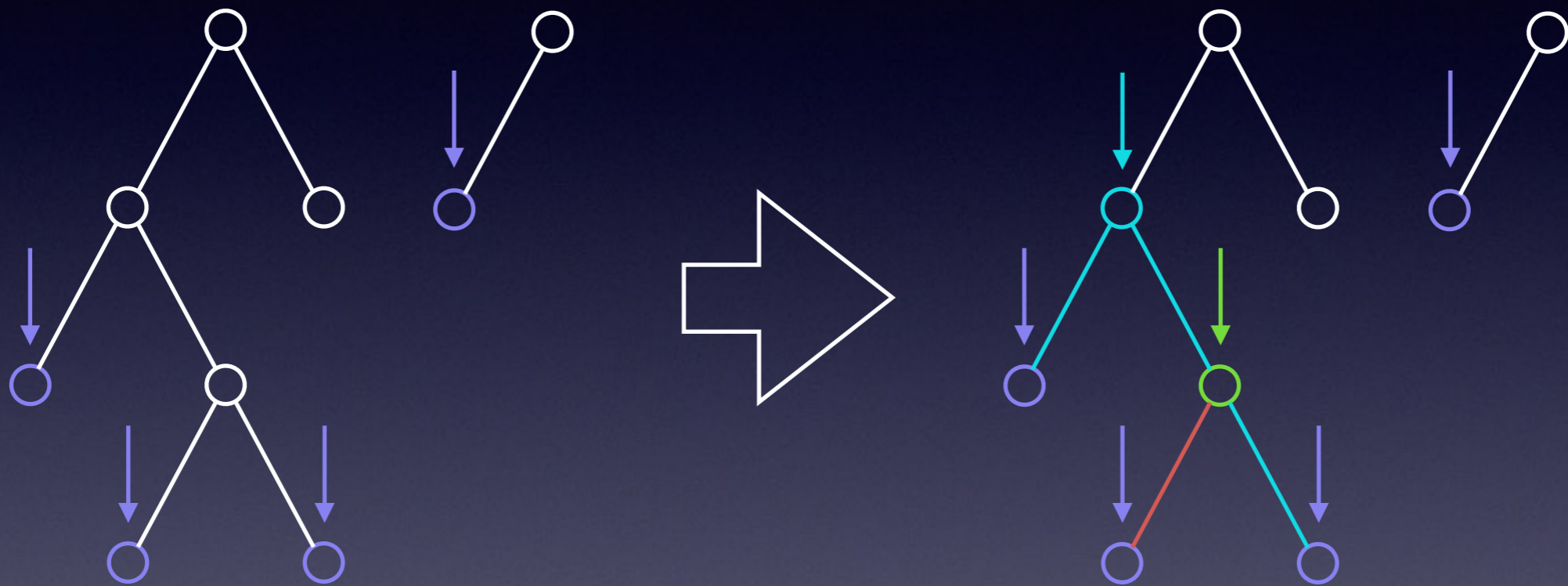


Parent_Cluster = set of nodes necessary to use the CPTs.

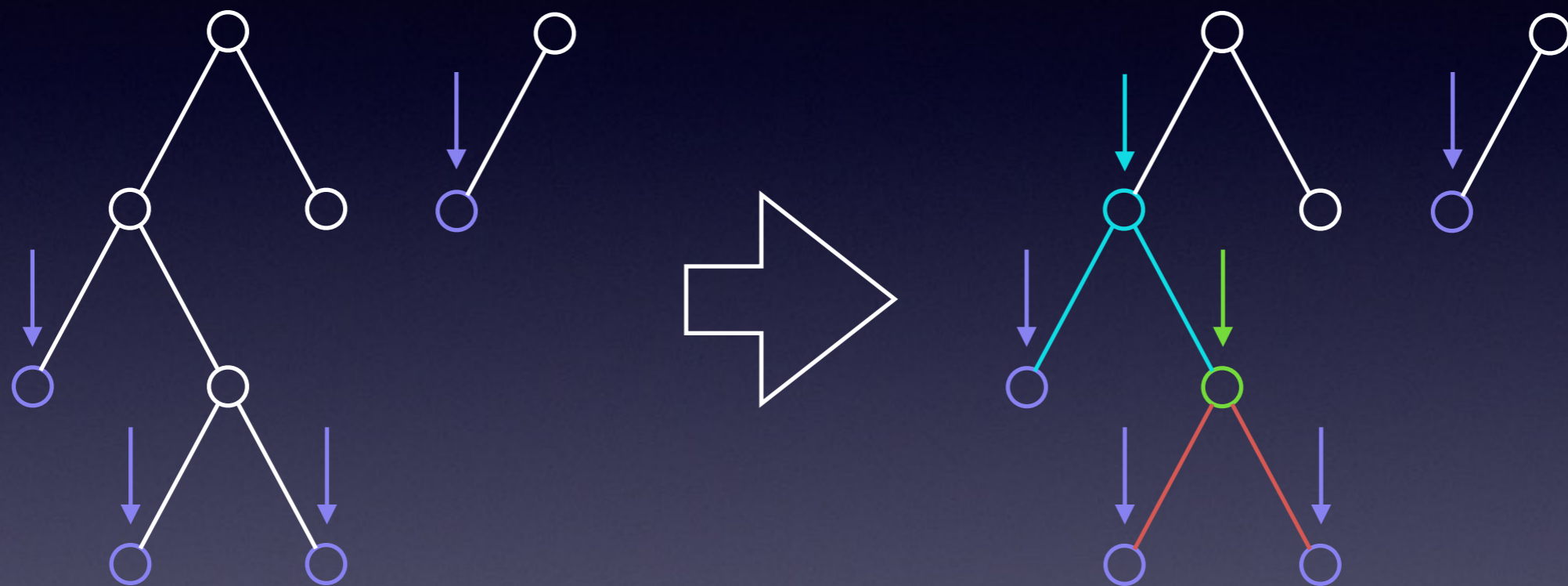
How to compute $P(\text{parents}(\text{Cluster}))$



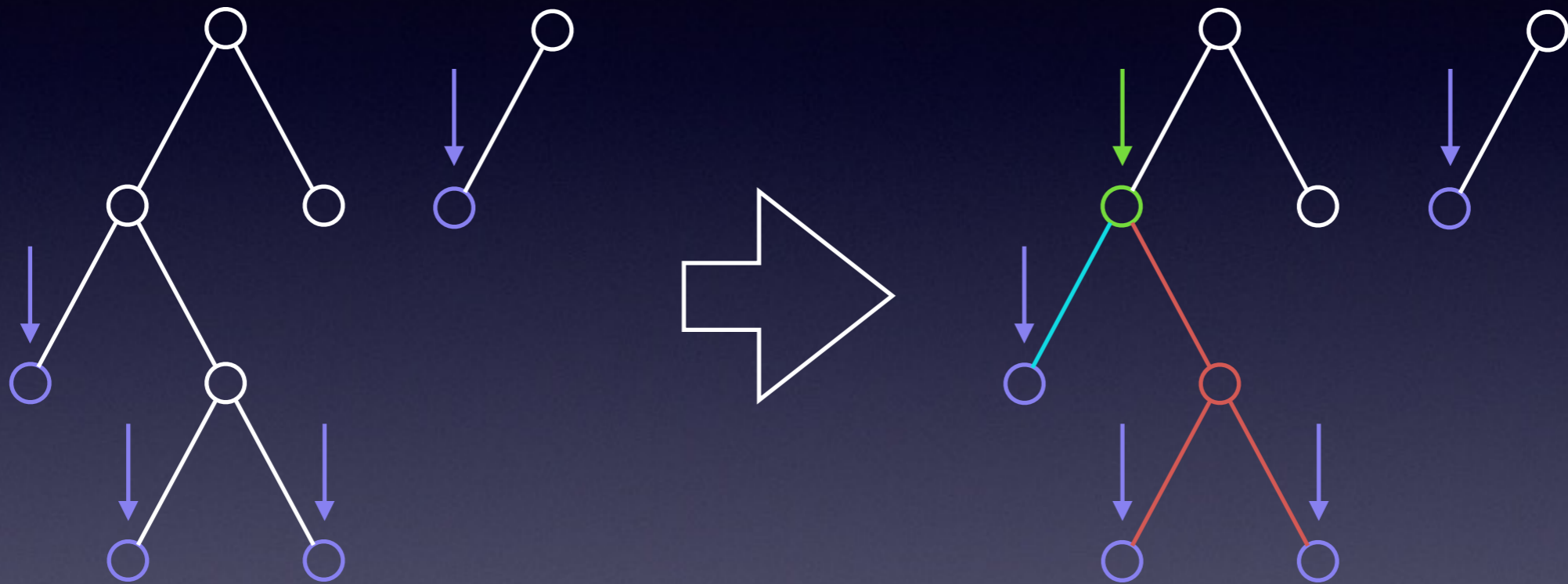
How to compute $P(\text{parents}(\text{Cluster}))$



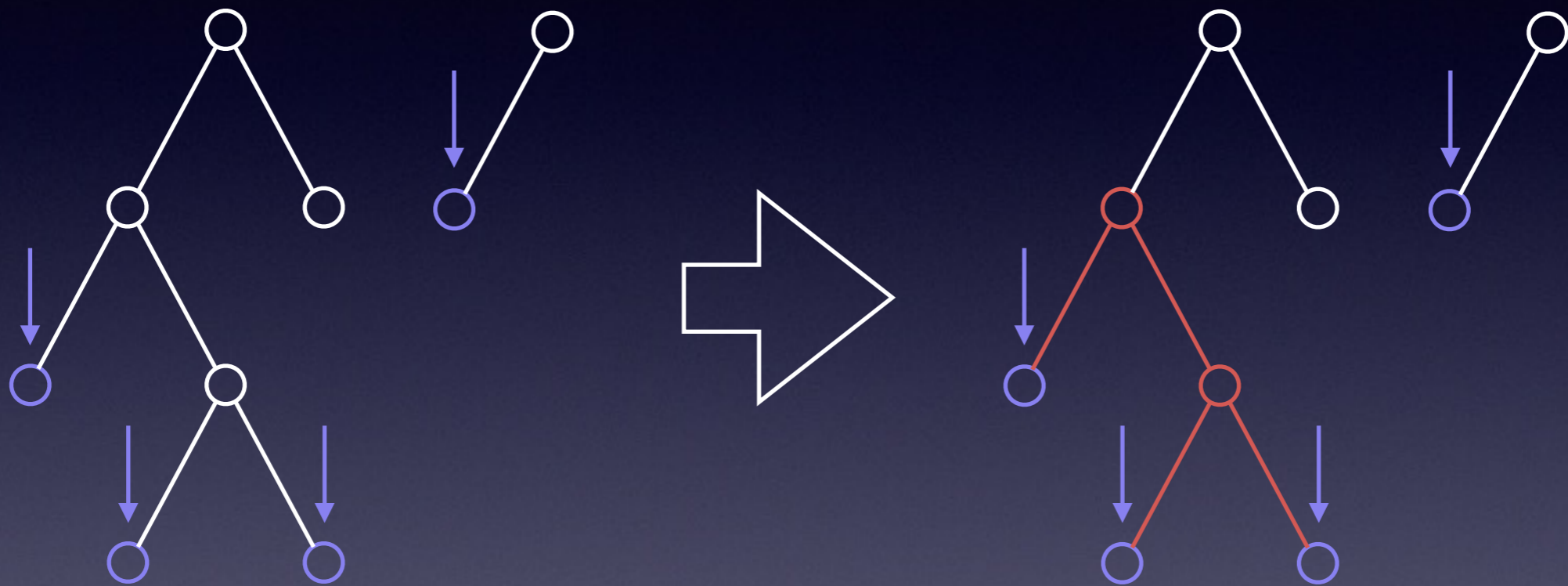
How to compute $P(\text{parents}(\text{Cluster}))$



How to compute $P(\text{parents}(\text{Cluster}))$



How to compute $P(\text{parents}(\text{Cluster}))$



Independence between trees

Complexity : Species x Values^{Parents_Cluster+1}

Algorithm comparison

	FF	ClusteredFF	Exact computation
Complexity	Species x Values ^{NbParents}	Species x Values ^{Parents_Cluster+1}	> Values ^{Species}
Accuracy	Low	? but better than FF	Exact

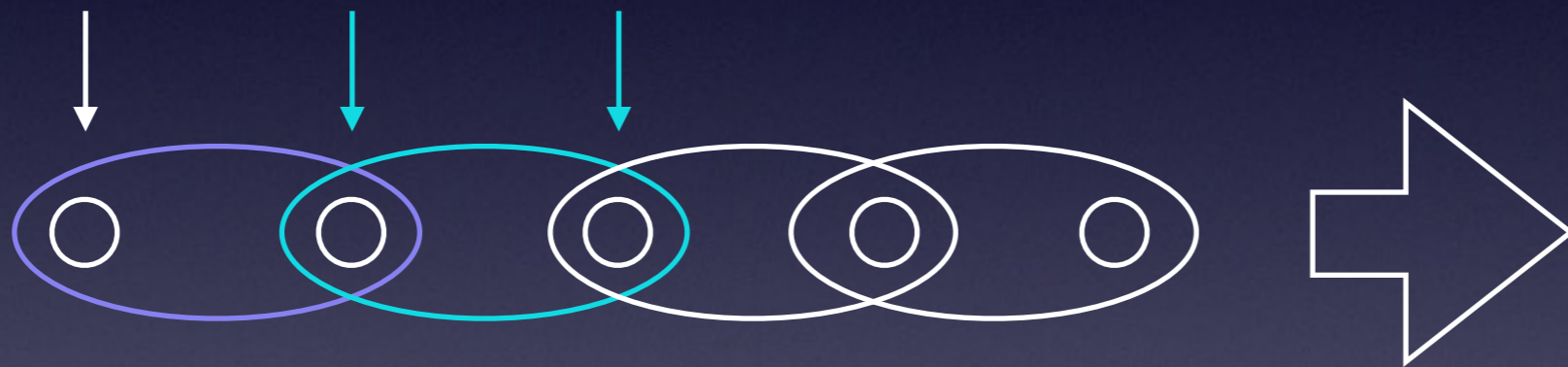
Conclusion

- Our program is currently still being written. Results will tell if the accuracy is good or not.
- After the first results are obtained we will upgrade it to accept bigger clusters and non-tree graphs

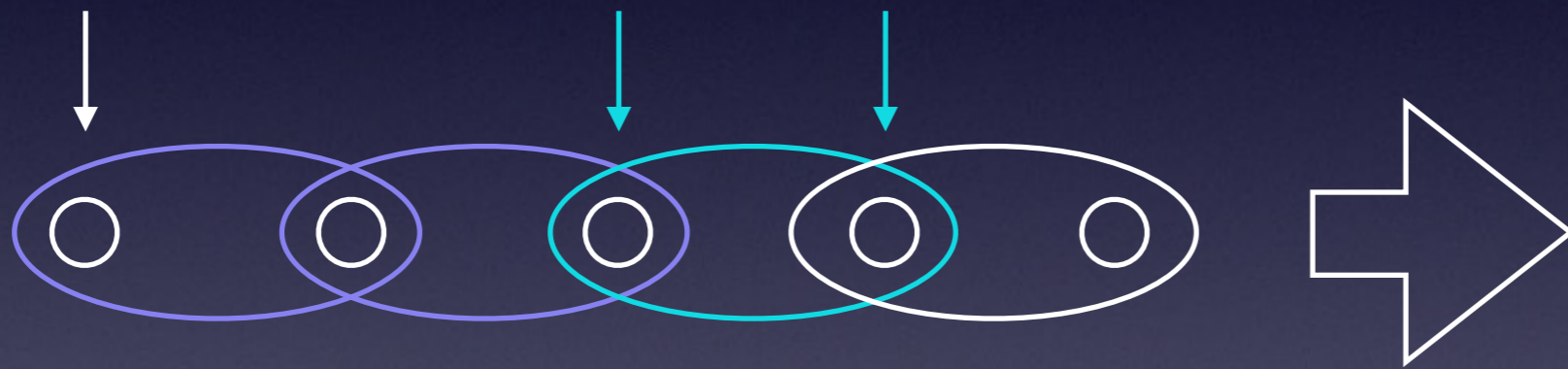
How our algorithm work



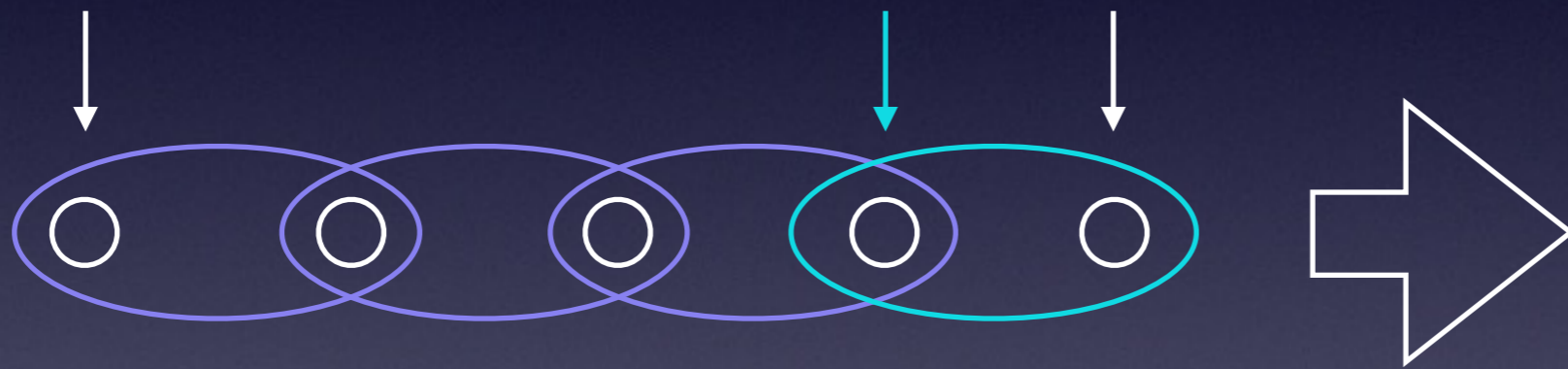
How our algorithm work



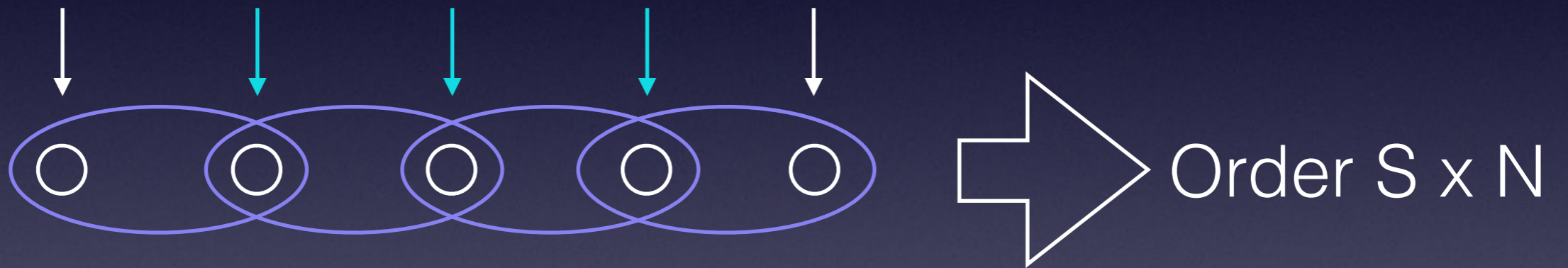
How our algorithm work



How our algorithm work



How our algorithm work

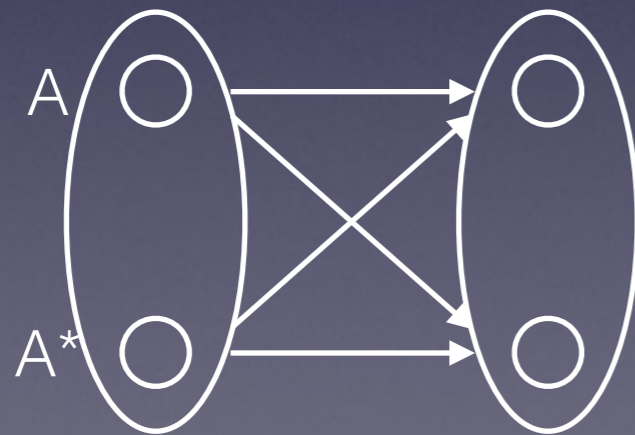


How our algorithm work

- For each time T groups of clusters are found
- Most efficient path is found to calculate each cluster
- Calculate probability using CPTs
- Results are saved, cluster probabilities are kept in memory

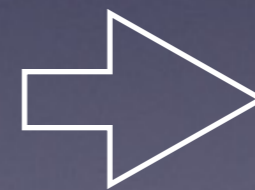
Clustered Factored Frontier

$A \longleftrightarrow A^*$



CPT:

98% : $A = h \ A^* = l \rightarrow A = h$
2% : $A = h \ A^* = l \rightarrow A = l$
2% : $A = l \ A^* = h \rightarrow A = h$
98% : $A = l \ A^* = h \rightarrow A = l$
2% : $A = h \ A^* = l \rightarrow A^* = h$
98% : $A = h \ A^* = l \rightarrow A^* = l$
98% : $A = l \ A^* = h \rightarrow A^* = h$
2% : $A = l \ A^* = h \rightarrow A^* = l$

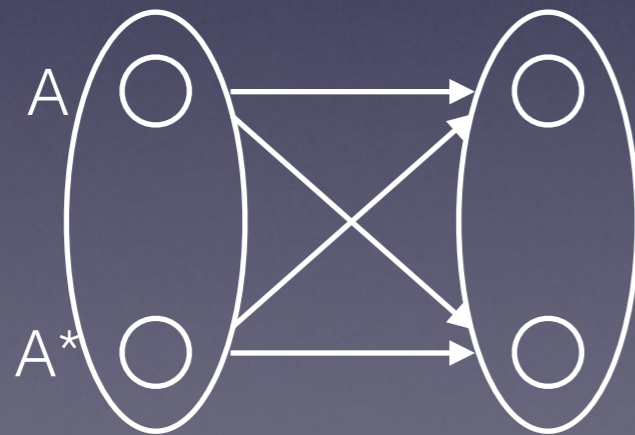


50% $A = h \ A^* = l$
50% $A = l \ A^* = h$:

96.04% $A = h, A^* = l$
0.04% $A = l, A^* = h$
1.96% $A = h, A^* = h$
1.96% $A = l, A^* = l$
0.04% $A = h, A^* = l$
96.04% $A = l, A^* = h$
1.96% $A = h, A^* = h$
1.96% $A = l, A^* = l$

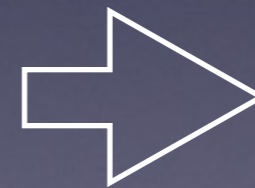
Clustered Factored Frontier

$A \longleftrightarrow A^*$



CPT:

98% : $A = h \ A^* = l \rightarrow A = h$
 2% : $A = h \ A^* = l \rightarrow A = l$
 2% : $A = l \ A^* = h \rightarrow A = h$
 98% : $A = l \ A^* = h \rightarrow A = l$
 2% : $A = h \ A^* = l \rightarrow A^* = h$
 98% : $A = h \ A^* = l \rightarrow A^* = l$
 98% : $A = l \ A^* = h \rightarrow A^* = h$
 2% : $A = l \ A^* = h \rightarrow A^* = l$



50% $A = h \ A^* = l$
 50% $A = l \ A^* = h$:

53.04% $A = h, A^* = l$
 53.04% $A = l, A^* = h$
 1.96% $A = h, A^* = h$
 1.96% $A = l, A^* = l$