16th Brainstorming Week on Membrane Computing (16BWMC)
January 30 – February 2 2018 , Sevilla, Spain

Decision P-trees and Random P-forests

José M. Sempere
Department of Information Systems and Computation
Universitat Politècnica de València

jsempere@dsic.upv.es

16th Brainstorming Week on Membrane Computing (16BWMC)
January 30 – February 2 2018 , Sevilla, Spain

Goals of this work: (1) To implement decision trees and random forests in P systems
(2) To implement a machine learning technique to obtain decision trees and random forests from given data through membrane and object rules

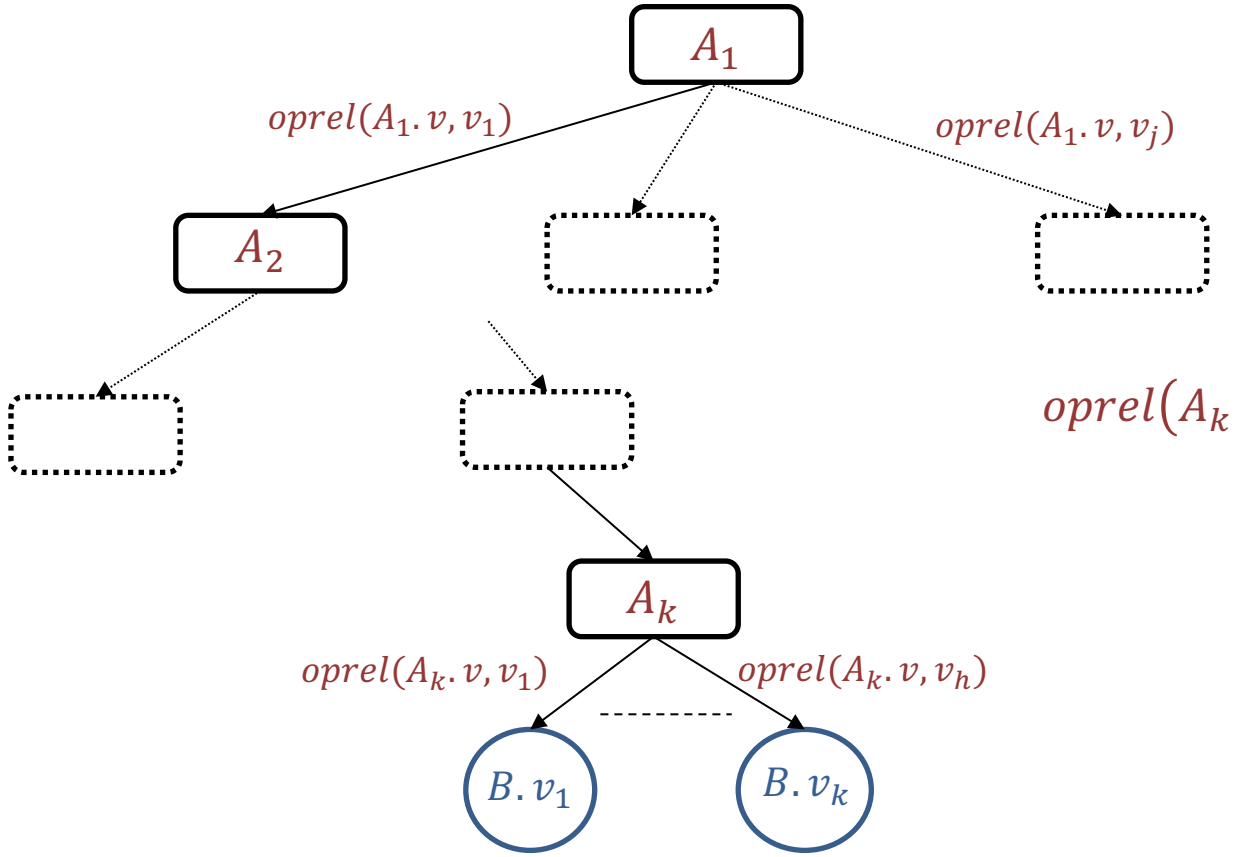
Previous Works

Decision Tree Models Induced by Membrane Systems (2015) J. Wang, J. Hu, M.J. Pérez-Jiménez, A.Riscos-Núñez ROMJIST Vol.18 No.3 pp 228-239

Wang *et al.*		**Our approach**
Tree-like objects	→	Trees defined by the membrane structure
External Induction Algorithm	→	Algorithm runs by P rules within an entropic manner

A Decision Tree is a representation for a discrete-valued function

$$f: A_1 \times A_2 \times \cdots \times A_n \rightarrow B$$



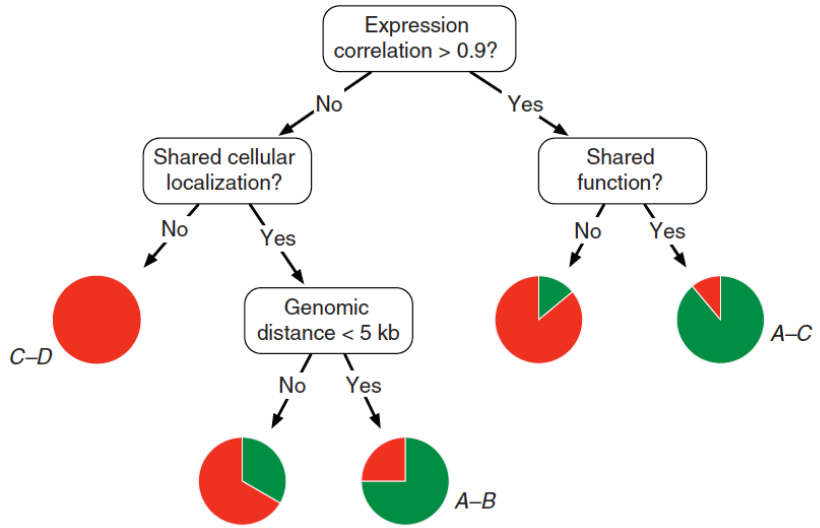
$$oprel(A_k.v, v_p) \in \{A_k.v = v_p, A_k.v \leq v_p, A_k.v \geq v_p, A_k.v \neq v_p, \ldots\}$$

An example: protein-protein interactions prediction

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



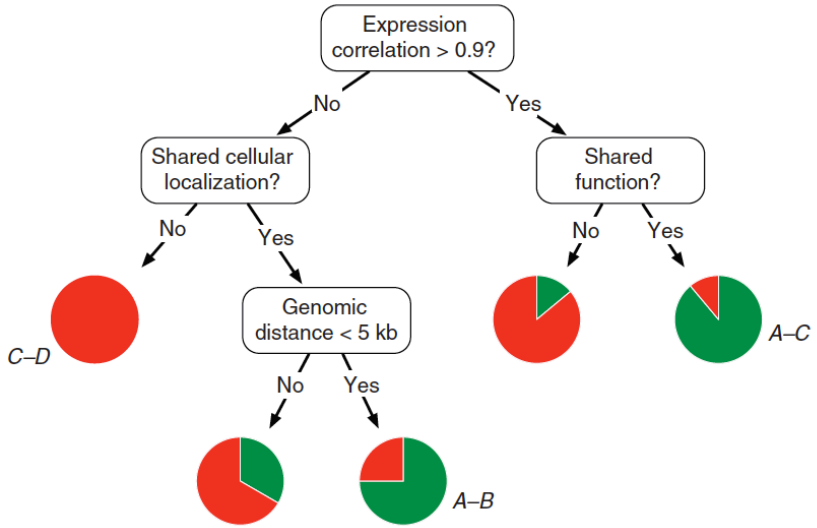
(From **"What are decision trees ?". Carl Kingsford & Steven L. Salzberg. Nature Biotechnology 26 No. 9, 1011 – 1013 (2008)**)

Implementing decision trees by membrane structures (I)

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



Objects alphabet

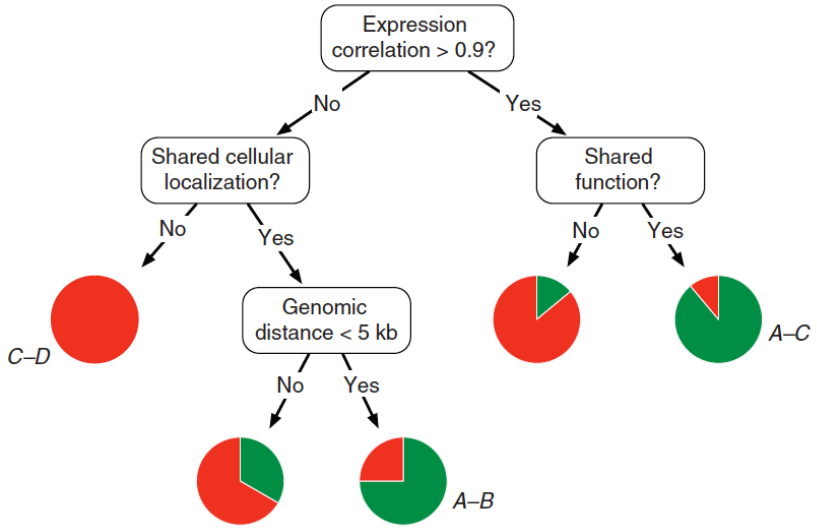
{ECno, ECyes,
Sfno, Sfyes,
SCLno,SCLyes,
GDL5yes,GDL5no
YES,NO}

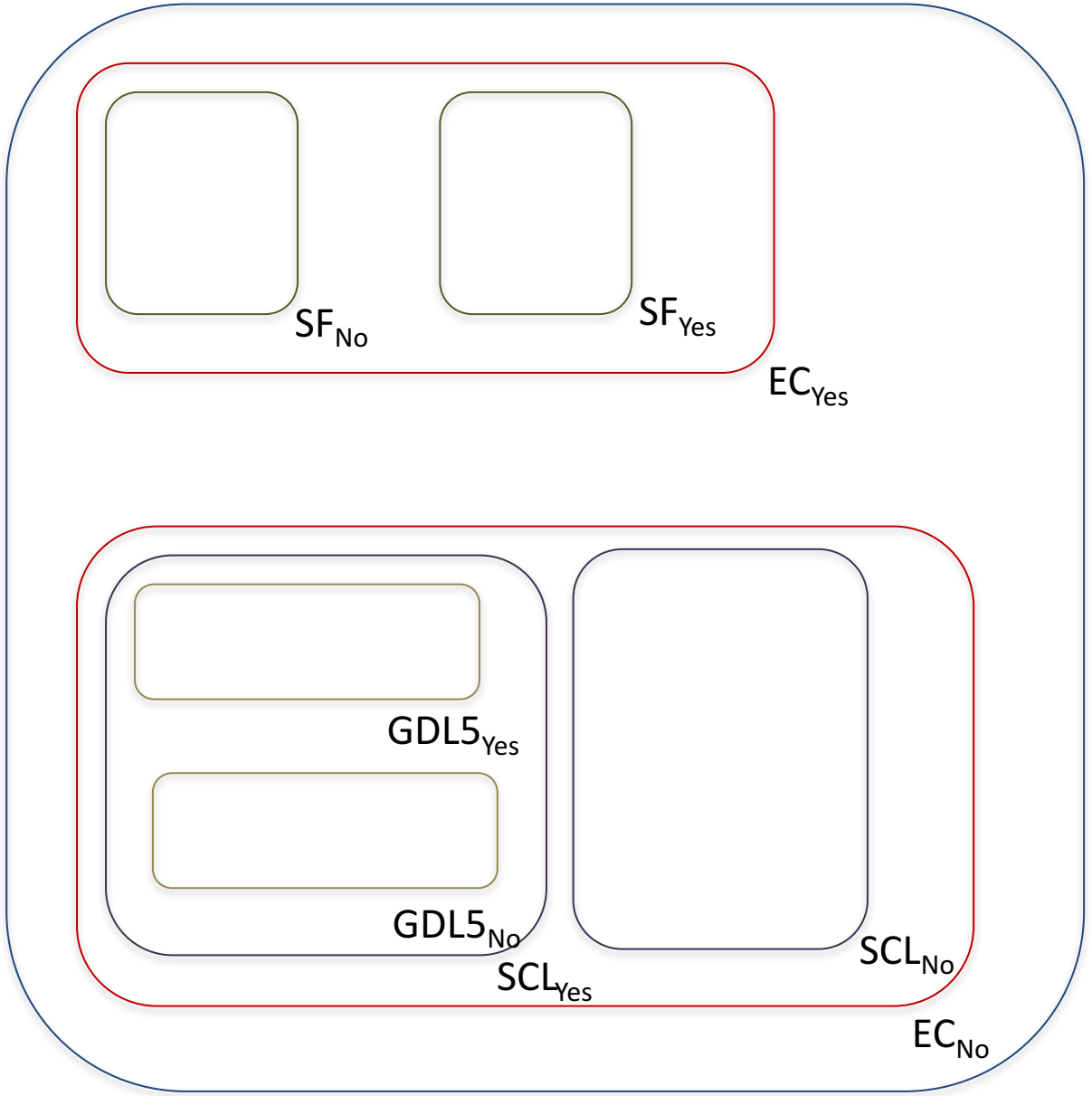
Implementing decision trees by membrane structures (II)

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



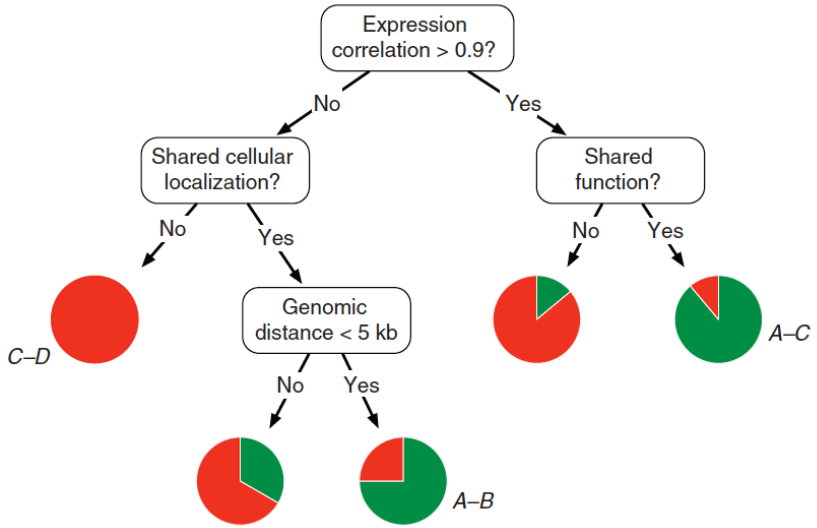


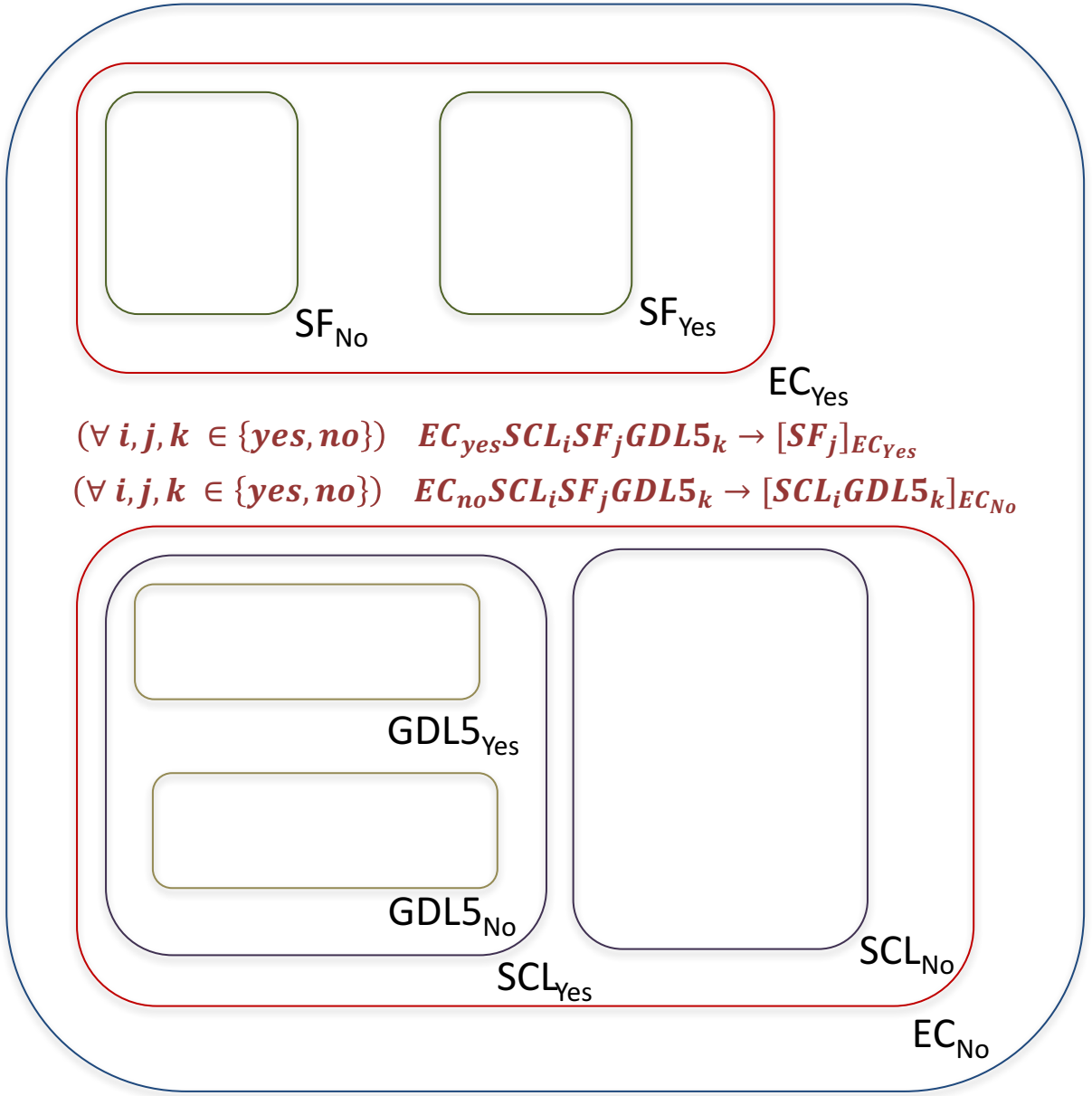
Implementing decision trees by membrane structures (III)

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



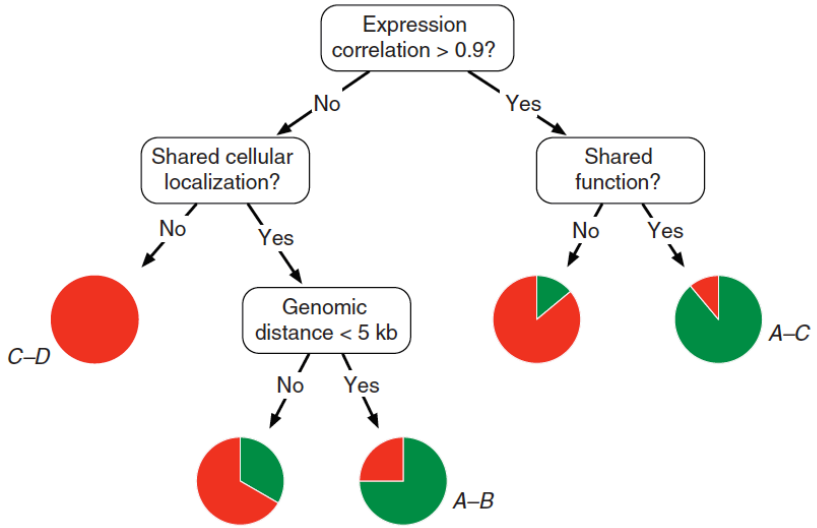


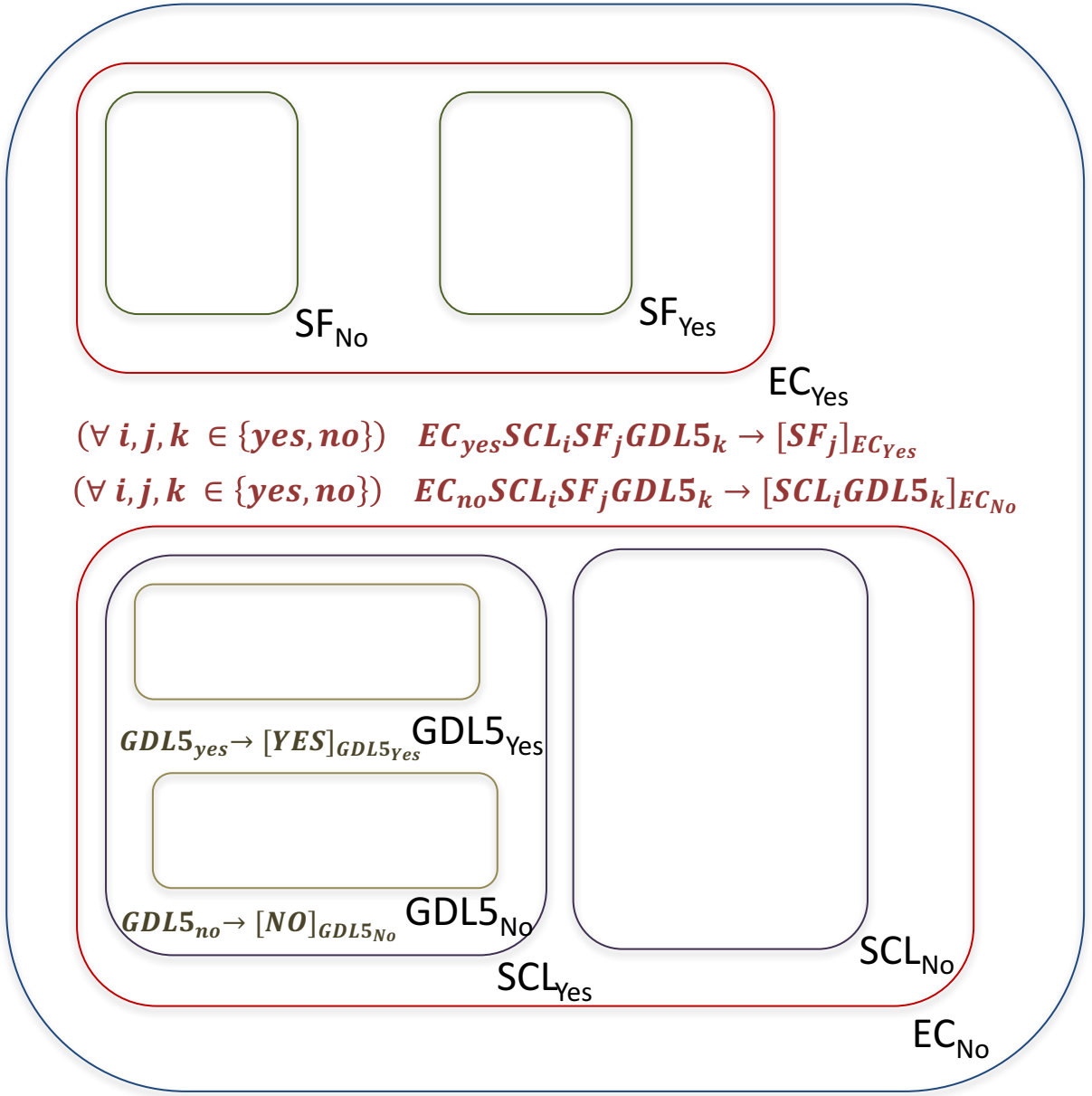
Implementing decision trees by membrane structures (IV)

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



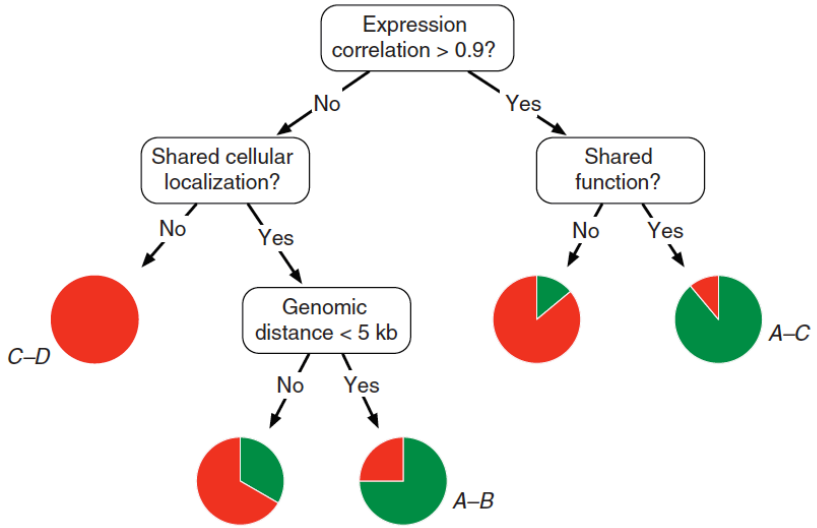


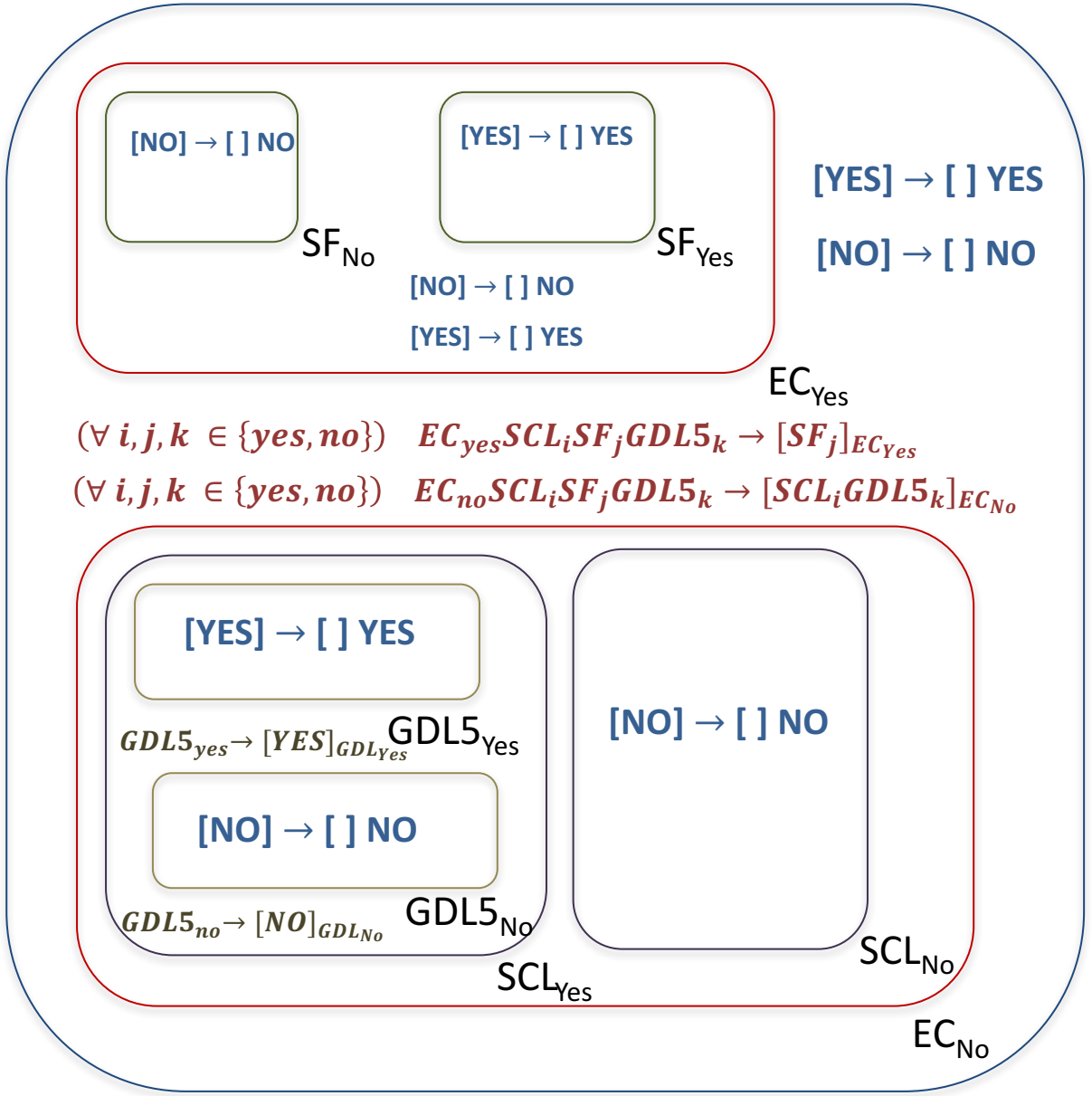
Implementing decision trees by membrane structures (V)

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



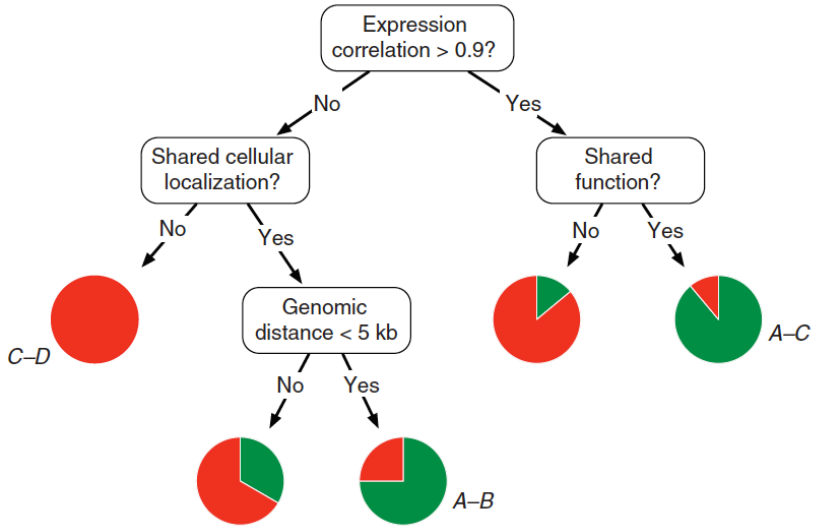


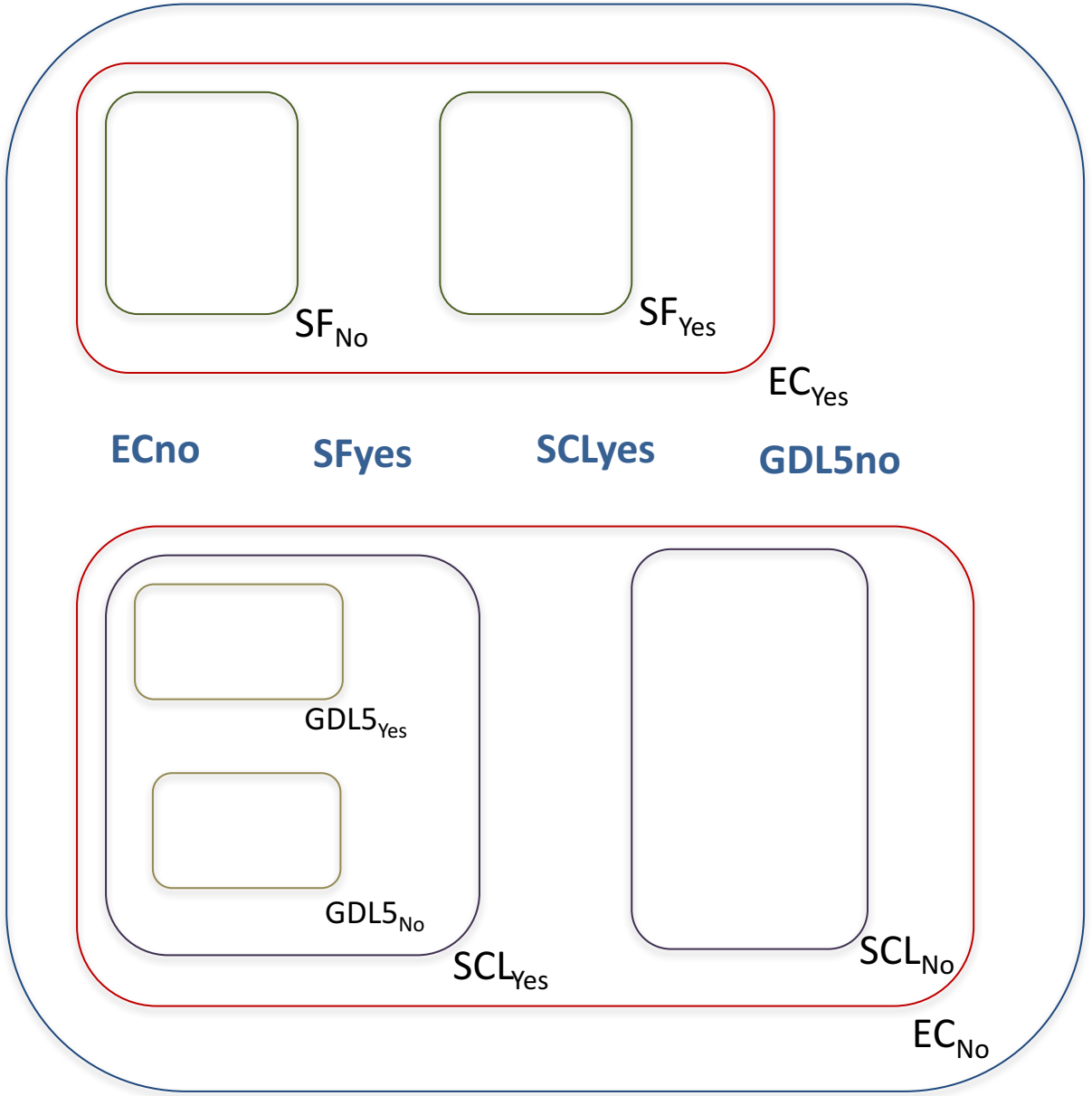
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



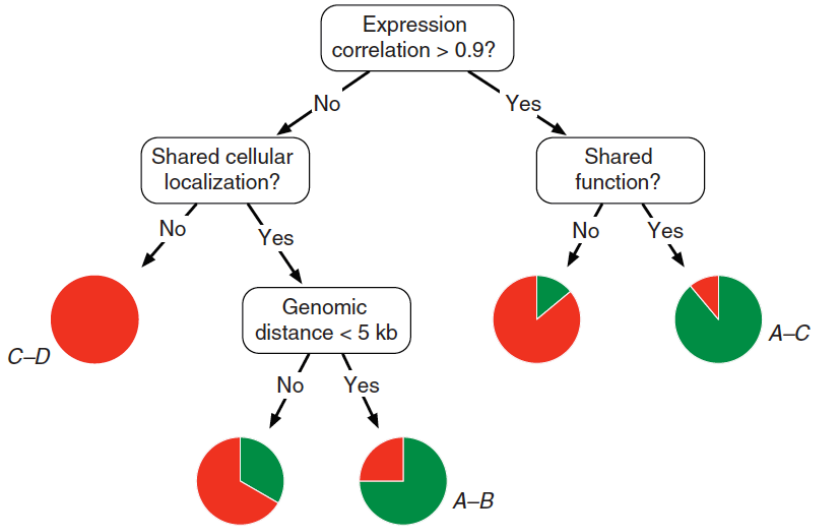


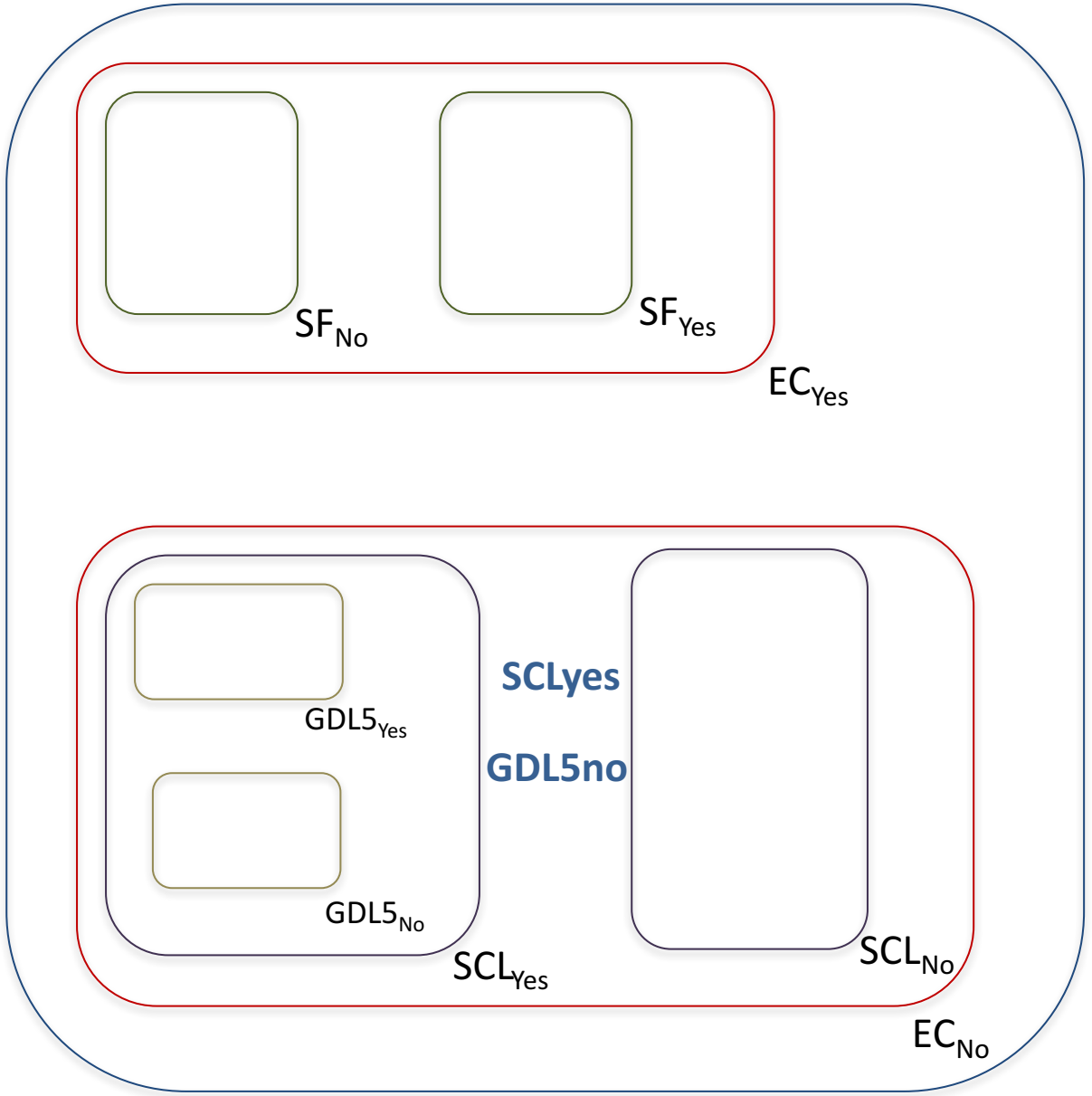
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



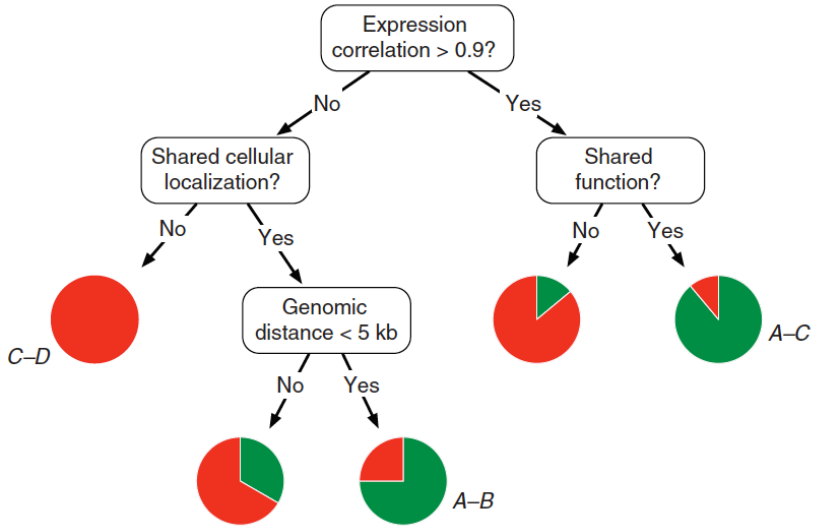


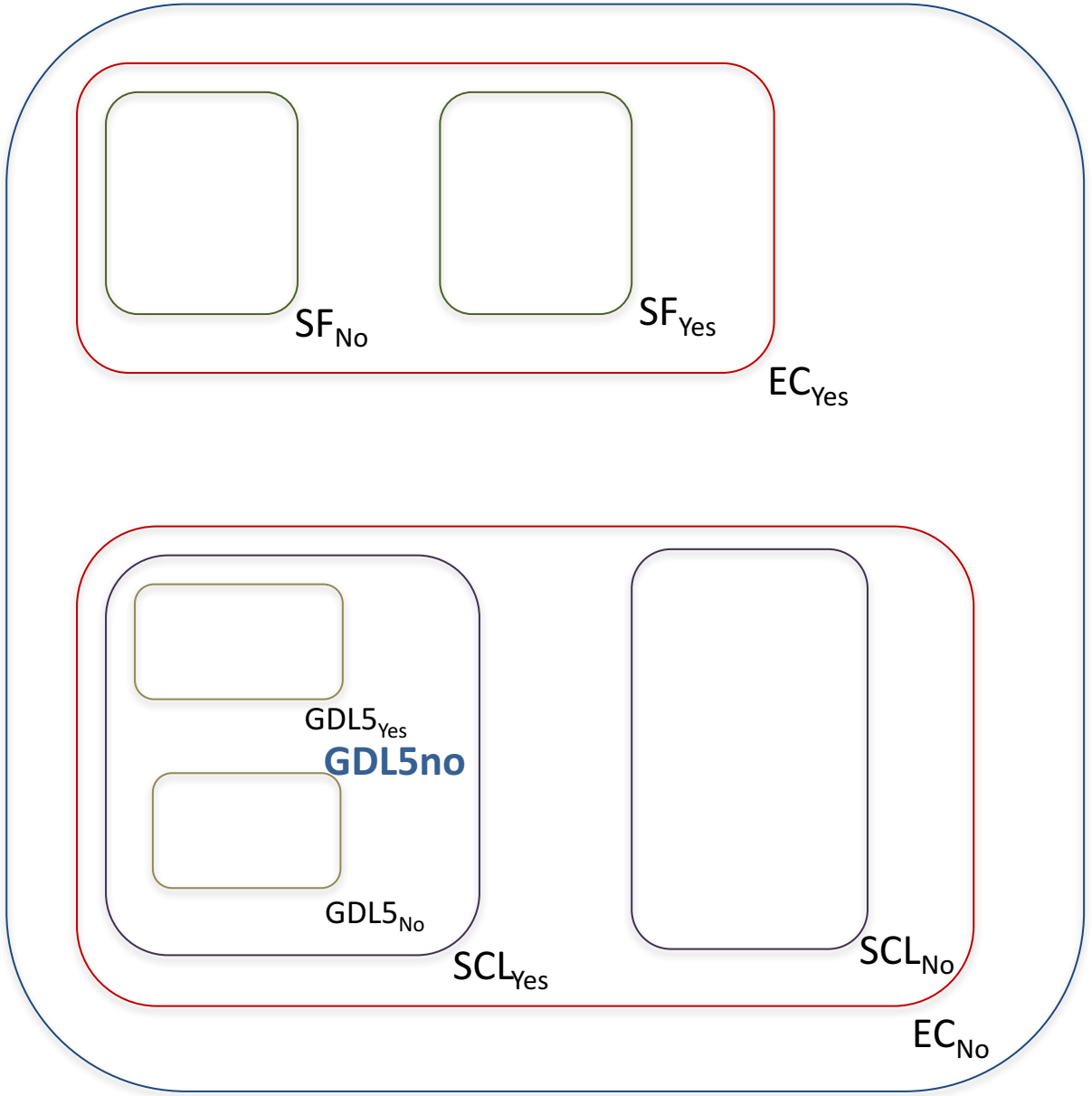
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



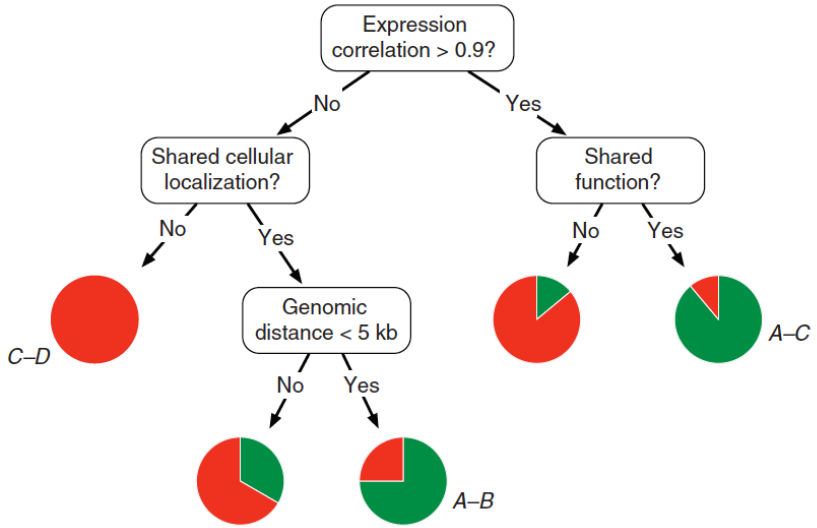


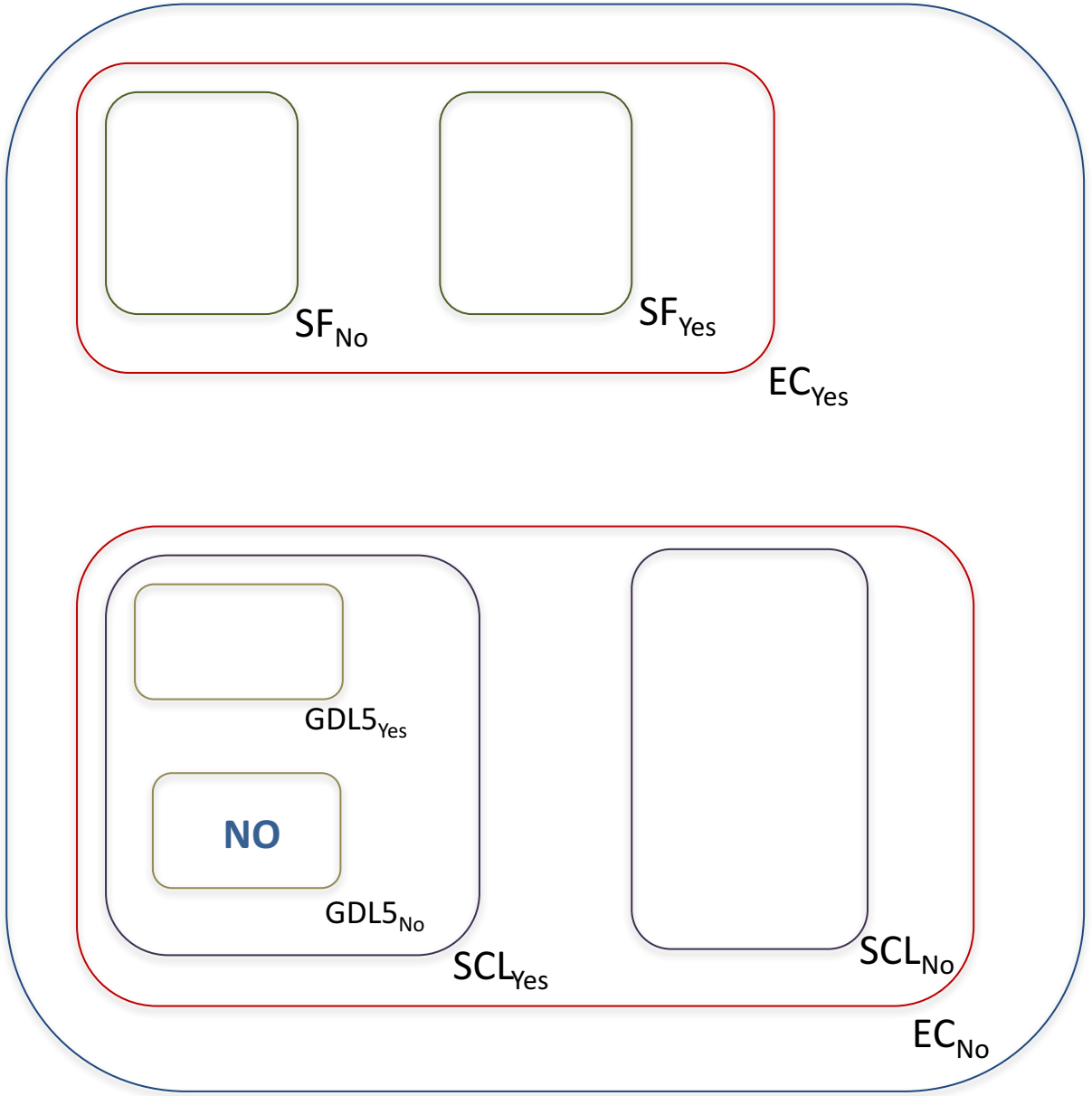
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



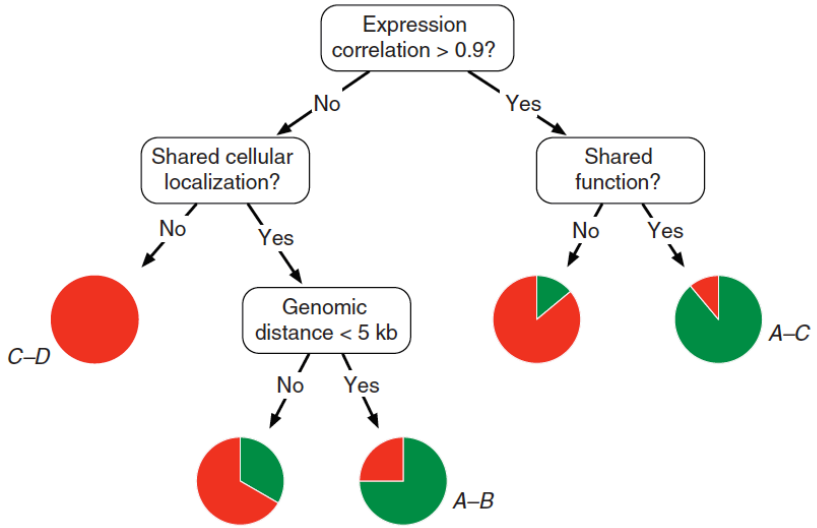


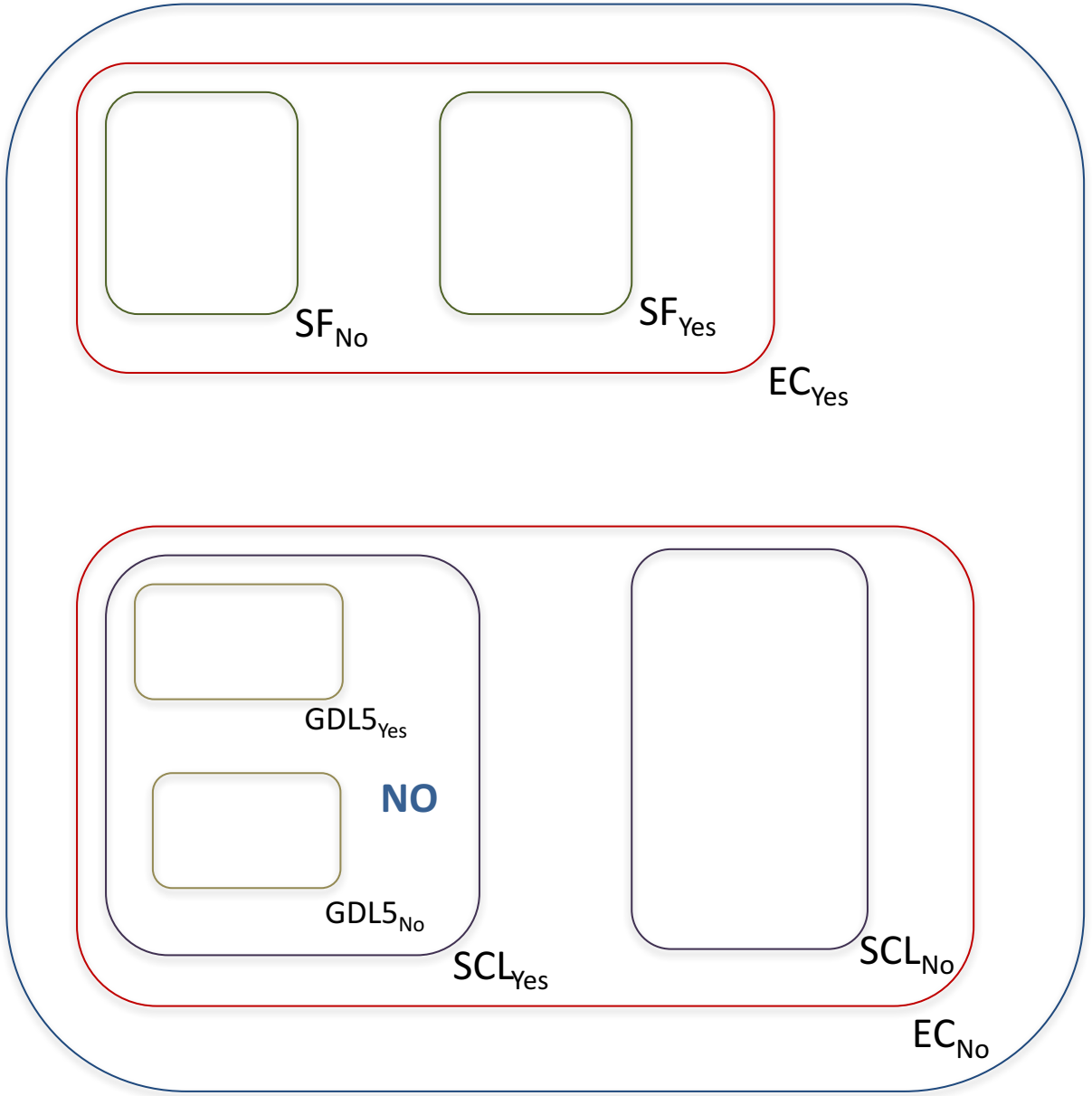
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



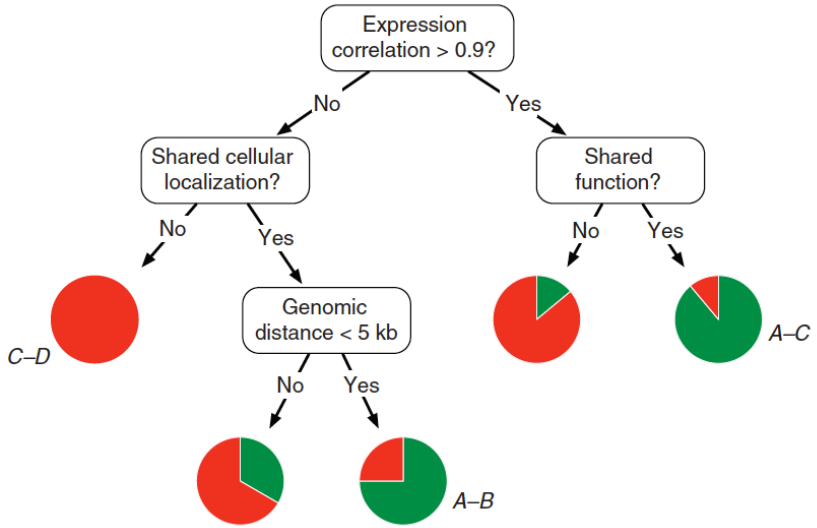


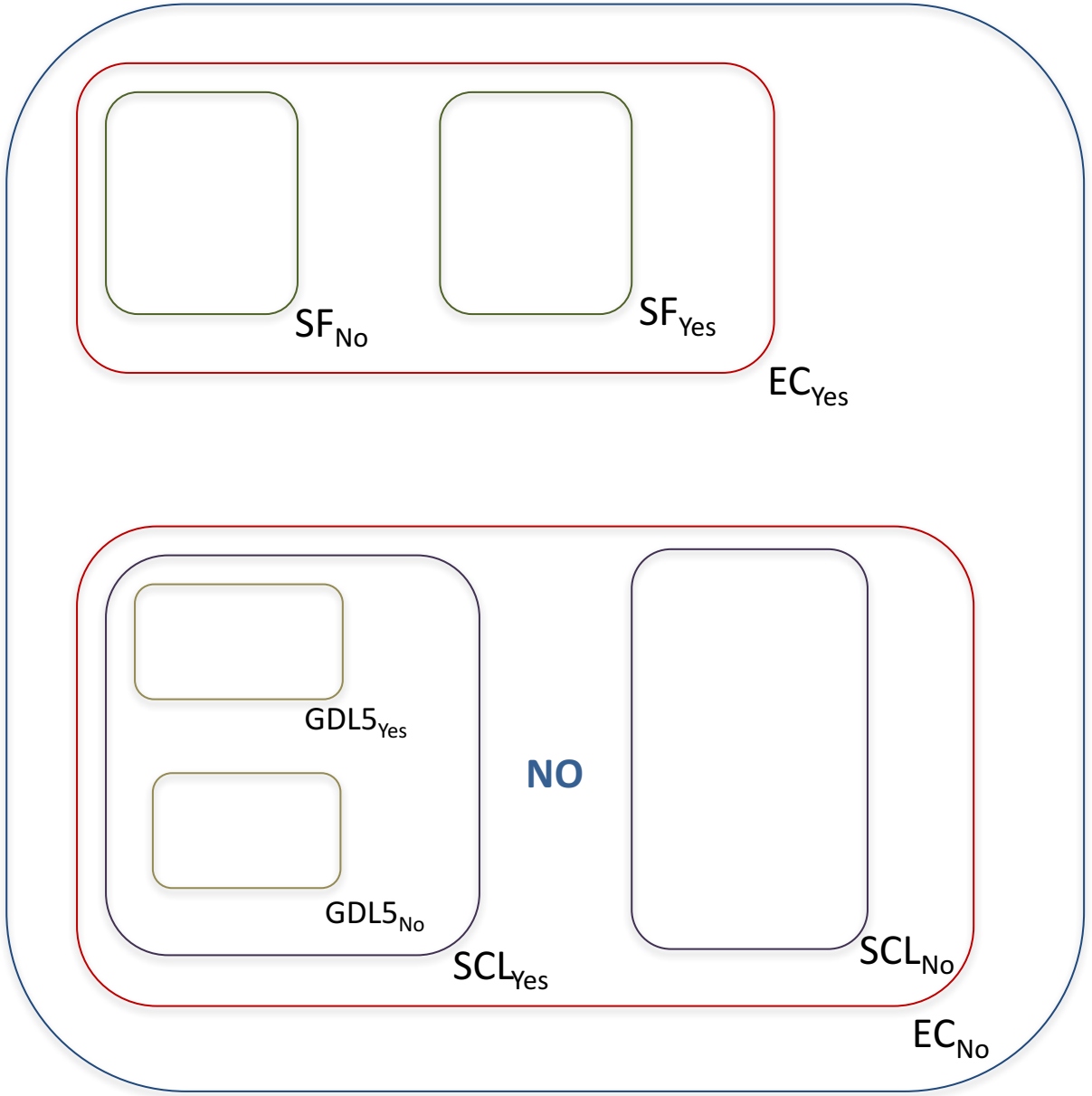
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



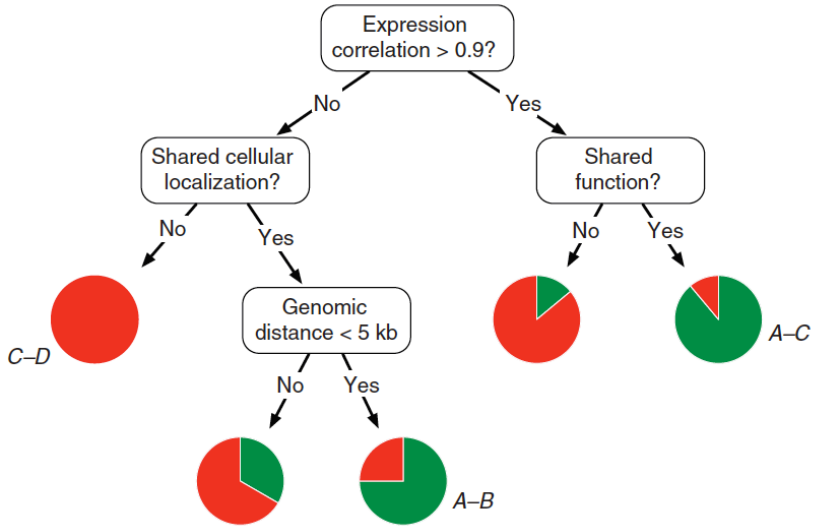


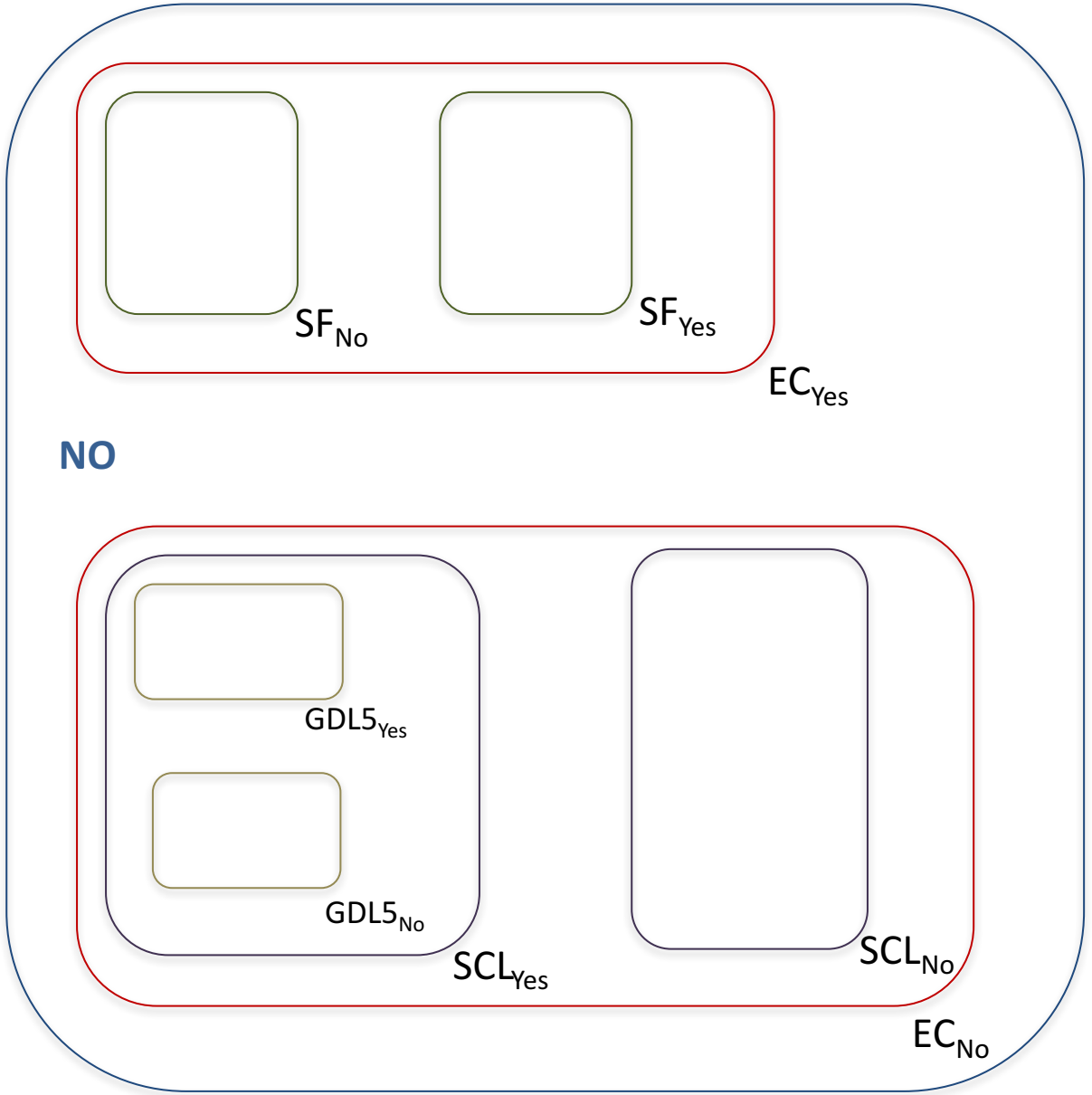
A parsing example

a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



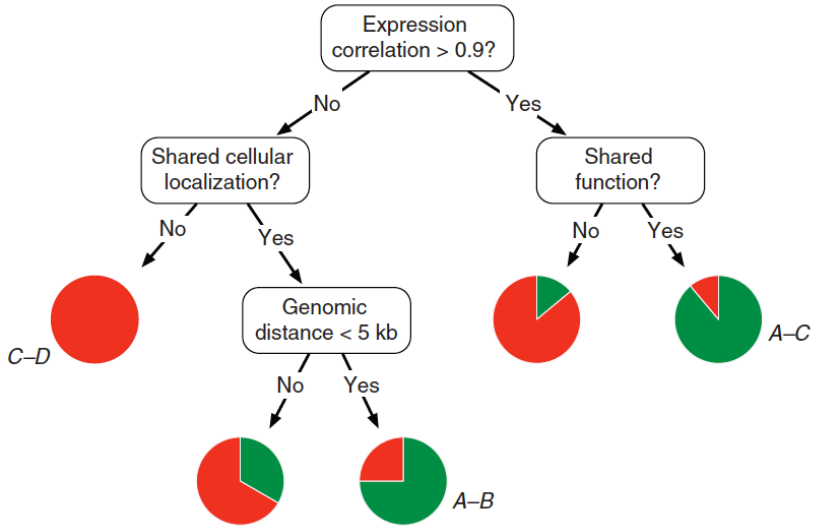


A parsing example

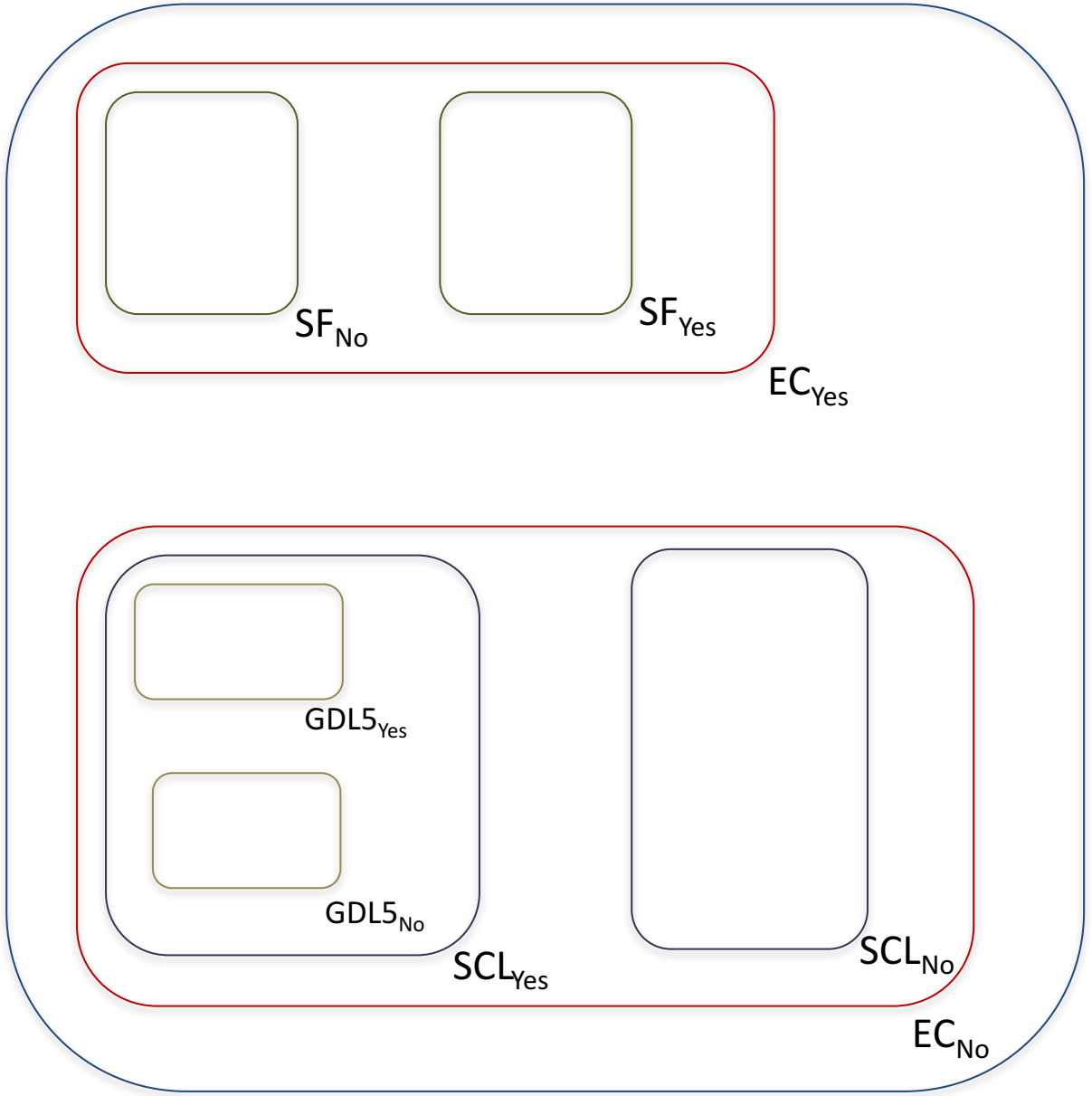
a

Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



NO



Applying a machine learning technique inside a P system

Sample	X	Y	Z	Decision
1	High	High	High	Yes
2	High	High	High	Yes
3	Low	High	Low	Yes
4	Medium	High	High	Not

Applying a machine learning technique inside a P system

Sample	X	Y	Z	Decision
1	High	High	High	Yes
2	High	High	High	Yes
3	Low	High	Low	Yes
4	Medium	High	High	Not

$$X^1_{High}\ Y^1_{High}\ Z^1_{High}\ Decision^1_{YES}$$

$$X^2_{High}\ Y^2_{High}\ Z^2_{High}\ Decision^2_{YES}$$

$$X^3_{Low}\ Y^3_{High}\ Z^3_{Low}\ Decision^3_{YES}$$

$$X^4_{Medium}\ Y^4_{High}\ Z^4_{Low}\ Decision^4_{NO}$$

Applying a machine learning technique inside a P system

Sample	X	Y	Z	Decision
1	High	High	High	Yes
2	High	High	High	Yes
3	Low	High	Low	Yes
4	Medium	High	High	Not

$X^1_{High}\ Y^1_{High}\ Z^1_{High}\ Decision^1_{YES}$

$X^2_{High}\ Y^2_{High}\ Z^2_{High}\ Decision^2_{YES}$

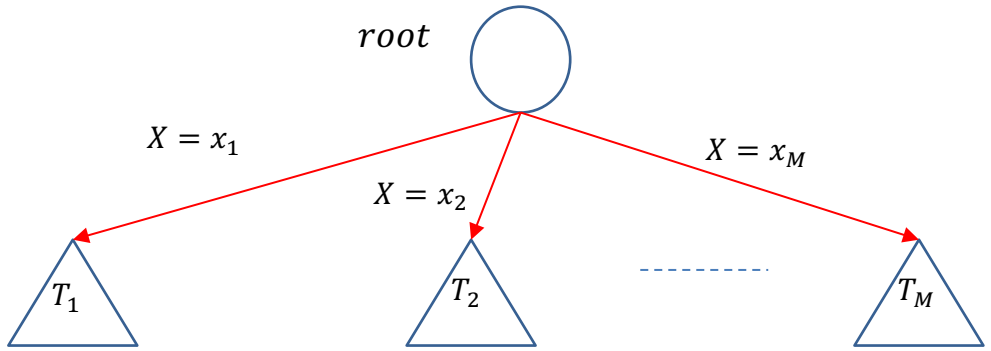
$X^3_{Low}\ Y^3_{High}\ Z^3_{Low}\ Decision^3_{YES}$

$X^4_{Medium}\ Y^4_{High}\ Z^4_{Low}\ Decision^4_{NO}$

A generic algorithm to build decision trees from data

Input: A finite set of supervised tuples E
Output: A decision tree T
Method:
1) Create an arbitrary root
2) If all the tuples belong to class C_j
   then *return*(root, C_j)
   else
   1. Select an attribute X with values $x_1, x_2, \dots, x_M$
   2. Make a partition of E according to the attribute value $E_1, E_2, \dots, E_M : E = \cup_{i=1}^{M} E_i$
   3. Build decision trees for every subset: $T_1, T_2, \dots, T_M$

endMethod



Applying a machine learning technique inside a P system

Sample	X	Y	Z	Decision
1	High	High	High	Yes
2	High	High	High	Yes
3	Low	High	Low	Yes
4	Medium	High	High	Not

$X^1_{High}\ Y^1_{High}\ Z^1_{High}\ Decision^1_{YES}$

$X^2_{High}\ Y^2_{High}\ Z^2_{High}\ Decision^2_{YES}$

$X^3_{Low}\ Y^3_{High}\ Z^3_{Low}\ Decision^3_{YES}$

$X^4_{Medium}\ Y^4_{High}\ Z^4_{Low}\ Decision^4_{NO}$

A generic algorithm to build decision trees from data

Input: A finite set of supervised tuples E
Output: A decision tree T
Method:
1) Create an arbitrary root
2) If all the tuples belong to class C_j
   then *return*(root, C_j)
   else
      1. Select an attribute X with values $x_1, x_2, \dots, x_M$
      2. Make a partition of E according to the attribute value $E_1, E_2, \dots, E_M : E = \cup_{i=1}^{M} E_i$
      3. Build decision trees for every subset: $T_1, T_2, \dots, T_M$

endMethod

Suppose that the attribute for the node root is X, then we apply the following three rules (membrane creation)

$X^1_{High}\ Y^1_{High}\ Z^1_{High}\ Decision^1_{YES}\ X^2_{High}\ Y^2_{High}\ Z^2_{High}\ Decision^2_{YES} \rightarrow [\ Y^1_{High}\ Z^1_{High}\ Decision^1_{YES} Y^2_{High}\ Z^2_{High}\ Decision^2_{YES}]_{X_{High}}$

$X^3_{Low}\ Y^3_{High}\ Z^3_{Low}\ Decision^3_{YES} \rightarrow [\ Y^3_{High}\ Z^3_{Low}\ Decision^3_{YES}]_{X_{Low}}$

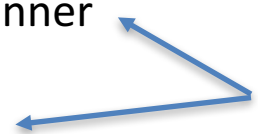
$X^4_{Medium}\ Y^4_{High}\ Z^4_{Low}\ Decision^4_{NO} \rightarrow [\ Y^4_{High}\ Z^4_{Low}\ Decision^4_{NO}]_{X_{Medium}}$

16th Brainstorming Week on Membrane Computing (16BWMC)
January 30 – February 2 2018 , Sevilla, Spain

Applying a machine learning technique inside a P system

How can the appropriate rules be selected ?

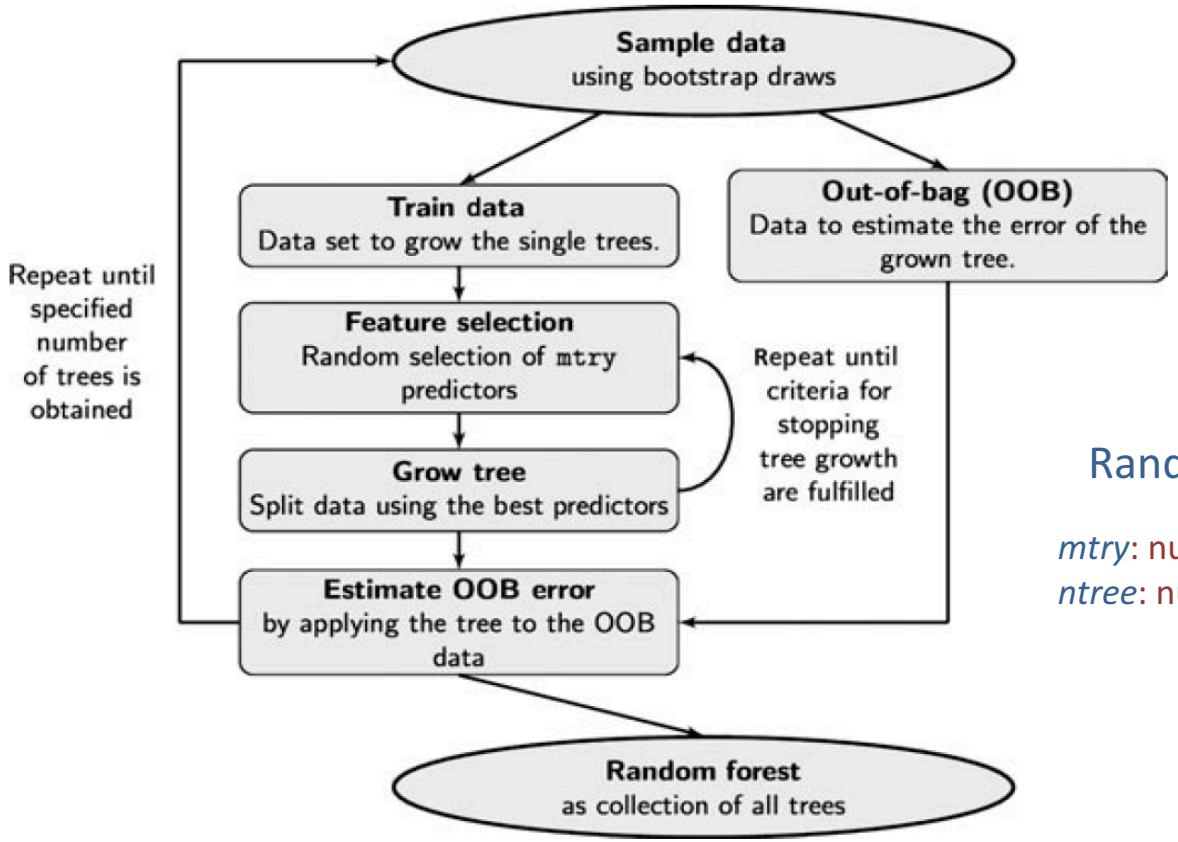
- External algorithm selects the rule and we apply a translation scheme from decision trees to P systems
- Apply a P system working within an entropic manner
- Apply rules according to a functional criterium



Work in Progress

$$f: M(V) \rightarrow R$$

From decision trees to random forests



Random Forest algorithm

mtry: number of features for node splitting
ntree: number of trees in the forest

From cell-like P systems to tissue-like P systems

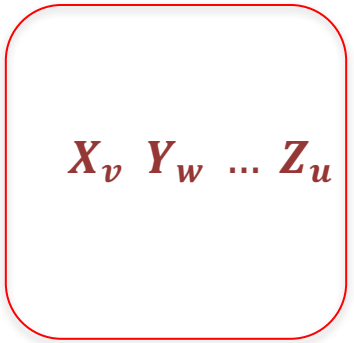
$X_v\ Y_w\ \dots\ Z_u$

Initially we have only one cell with cell-creation rules and all the input data

$$X_v\ Y_w\ \dots\ Z_u \rightarrow X_v\ Y_w\ \dots\ Z_u(X_v\ Y_w)$$

$$X_v\ Y_w\ \dots\ Z_u \rightarrow X_v\ Y_w\ \dots\ Z_u(Y_w\ Z_u)$$

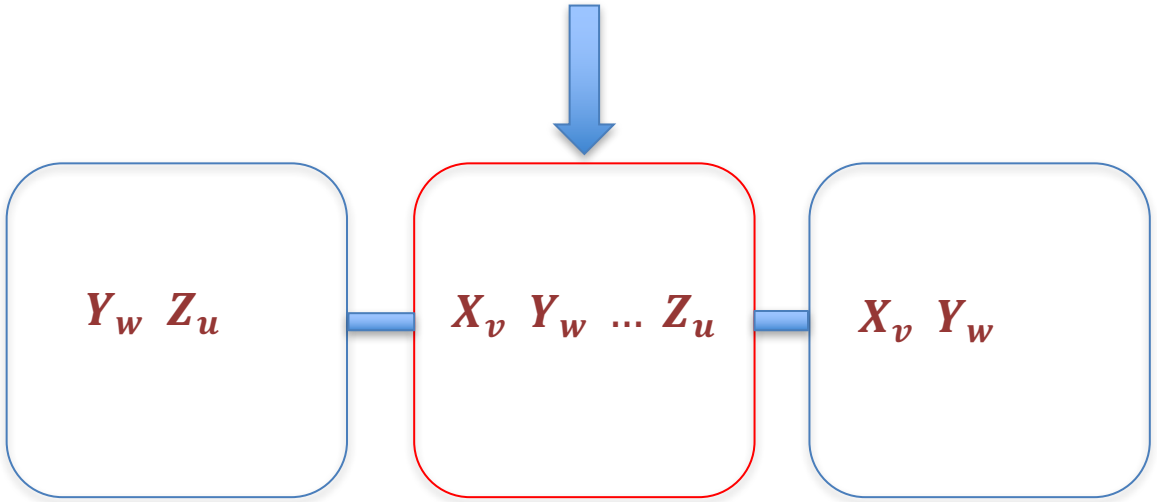
From cell-like P systems to tissue-like P systems



Initially we have only one cell with cell-creation rules and all the input data

$$X_v\ Y_w\ \ldots\ Z_u \rightarrow X_v\ Y_w\ \ldots\ Z_u(X_v\ Y_w)$$

$$X_v\ Y_w\ \ldots\ Z_u \rightarrow X_v\ Y_w\ \ldots\ Z_u(Y_w\ Z_u)$$



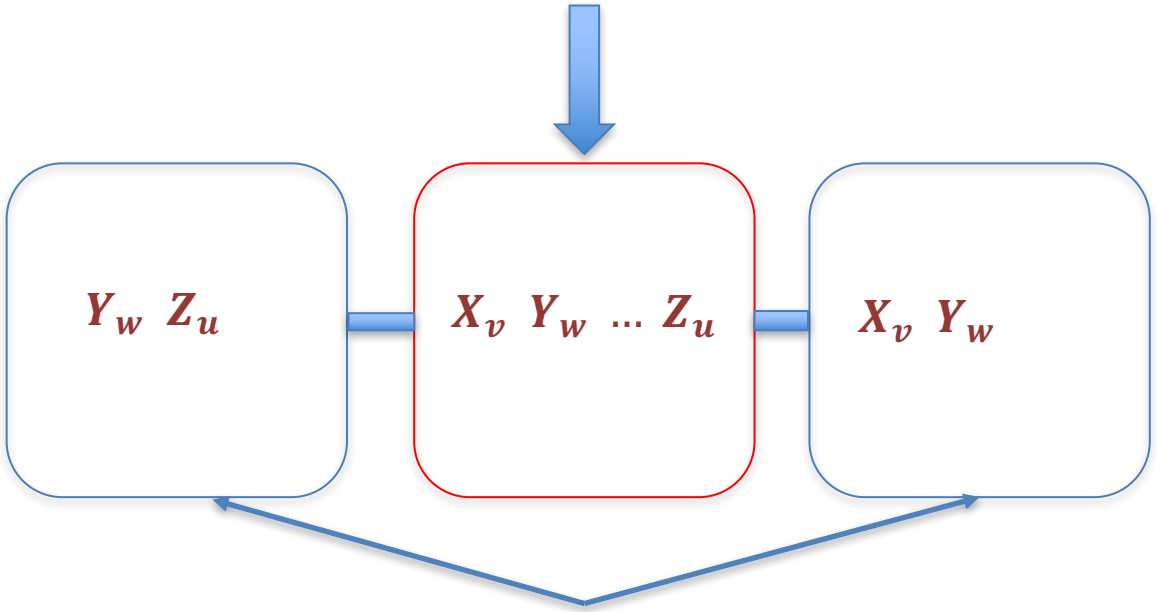
From cell-like P systems to tissue-like P systems

$X_v\ Y_w\ \ldots\ Z_u$

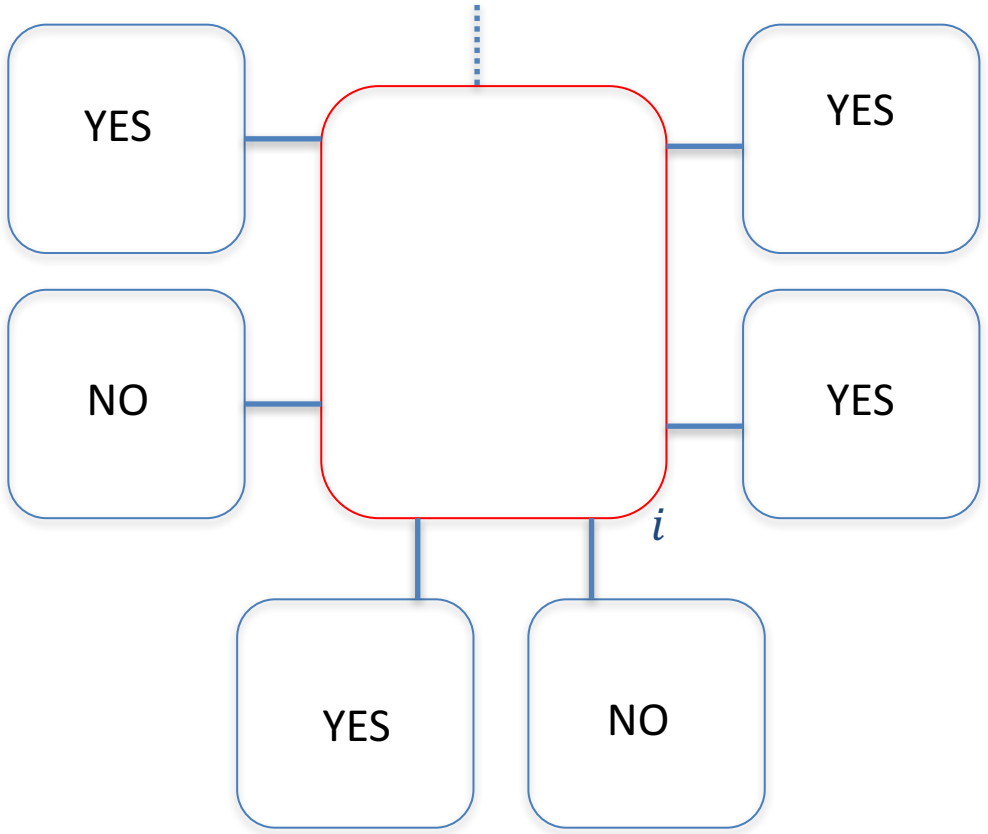
Initially we have only one cell with cell-creation rules and all the input data

$$X_v\ Y_w\ \ldots\ Z_u \rightarrow X_v\ Y_w\ \ldots\ Z_u(X_v\ Y_w)$$

$$X_v\ Y_w\ \ldots\ Z_u \rightarrow X_v\ Y_w\ \ldots\ Z_u(Y_w\ Z_u)$$



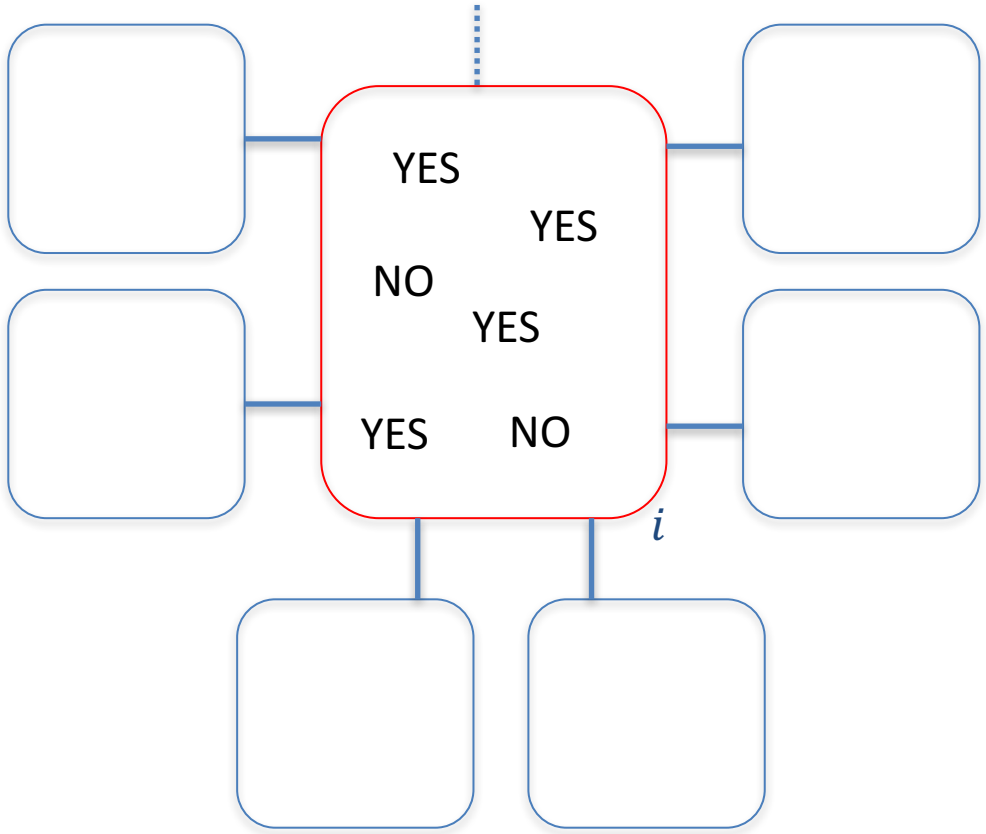
Apply rules to create a decision tree

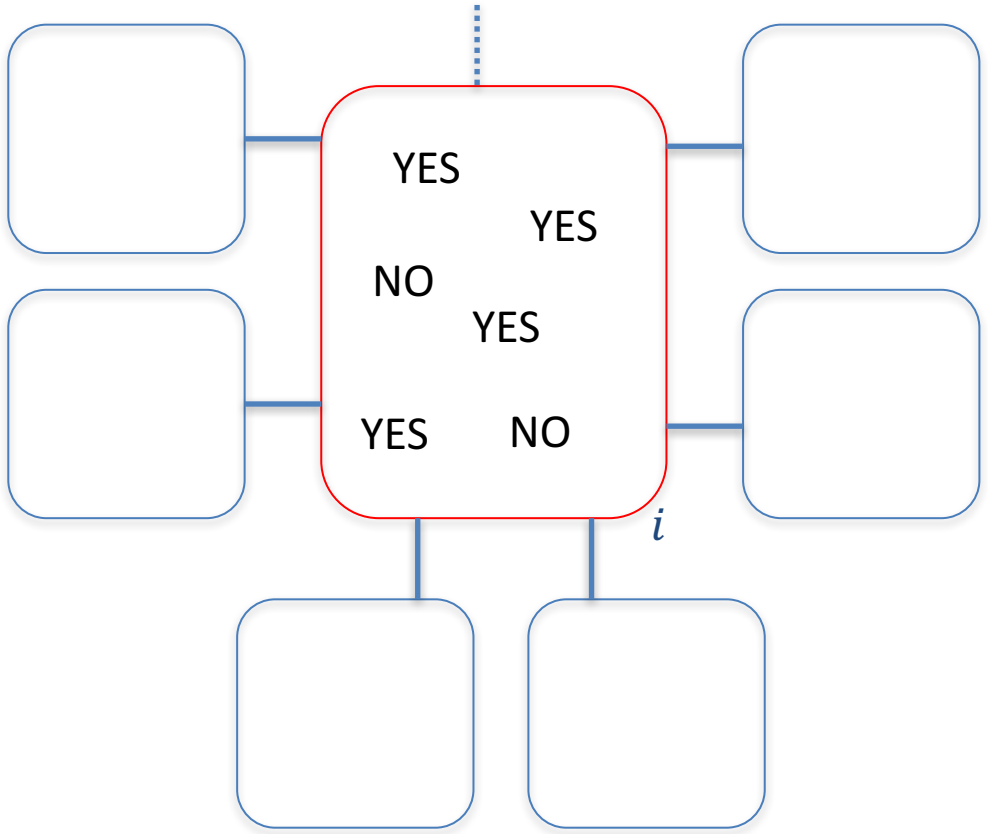


Rules inside every P system at skin region (communication rules)

YES $\rightarrow (\boldsymbol{YES})_i$

NO $\rightarrow (\boldsymbol{NO})_i$

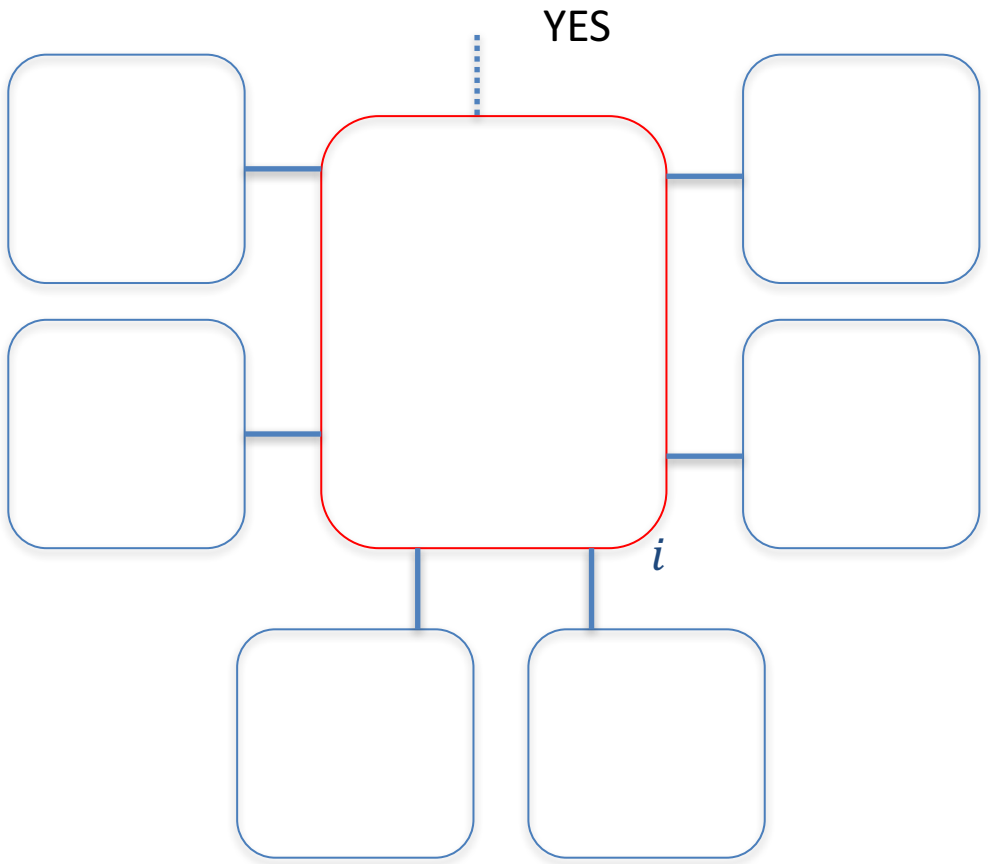

YES
YES
NO
YES
YES
NO
i



Majority rules

$$\#YES > \#NO \rightarrow YES_{out}$$

$$\#NO > \#YES \rightarrow NO_{out}$$



Majority rules

$$\#YES > \#NO \rightarrow YES_{out}$$

$$\#NO > \#YES \rightarrow NO_{out}$$

THAT'S ALL

THANK YOU !!!