

# WAVE K,D-RANDOM FAMILY SEARCH

Parcours interactif d'un espace de solutions

Etienne Lehembre<sup>1</sup>

<sup>1</sup>Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, FRANCE

`etienne.lehembre@unicaen.fr`



1. Sujet et enjeux
2. Interactivité et adaptation
3. Oracle et expériences
4. Conclusion

## Données :

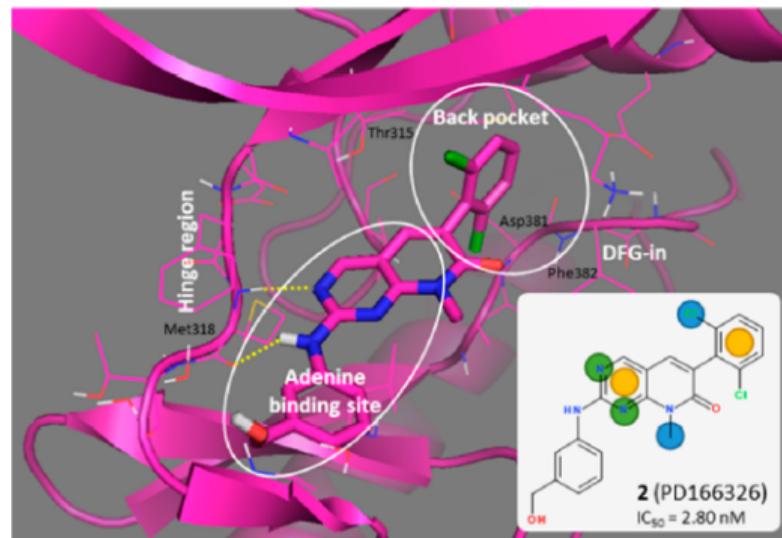
- ▶ 1485 molécules
- ▶ 1 récepteur : BCR-ABL

## Motifs :

- ▶ 112 363 pharmacophores
- ▶ 15 533 classes d'équivalence

## But :

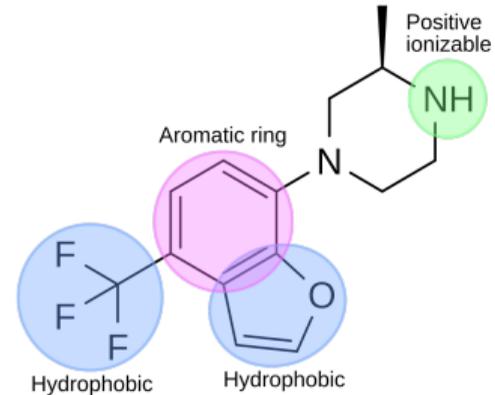
Guider l'exploration d'un ensemble de solutions par un expert.



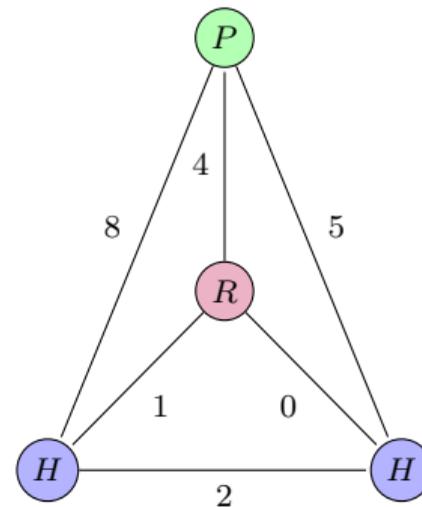
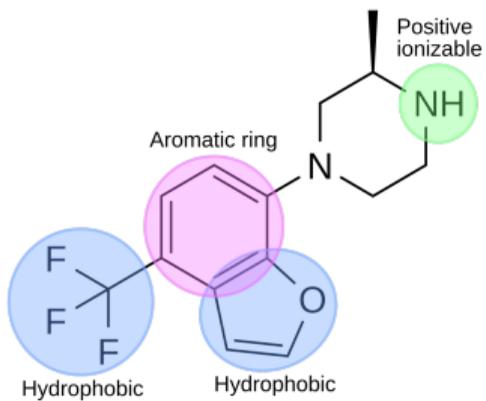
*J.-P. METIVIER, B. CUISSART, R. BUREAU, and A. LEPAILLEUR.*

## Description des molécules

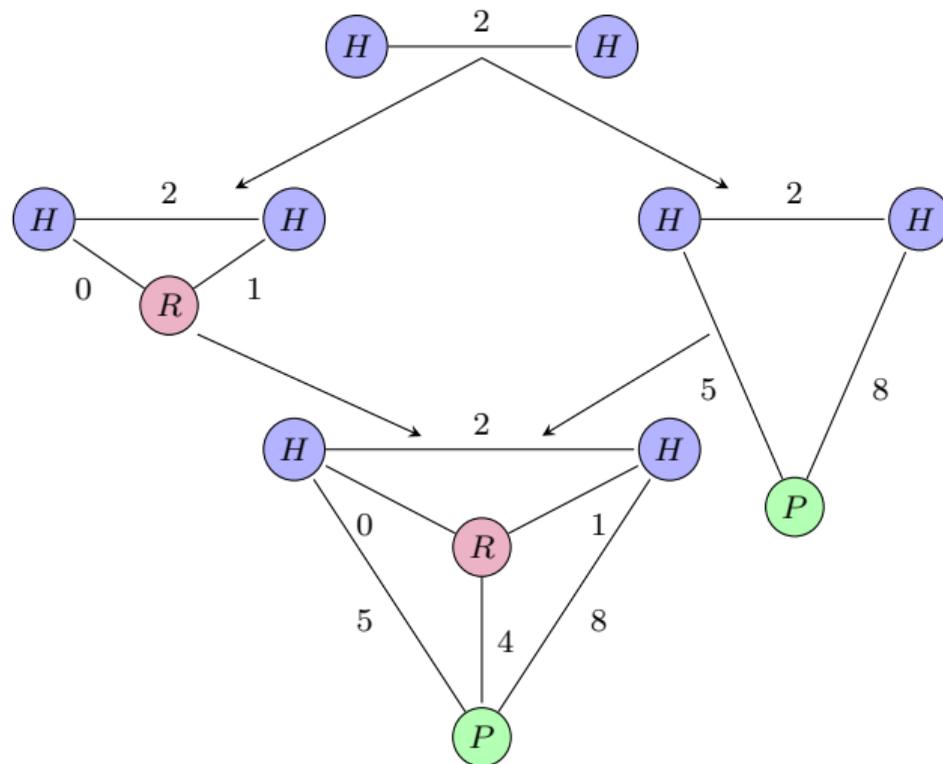
- ▶ Marqueurs pharmacophoriques.
- ▶ Nombre d'arêtes entre les marqueurs.



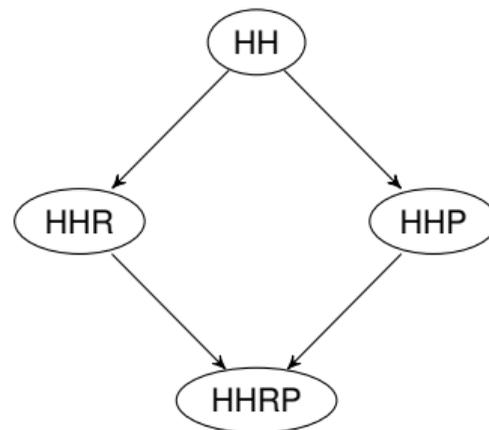
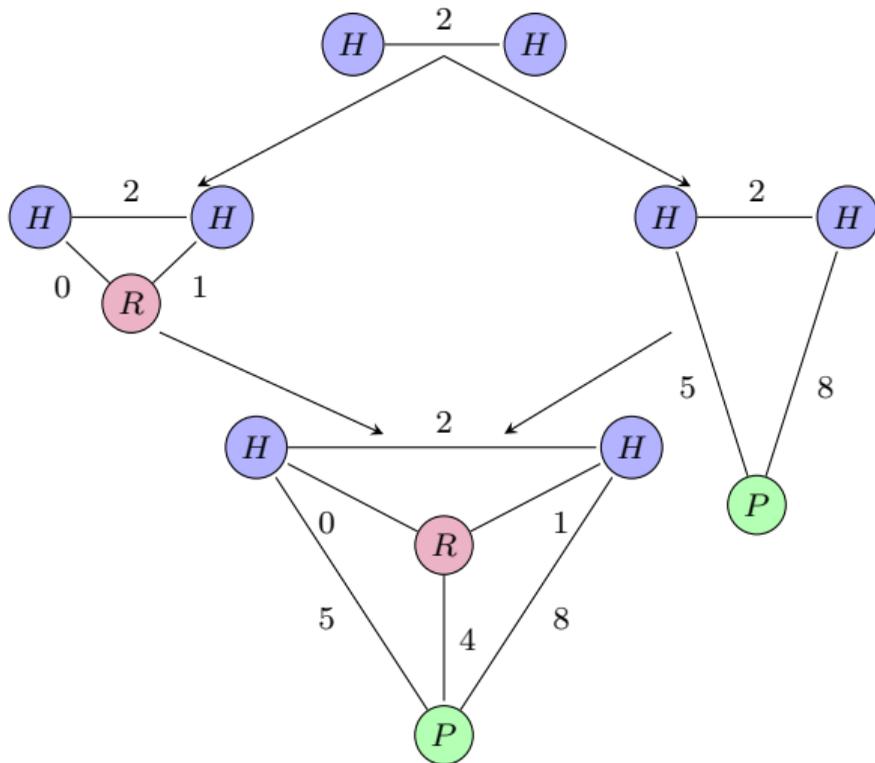
# Des molécules aux pharmacophores



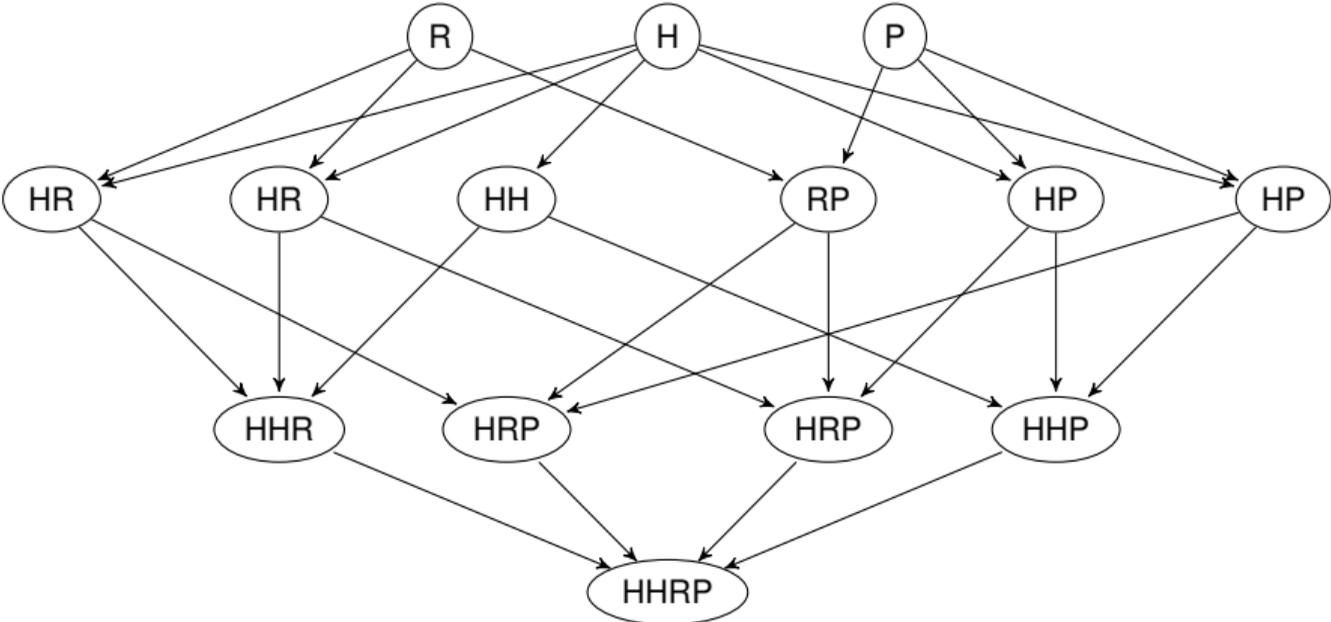
# Des molécules aux pharmacophores



# Pharmacophores abstraction



# Structuration des pharmacophores



1. Sujet et enjeux

2. Interactivité et adaptation

3. Oracle et expériences

4. Conclusion

## Actions souhaitées

- ▶ Découpe de zones de sélection et d'exclusion.
- ▶ Utilisation des pharmacophores comme descripteur de l'espace des solutions.

## Réponses et conséquences

Oui, je veux explorer la famille de ce motif	=	+pds et sélection
Oui, je trouve ce motif intéressant	=	+pds
Je ne sais pas quoi en penser	=	rien
Non, je ne trouve pas ce motif intéressant	=	-pds
Non, je ne veux pas explorer la famille de ce motif	=	-pds et exclusion

## Description :

Les **couches du DAG** (Directed Acyclic graph) sont découpées selon **l'ordre des pharmacophores** en partant de **l'ordre minimal vers l'ordre maximal**.

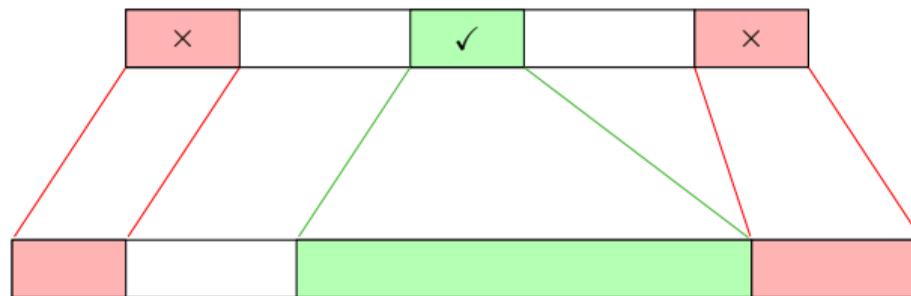
Pharmacophores d'ordre  $k$

Pharmacophores d'ordre  $k + 1$

Pharmacophores d'ordre  $k + 2$

## Exemple :

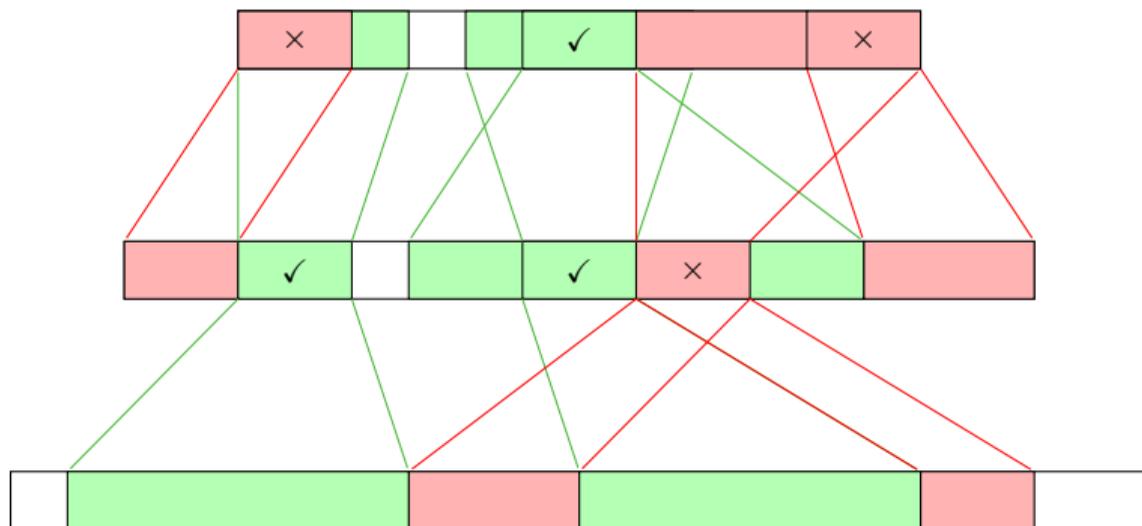
Trois propositions sont soumises à l'expert. Il **valide une proposition** et en **rejette deux**. On découpe donc **une zone d'intérêt** et **deux zones d'exclusion**.



## Exemple :

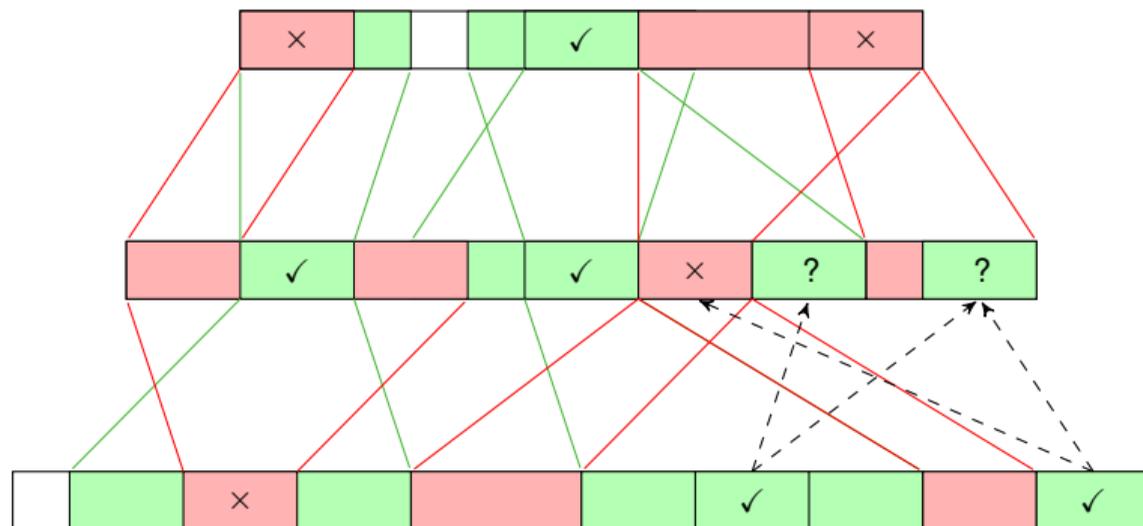
On **descend d'une couche** avec **deux propositions dans la zone d'intérêt** et **une dans la zone neutre**.

Les **zones d'intérêt** et **d'exclusion** peuvent se **superposées**.



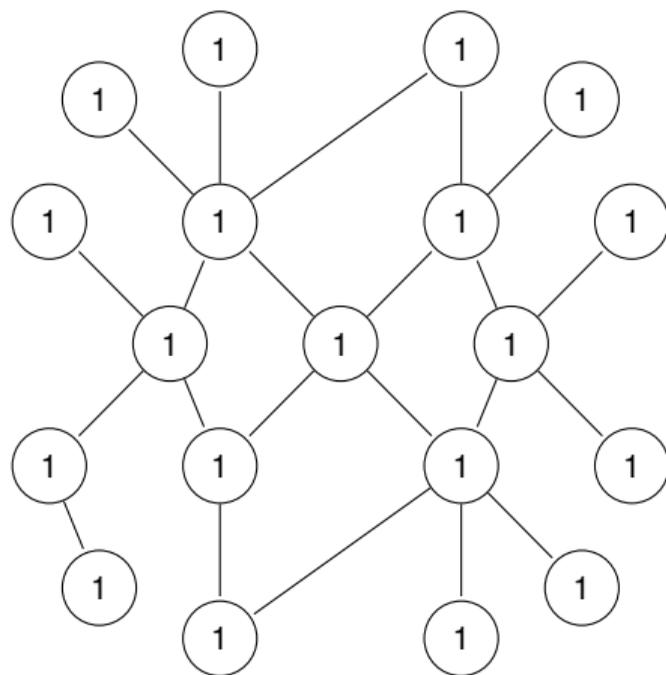
## Description :

On est à la **couche d'ordre maximale** du DAG. On va donc **remonter** itérativement vers les **couches d'ordre inférieur** tout en offrant des propositions à l'expert.



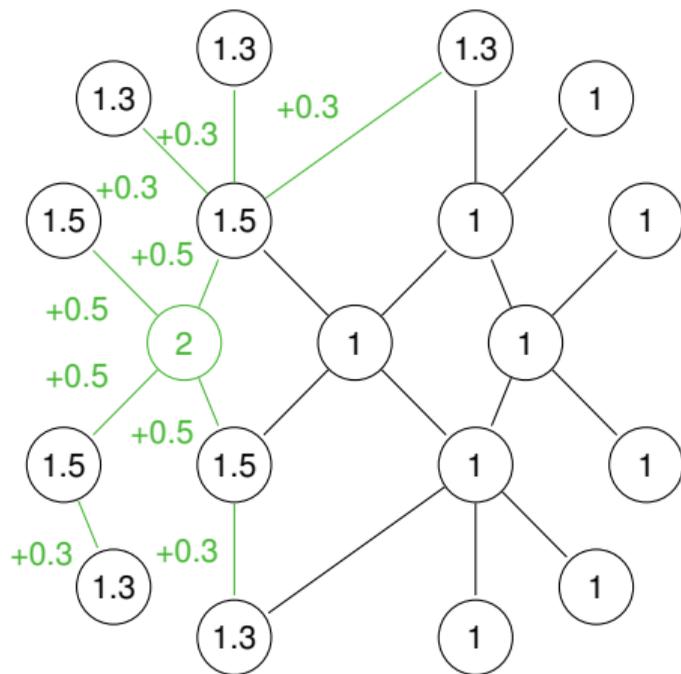
## Description :

Le DAG n'a **pas encore** été **affecté** par les retours de l'expert. Les **poids** des pharamacophores sont donc encore **neutres** (égaux à 1).



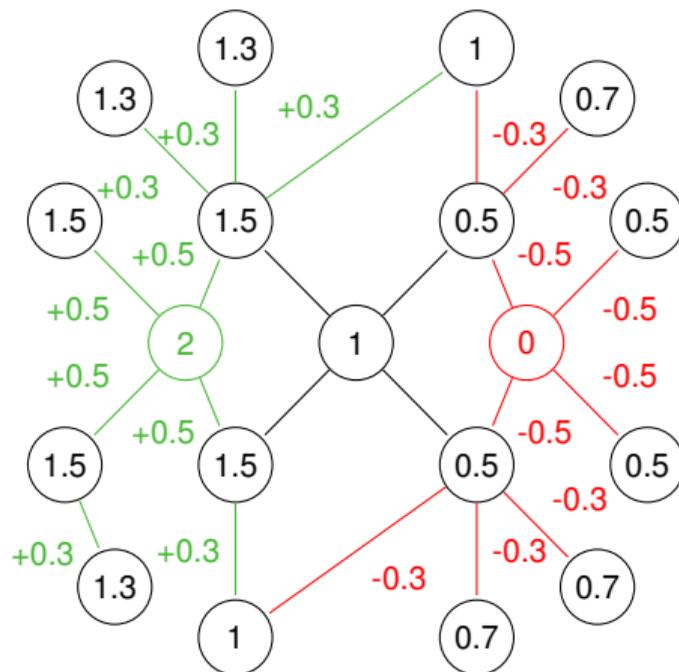
## Exemple :

L'expert a donné un **retour positif**. On **augmente donc le poids** du pharmacophore proposé de 1 et les **poids de sa ligné** (ensemble des ancêtres et descendants) avec un **impact décroissant**.



## Exemple :

L'expert a donné un **retour négatif**. On **diminue donc le poids** du pharmacophore proposé de 1 et les **poids de sa ligné** (ensemble des ancêtres et descendants) avec un **impact décroissant**.



## Fonction d'évaluation

$$fa(X) = \sum_{i=1}^k (Weight(L_i^+) + Weight(L_i^?) + (Weight(L_i^*) - |Weight(L_i^+) - Weight(L_i^-)|)) * (k - i + 1)$$

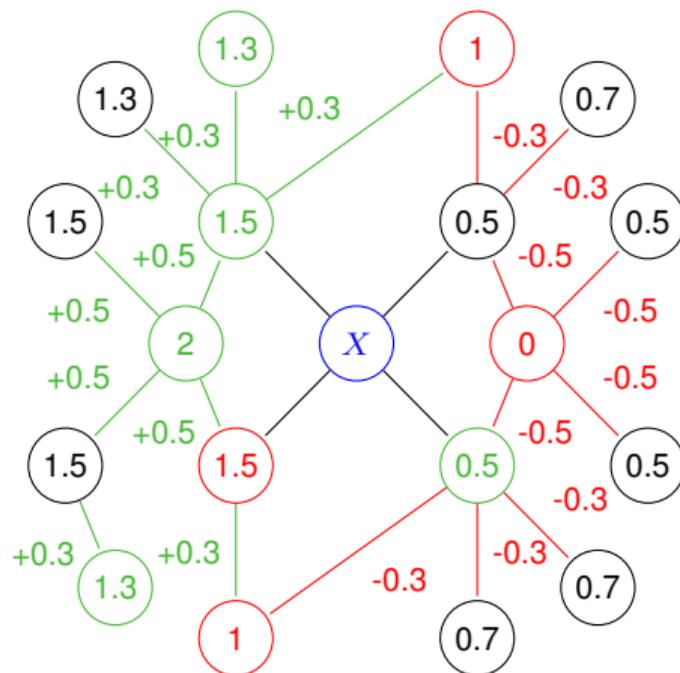
## Légende

- ▶  $L_i^+$  est l'ensemble des éléments **positifs** à la profondeur  $i$ .
- ▶  $L_i^-$  est l'ensemble des éléments **négatifs** à la profondeur  $i$ .
- ▶  $L_i^?$  est l'ensemble des éléments non visités à la profondeur  $i$ .
- ▶  $L_i^*$  est l'ensemble des éléments **positifs**, **négatifs** ou **neutres** à la profondeur  $i$ .

## Décomposition des influences

Type	Prof. 1	Prof. 2
$Weight(L_i^+)$	2	1.3
$Weight(L_i^?)$	0.5	3.4
$Weight(L_i^-)$	1.5	2
$Weight(L_i^*)$	3.5	3.3

On **décompose** l'influence des pharmacophores sur l'évaluation en fonction de la **profondeur relative** (Prof.1, Prof.2) et du type de **retour de l'expert** ( $L_i^+$ ,  $L_i^-$ , etc...).

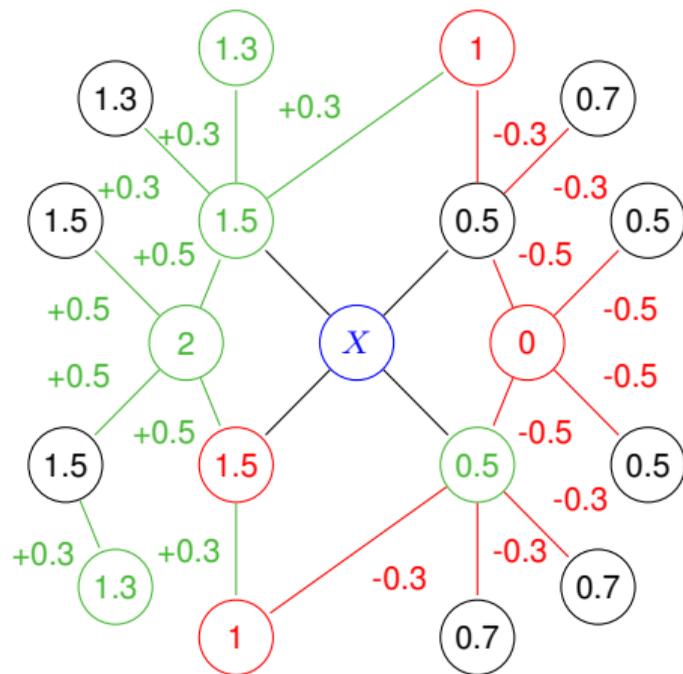


## Décomposition des influences

Type	Prof. 1	Prof. 2
$Weight(L_i^+)$	2	1.3
$Weight(L_i^?)$	0.5	3.4
$Weight(L_i^-)$	1.5	2
$Weight(L_i^*)$	3.5	3.3

## Évaluation du sommet

$$fa(X) = 2 + 0.5 + (3.5 - |2 - 1.5|) * 2 \\ + 1.3 + 3.4 + (3.3 - |1.3 - 2|) * 1$$

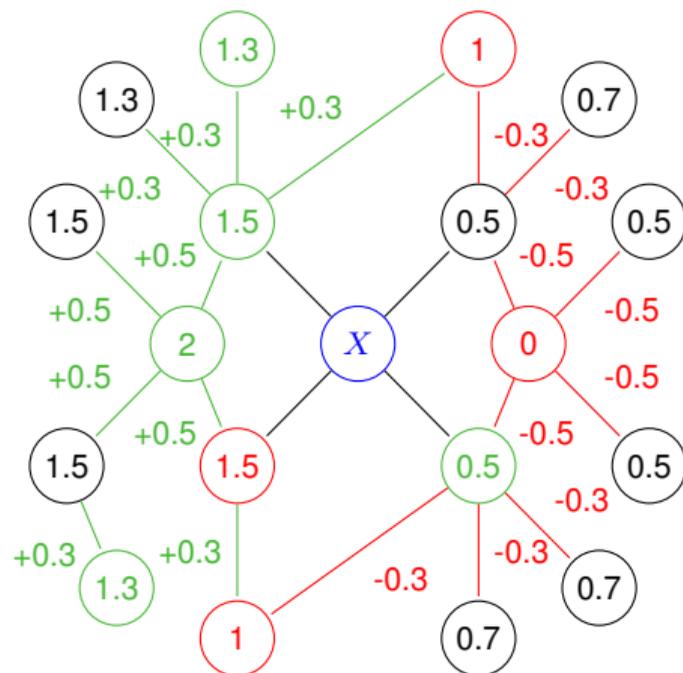


## Décomposition des influences

Type	Prof. 1	Prof. 2
$Weight(L_i^+)$	2	1.3
$Weight(L_i^?)$	0.5	3.4
$Weight(L_i^-)$	1.5	2
$Weight(L_i^*)$	3.5	3.3

## Évaluation du sommet

$$fa(X) = 11 + 7.3$$



Avec  $fa(X_2) < fa(X_1) < fa(X_4) < fa(X_3)$



Zone sélectionnée ( $k$  tirages)



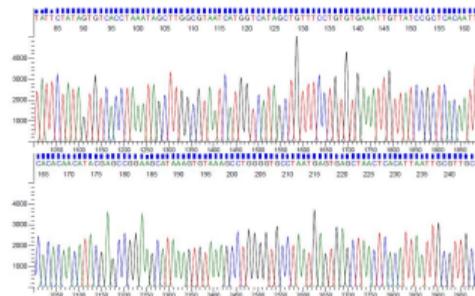
Zone inconnue ( $d$  tirages)

## Explication :

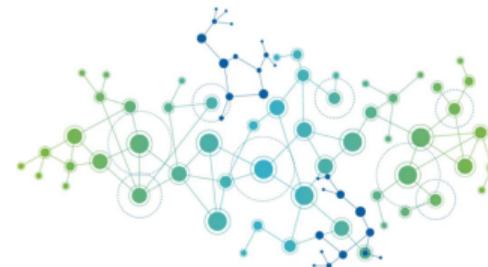
Plus  $fa(X_i)$  est grand, plus  $X_i$  aura de chance d'être choisit. On peut donc passer de la valeur de  $fa(X_i)$  à un pourcentage de sélection dépendant la zone.

	Support	Frequent	Gen	Closed	Max
6	t1, t2, t3, t4, t5, t6	c	c	c	c
5	t1, t2, t3, t4, t5	e, ce	e	ce	
4	t1, t3, t4, t5	a, ac, ae, ace	a	ace	
4	t1, t3, t5, t6	b, bc	b	bc	
4	t2, t4, t5, t6	d, cd	d	cd	
3	t1, t3, t5	ab, abc, abe	ab	abce	abce
3	t2, t4, t5	be, bce, abce	be	abce	abce
3	t2, t4, t5	de, cde	de	cde	cde

Itemsets



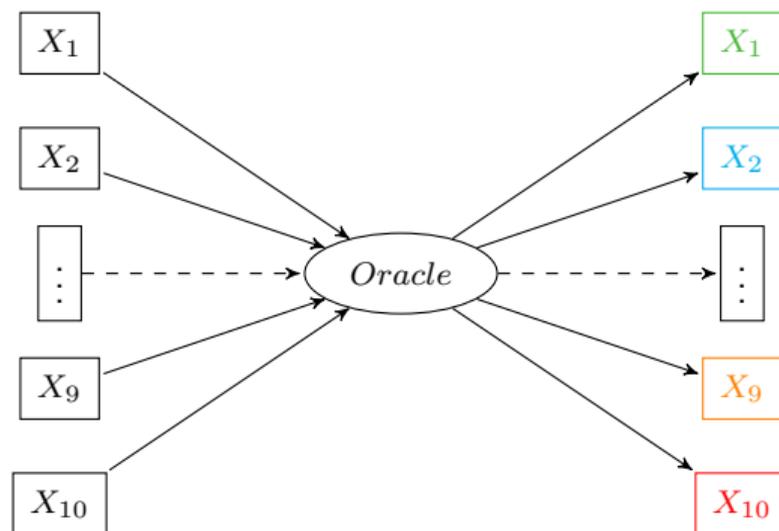
Séquences



Graphes

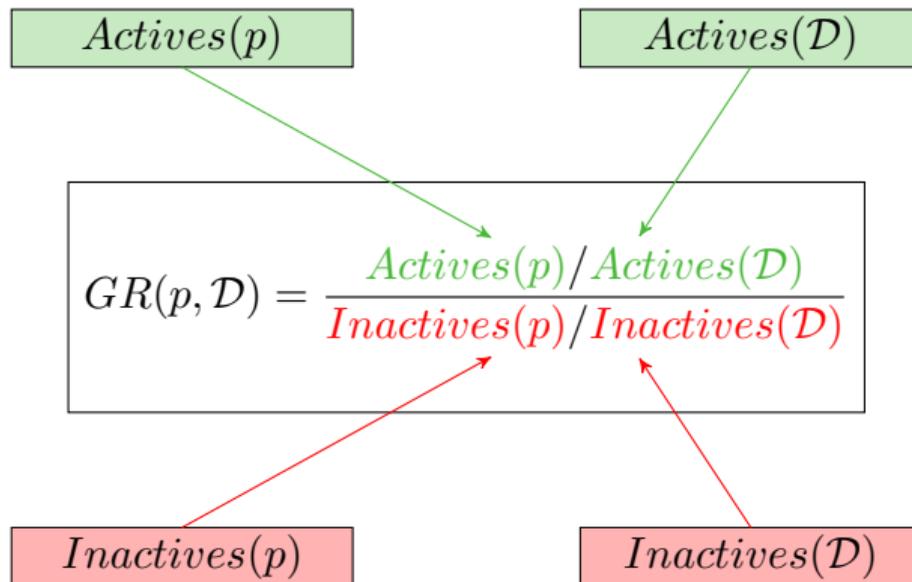
# Plan de l'exposé

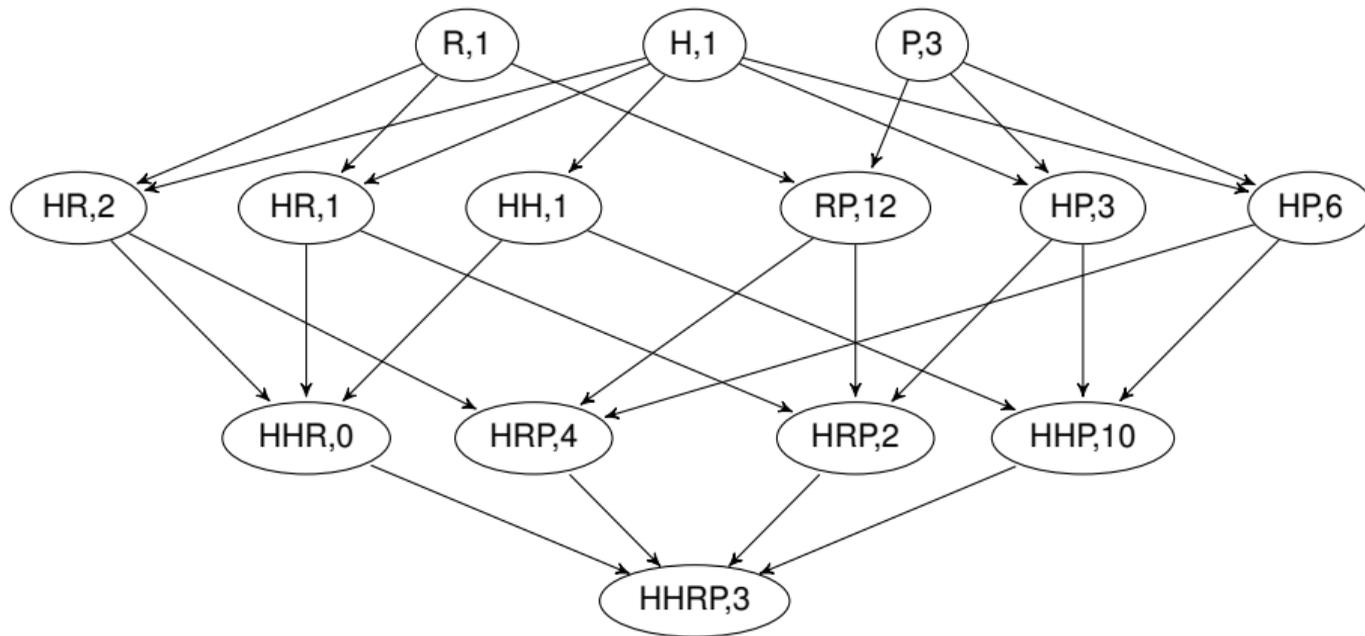
1. Sujet et enjeux
2. Interactivité et adaptation
- 3. Oracle et expériences**
4. Conclusion



## Mesures existantes :

- ▶ Growth Rate (taux de croissance)
- ▶ Confidence (Confiance)
- ▶ WRAcc
- ▶ ChiSquare
- ▶ ...





# Composition des couches

Rejected  $\approx$  5%

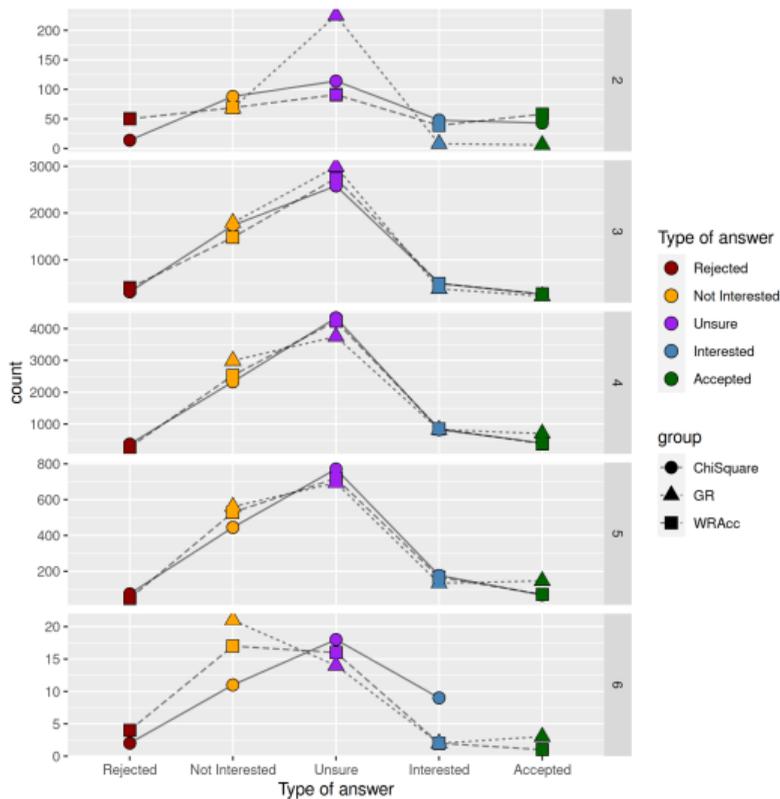
Not interested  $\approx$  30%

Unsure  $\approx$  50%

Interested  $\approx$  10%

Accepted  $\approx$  5%

*Pourcentage sur la représentation ordonnée des qualités.*



Échantillon(S)



$$S_A = 40\%$$

$$E_A = 10\%$$



Couche du DAG(E)

## Formule du Lift

$$Lift(S, E)_A = \frac{|S_A|/|S|}{|E_A|/|E|} \quad (1)$$

$$= \frac{40\%}{10\%} \quad (2)$$

$$= 4 \quad (3)$$

# Heatmap WRFs avec WRAcc

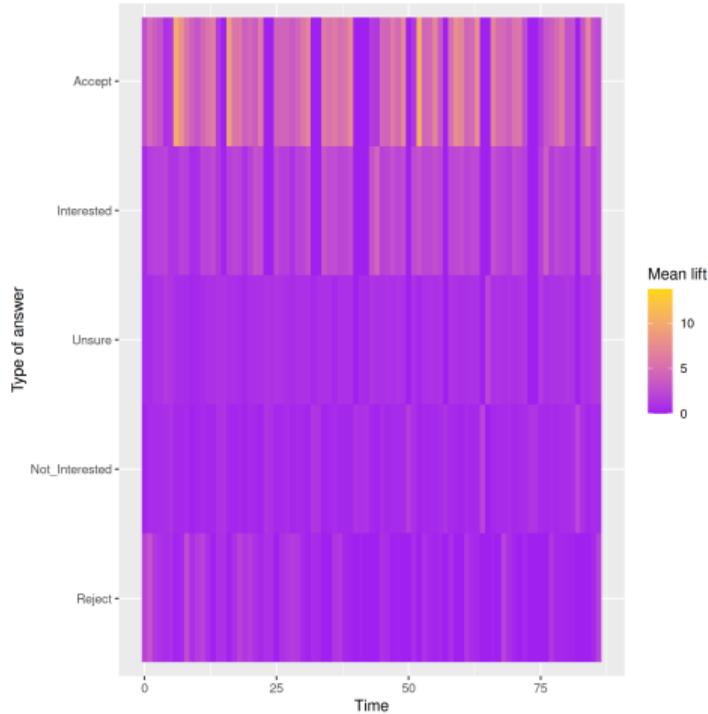


FIGURE – WRFS dirigé

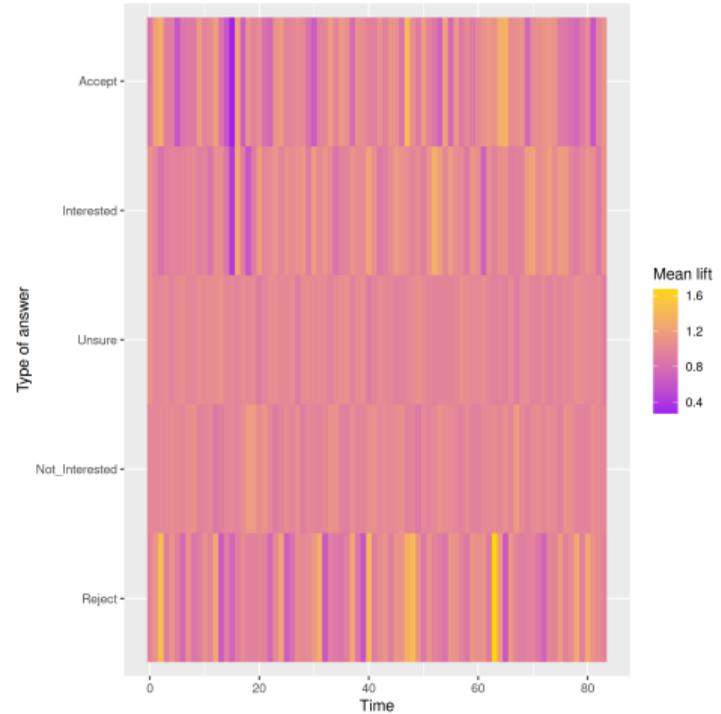


FIGURE – WRFS aléatoire

# Heatmap WRFs avec WRAcc

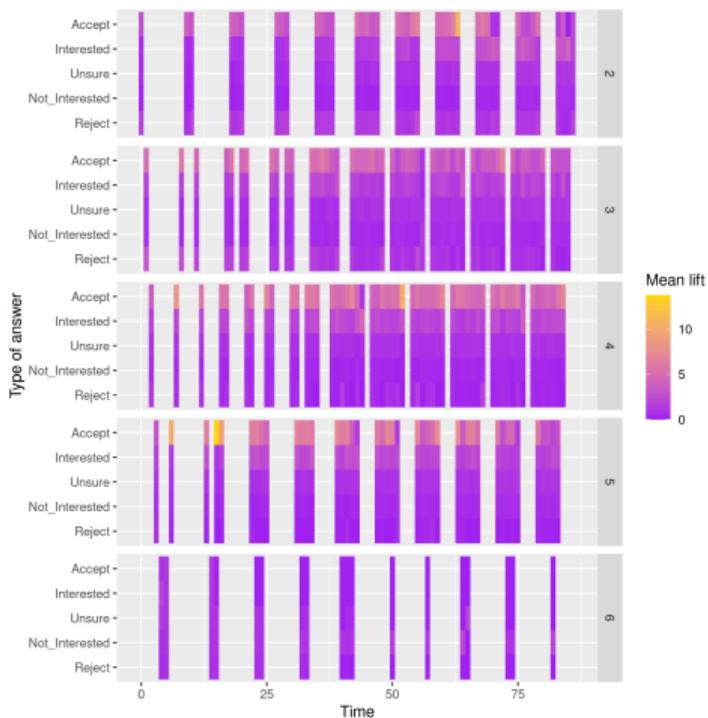


FIGURE – WRFs dirigé

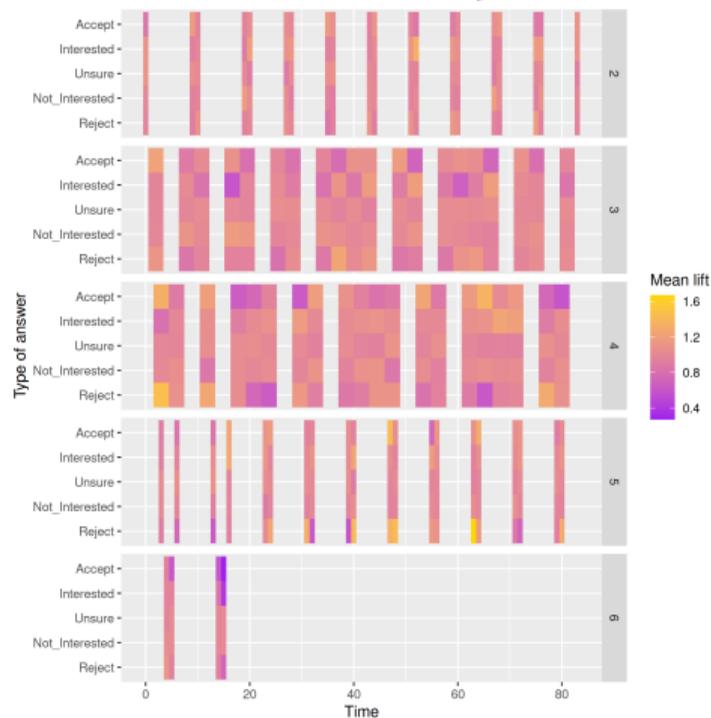


FIGURE – WRFs aléatoire

# Heatmap WRFs avec WRAcc

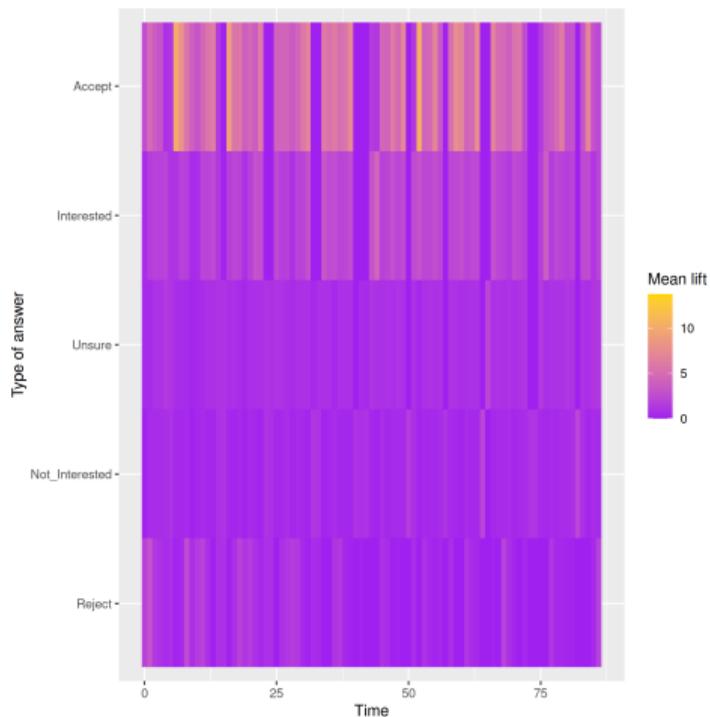


FIGURE – WRFs avec évaluation

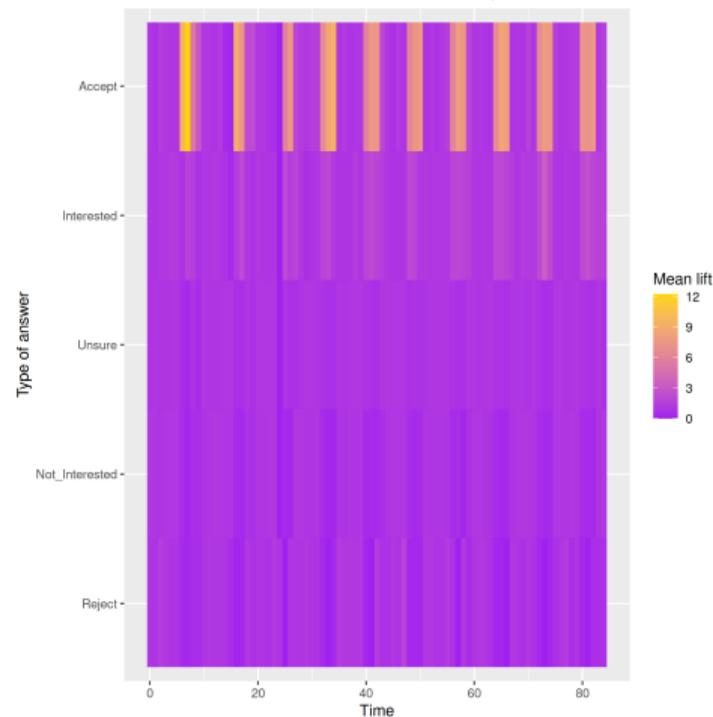


FIGURE – WRFs sans évaluation

# Heatmap WRFs avec WRAcc

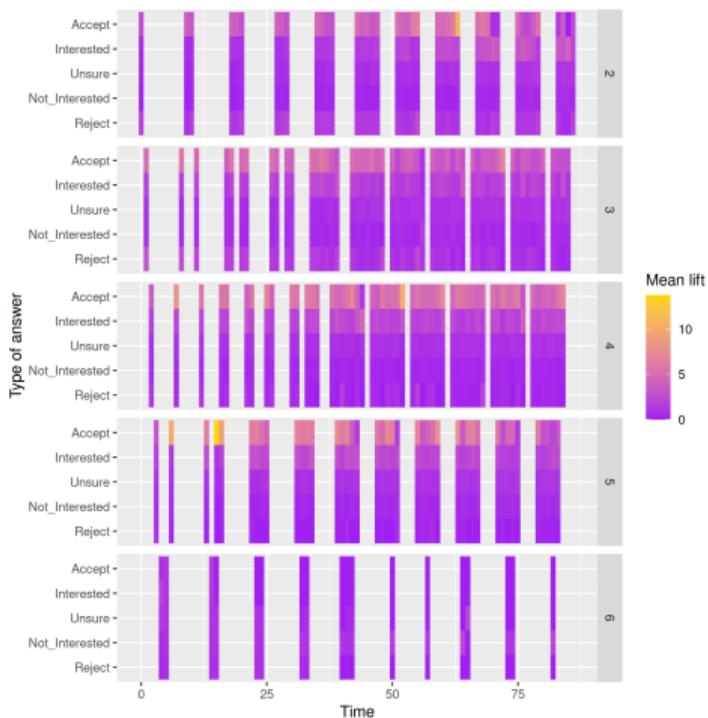


FIGURE – WRFs avec évaluation

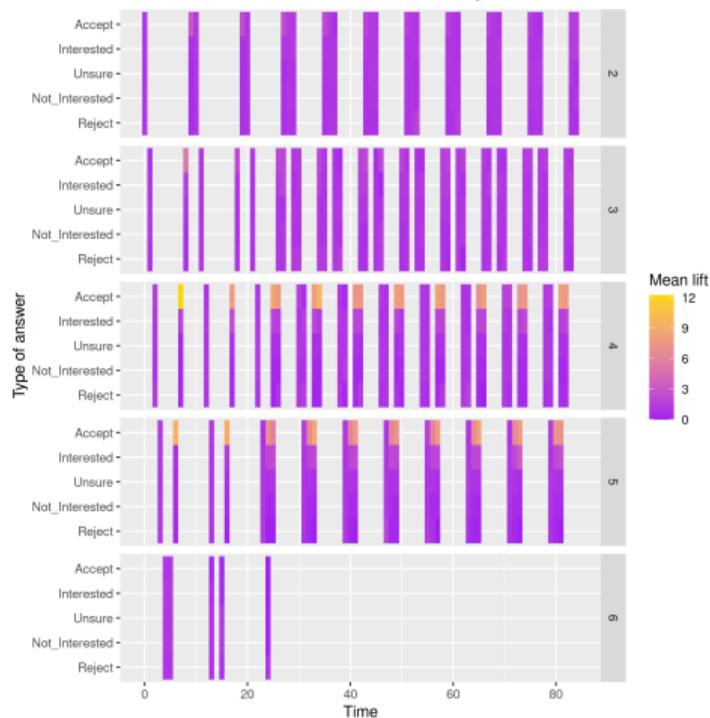


FIGURE – WRFs sans évaluation

# Heatmap WRFS avec WRAcc

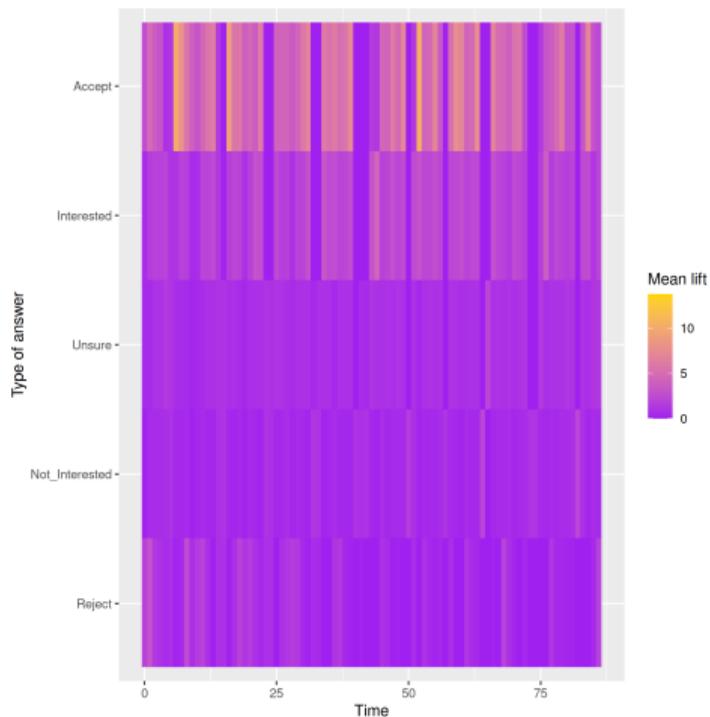


FIGURE – WRFS avec sélection

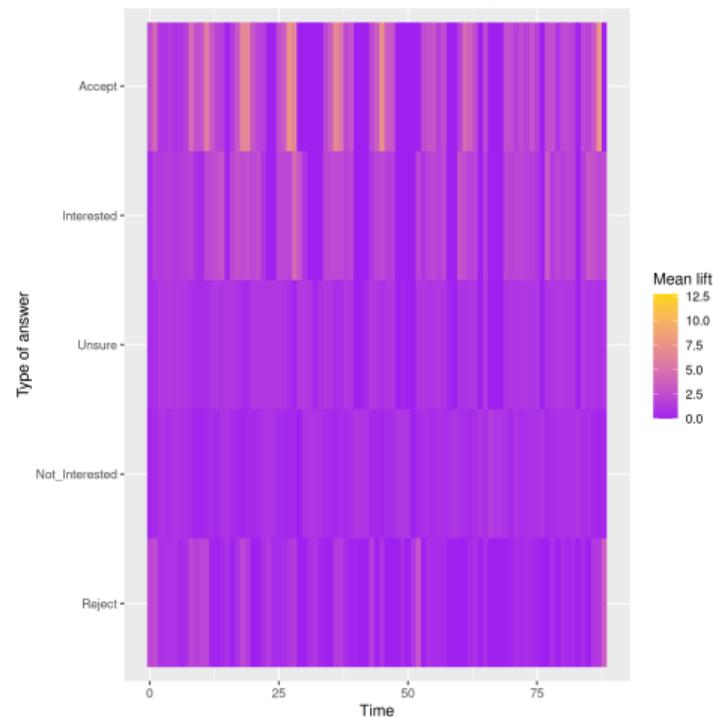


FIGURE – WRFS sans sélection

# Heatmap WRFs avec WRAcc

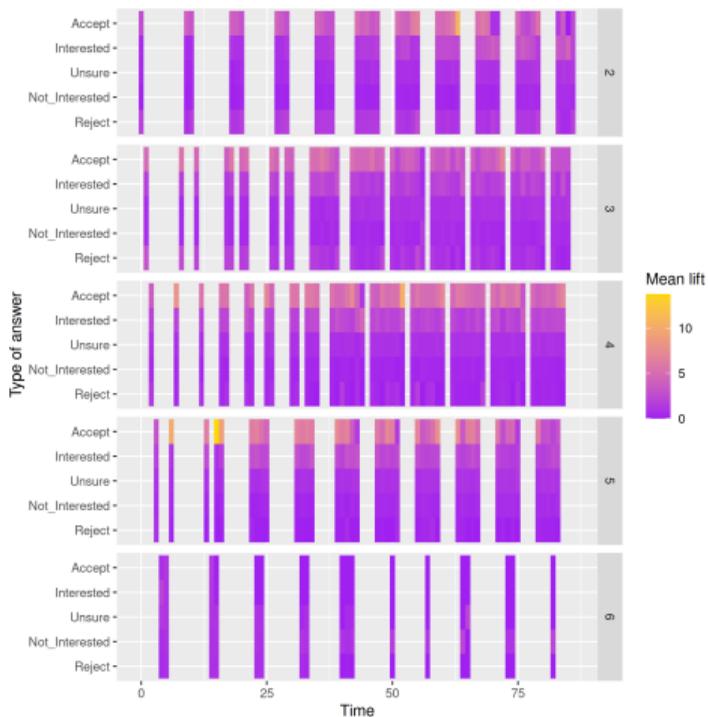


FIGURE – WRFs avec sélection

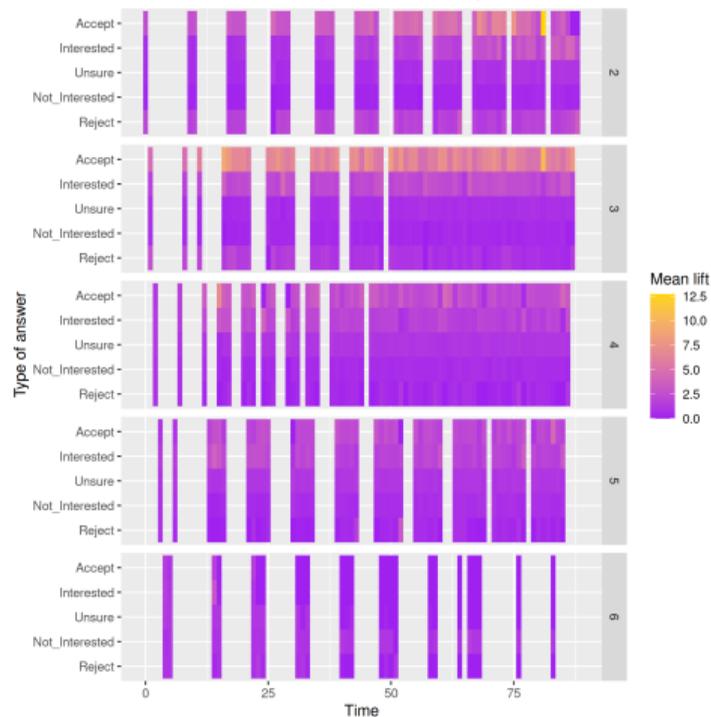


FIGURE – WRFs sans sélection

# Heatmap wrfs vs pog

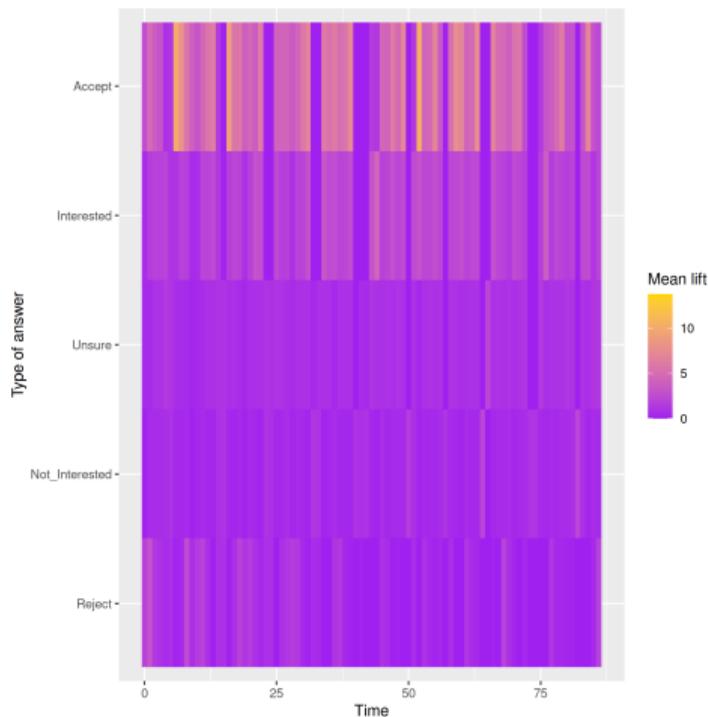


FIGURE – WRFS avec WRAcc

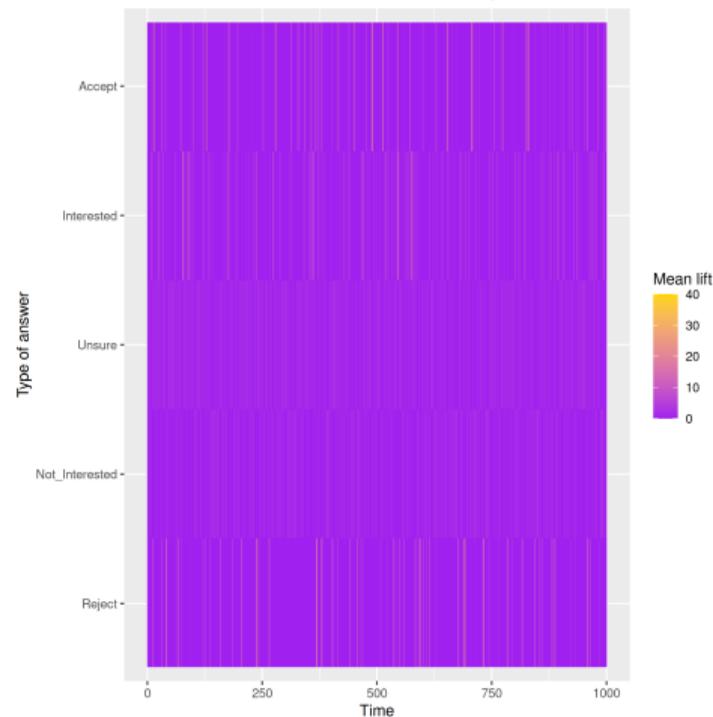


FIGURE – POG avec WRAcc

# Heatmap wrfs vs pog

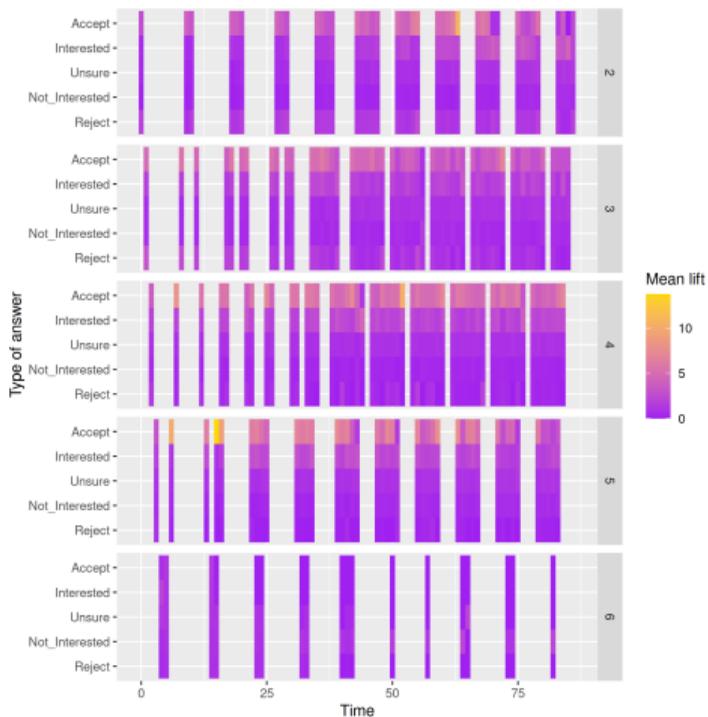


FIGURE – WRFS avec WRAcc

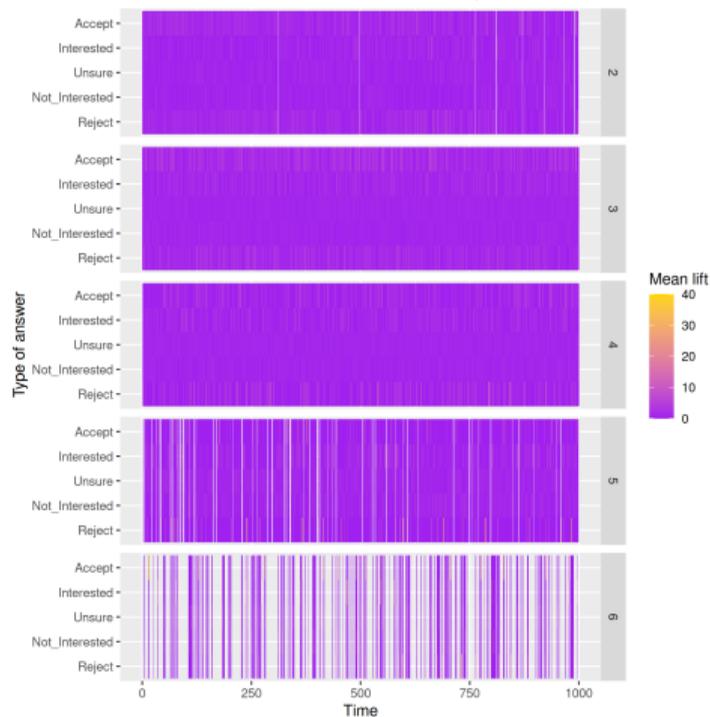


FIGURE – POG avec WRAcc

# Area wrfs vs pog : 200 premières questions

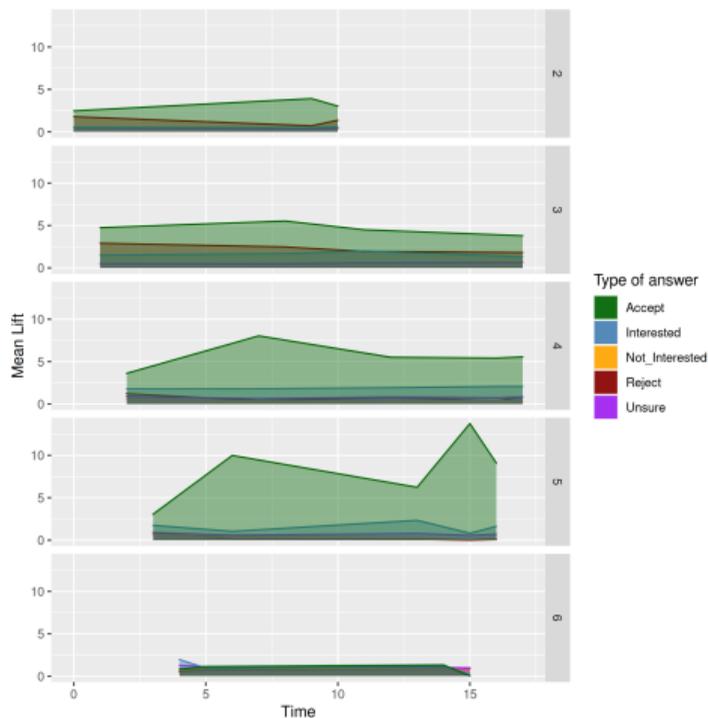


FIGURE – WRFS avec WRAcc

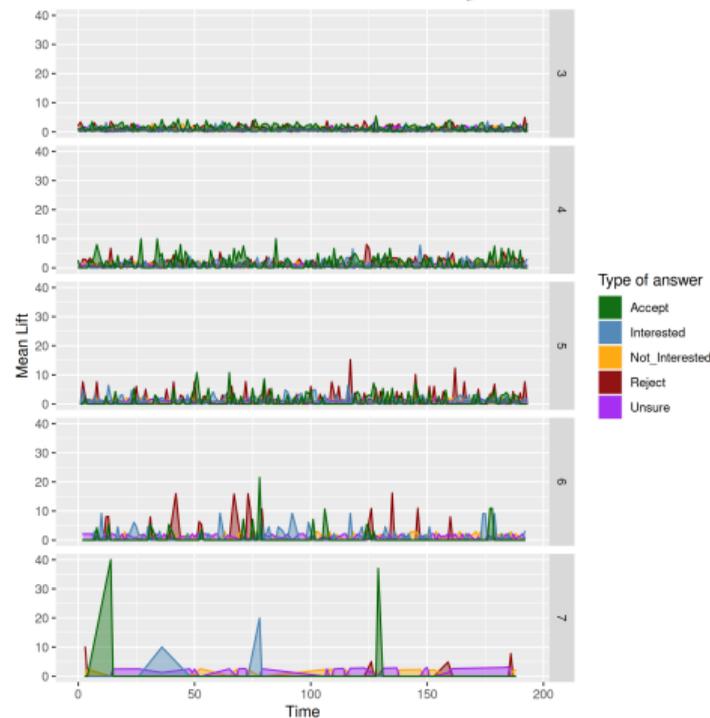
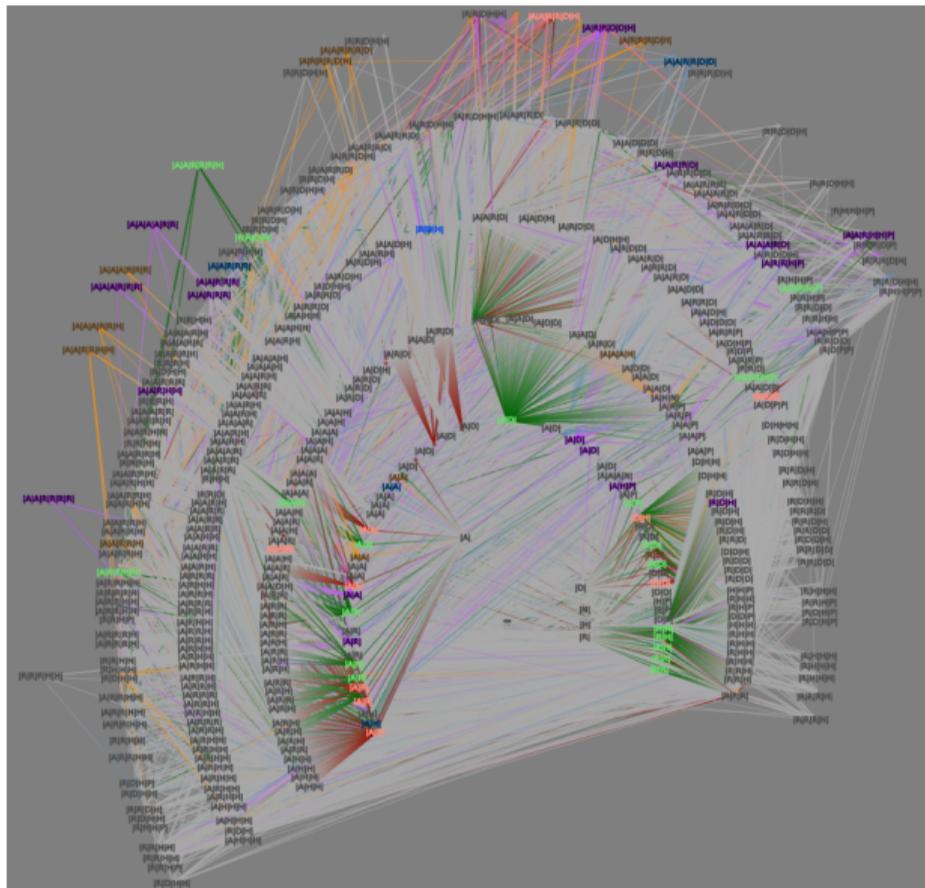
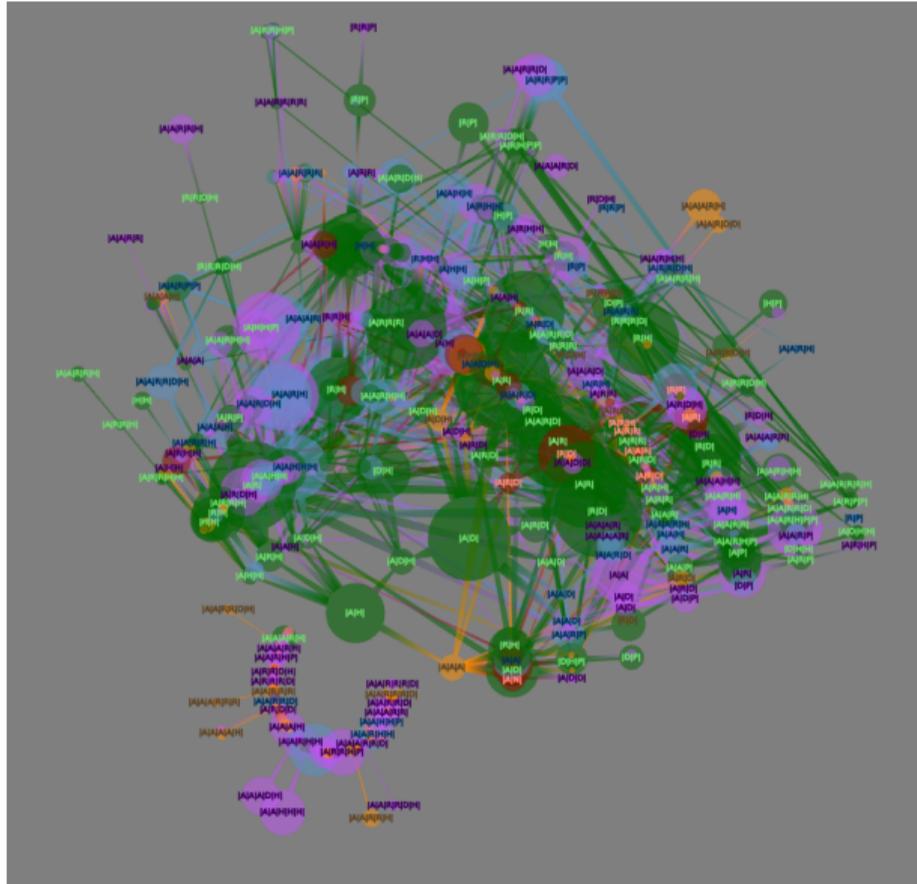


FIGURE – POG avec WRAcc

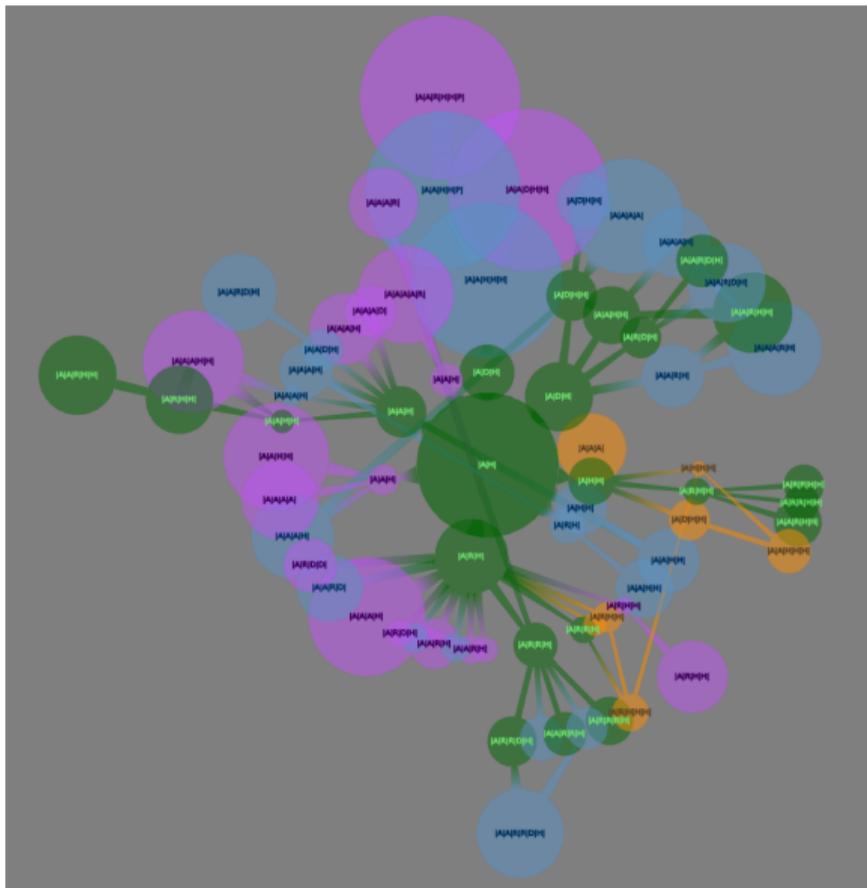
# BCR-ABL network with WRFs and WRAcc



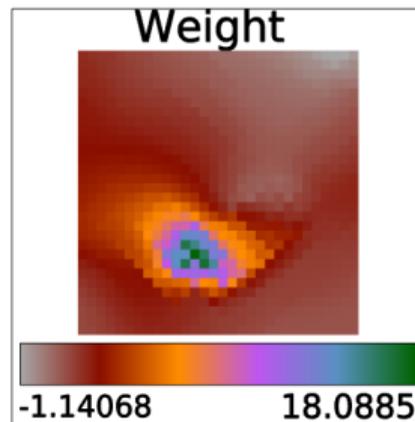
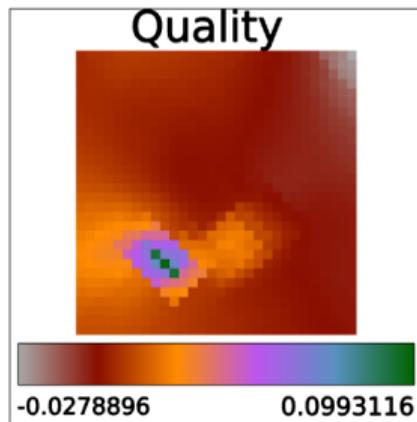
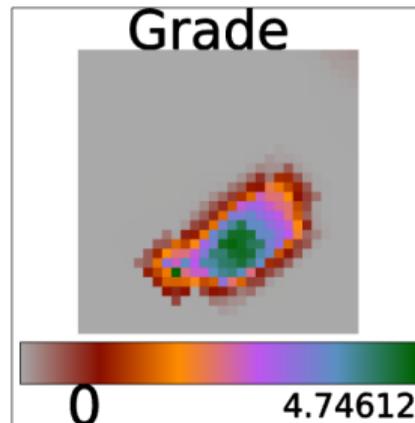
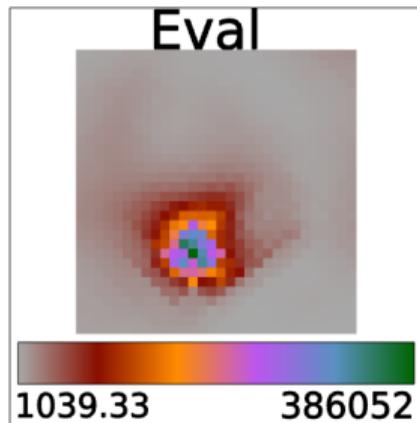
# BCR-ABL visited vertices



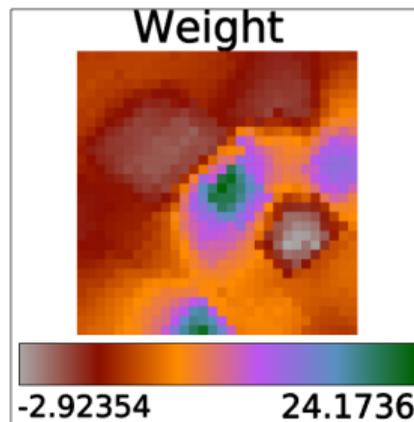
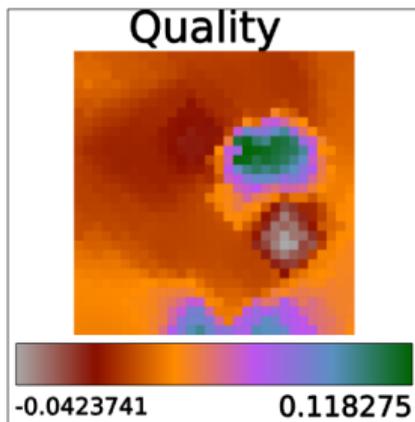
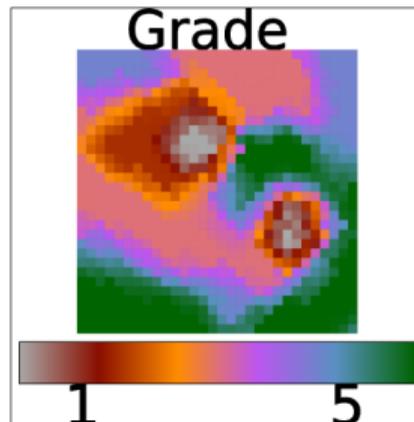
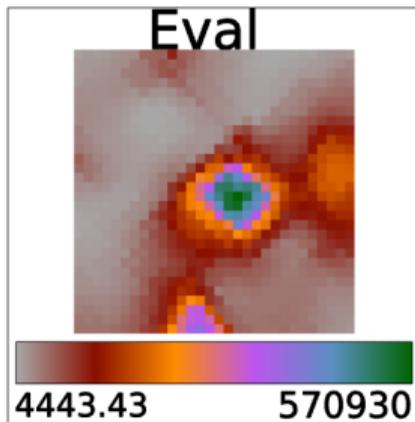
# BCR-ABL visited vertices from AH



# BCR-ABL SOM overview



# BCR-ABL SOM visited



1. Sujet et enjeux
2. Interactivité et adaptation
3. Oracle et expériences
4. Conclusion

## Avancés

- ▶ L'exploitation de zone de prédilection.
- ▶ Une propagation de l'intérêts utilisateur (poids).
- ▶ L'exploitations des poids transformant les motifs en descripteurs.

## Possibilités

- ▶ Approfondissement de la fonction d'acceptation.
- ▶ Mise au point d'un Oracle plus complexe.
- ▶ Développement d'une interface utilisateur pour permettre le début des tests expert.

# Run-times

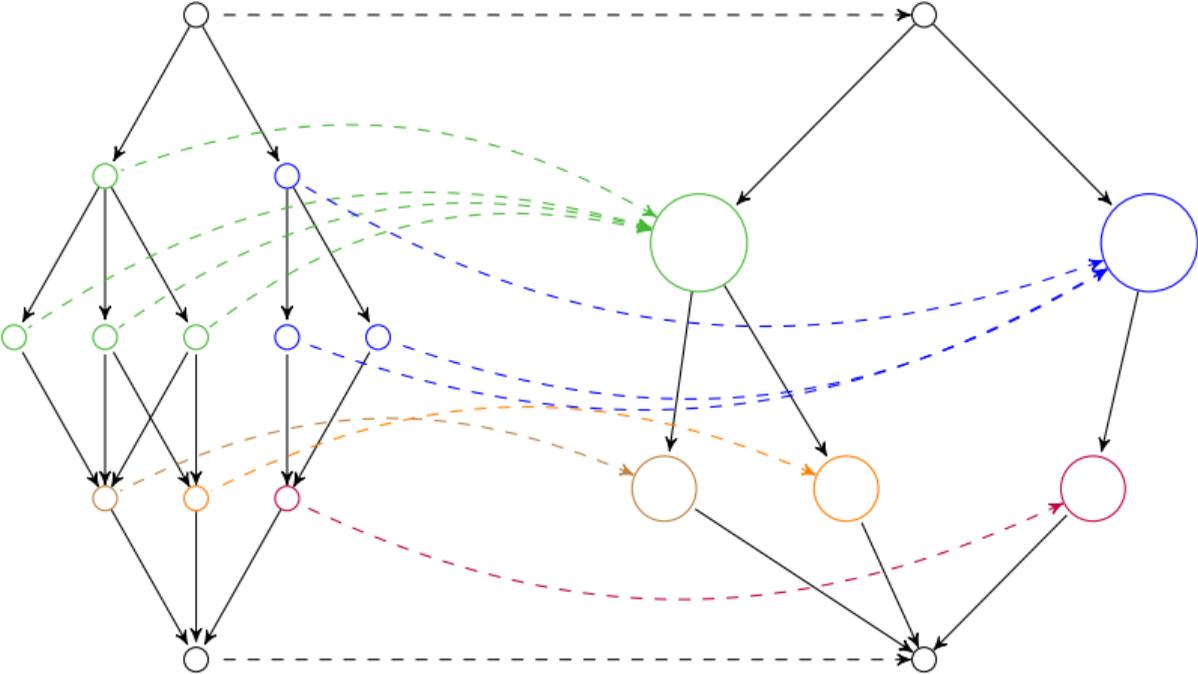
## Configuration

8 cores : 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz

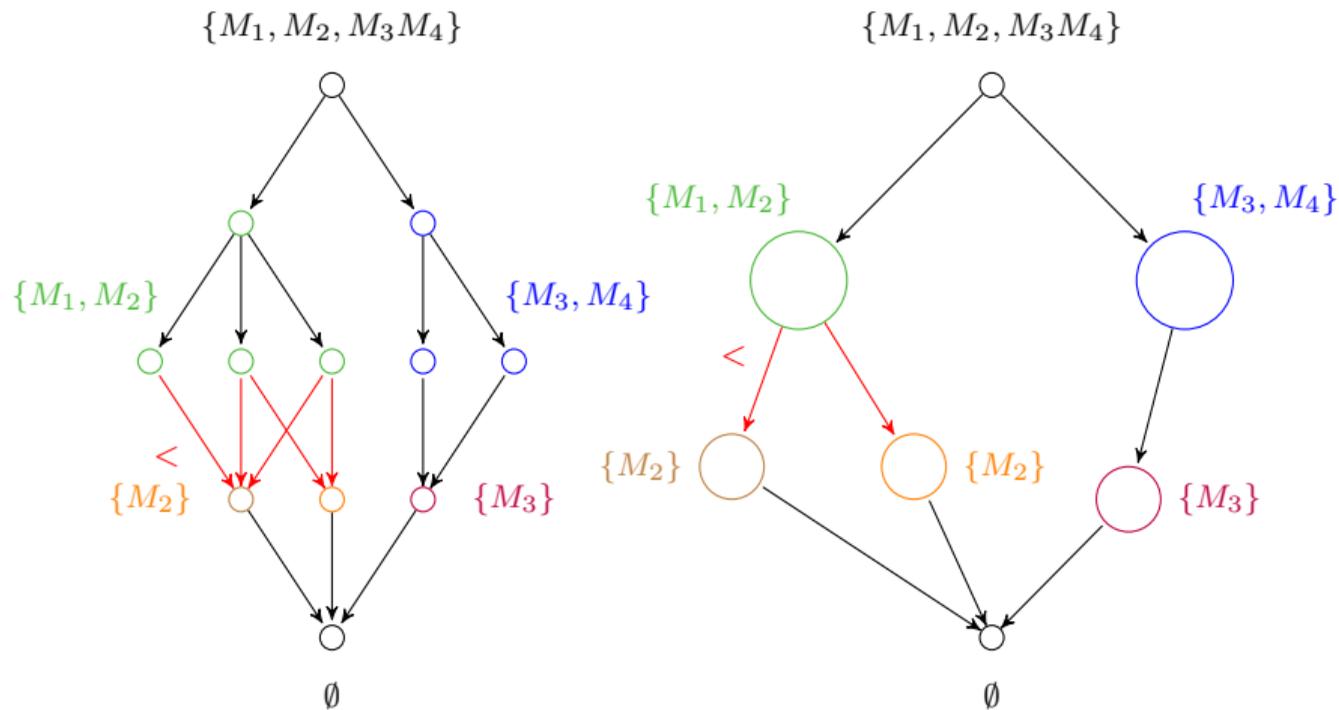
## Run-times

- ▶ Creating DAG : under 4 minutes.
- ▶ Creating equivalence classes : under 4 minutes (with 4 thread).
- ▶ Applying quality measure : few seconds.
- ▶ Applying WRFS for 1000 questions : less than a minute.

# Defining Structured Equivalence Classes



# Defining Structured Equivalence Classes



# References

- ▶ J.-P. METIVIER, B. CUISSART, R. BUREAU, and A. LEPAILLEUR. The pharmacophore network: A computational method for exploring structure-activity relationships from a large chemical data set.
- ▶ [https://www.cms.waikato.ac.nz/~abifet/book/chapter\\_10.html](https://www.cms.waikato.ac.nz/~abifet/book/chapter_10.html)
- ▶ <https://dnacore.missouri.edu/dataevaluation.html>
- ▶ <https://levelup.gitconnected.com/reactive-apis-for-graph-data-vert-x-neo4j-335521321fa3>
- ▶ <https://commons.wikimedia.org/wiki/File:Pharmacophore-5-HT2C.svg>