



HAL
open science

Identification de concepts remarquables à partir de leur voisinage dans un treillis

Etienne Lehembre, Jean-Luc Lamotte, Christophe Couronne, Bruno Crémilleux, Jean-Luc Manguin, Bertrand Cuissart, Abdelkader Ouali, Albrecht Zimmermann

► To cite this version:

Etienne Lehembre, Jean-Luc Lamotte, Christophe Couronne, Bruno Crémilleux, Jean-Luc Manguin, et al.. Identification de concepts remarquables à partir de leur voisinage dans un treillis. Conférence francophone sur l'Extraction et la Gestion des Connaissances, Jan 2022, Blois, France. hal-03764657

HAL Id: hal-03764657

<https://hal.science/hal-03764657v1>

Submitted on 21 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IDENTIFICATION DE CONCEPTS REMARQUABLES À PARTIR DE LEUR VOISINAGE DANS UN TREILLIS

Etienne Lehembre^a, Jean-Luc Lamotte^b, Abdelkader Ouali^a, Christophe Couronne^a, Bruno Crémilleux^a, Jean-Luc Manguin^a, Albrecht Zimmermann^a, Bertrand Cuissart^a

^a : GREYC, CNRS UMR 6072, UNICAEN, Normandie Univ. Caen, France

^b : CERMN, EA 4258 FR CNRS 3038 INC3M SF 4206 ICORE, UNICAEN, Normandie Univ. Caen, France

auteur correspondant : etienne.lehembre@unicaen.fr

Calculer des configuration structurelles clés à partir des ligands est l'un des challenges majeur de la conception pharmaceutique. Dans le projet *Involvd*¹, nous avons conçu une méthode de fouille de pharmacophores afin de découvrir les configurations structurelles pouvant être liée à des activités biologiques.

1. Détails de la méthode

- 1) Définir les **marqueurs pharmacophoriques**.
- 2) Représenter me jeu de données moléculaire par des **graphes pharmacophoriques**.
- 3) Extraction des **pharmacophores** en minant les graphes pharmacophoriques [2] via Norns^a.
- 4) Structuration des pharmacophores en **treillis de concepts** [1].
- 5) Calcul de la qualité des pharmacophores avec le *taux de croissance*
- 6) *Extraction des fratries et sélection des éléments remarquables.*

^a. <https://forge.greyc.fr/projects/norns>

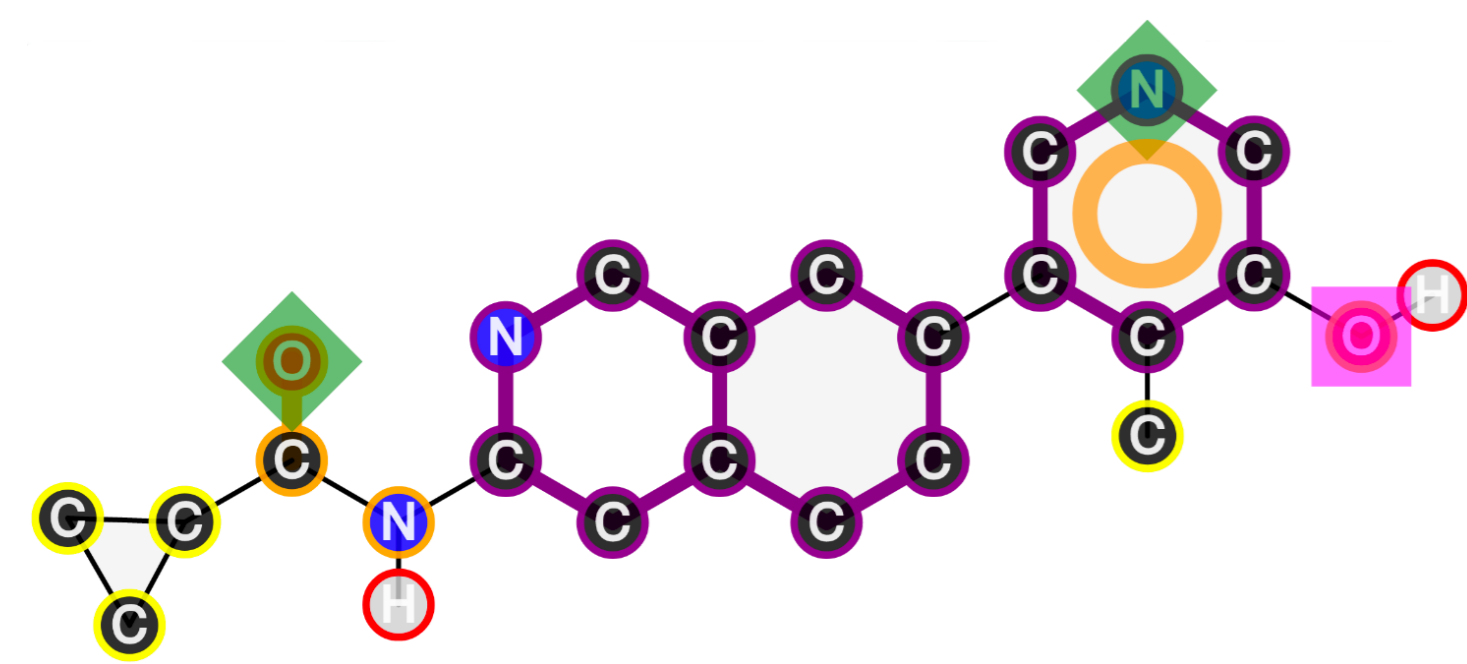


Figure 1: Molécule et son graphe pharmacophorique

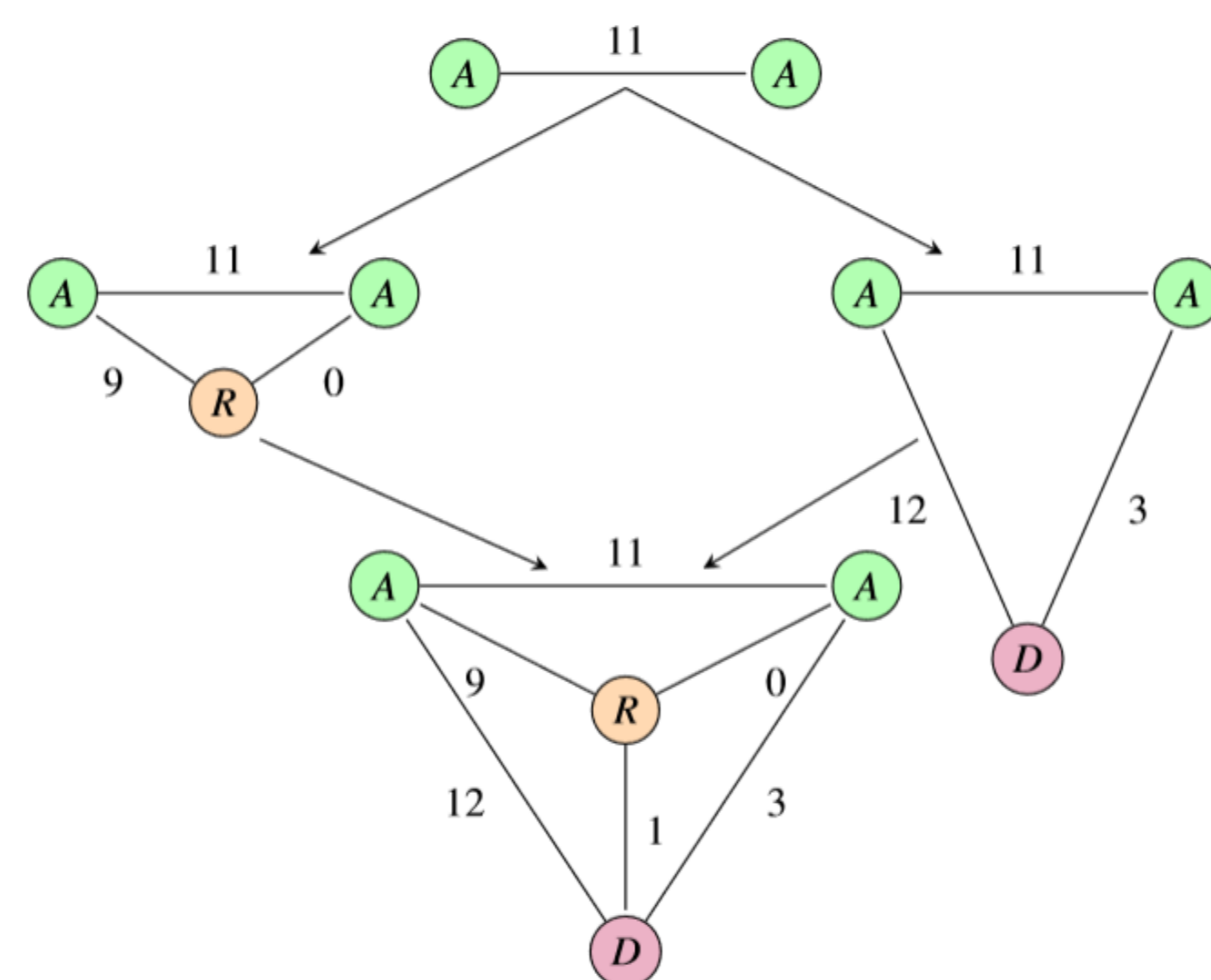


Figure 2: Inclusion des pharmacophores

2. Du treillis des pharmacophores au treillis condensé

Afin de structurer mes pharmacophores fréquents composant les molécules, nous utilisons un **treillis**.

- 1) Relation d'ordre : inclusion de sous-graphes.
- 2) Objet / Attributs : pharmacophores / molécules.

Nous condensions alors le treillis en construisant des classes d'équivalence qui regroupe les pharmacophores apparaissant dans les me molécules dès qu'ils constituent un sous-treillis connecté.

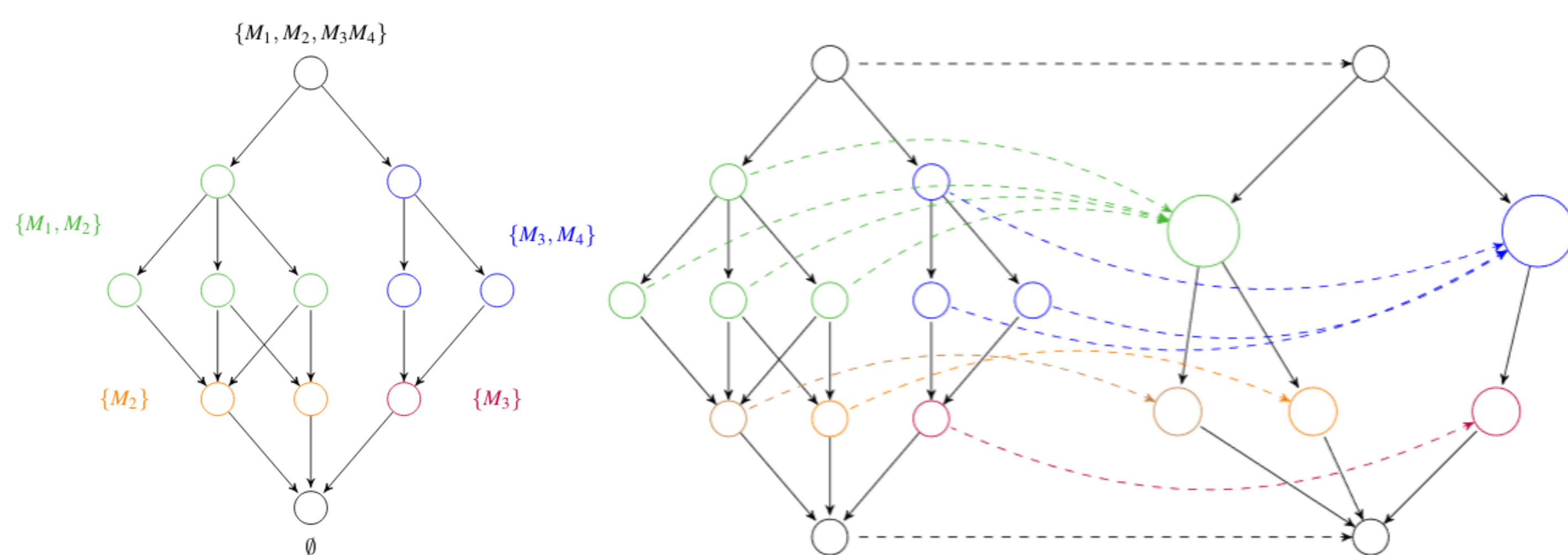


Figure 3: Treillis de Pharmacophores (gauche) et son treillis condensé (droit)

Dans la figure gauche, les nœuds oranges ont été séparés car :

- 1) Ils ne sont pas connecté par un parent ou un enfant.
- 2) L'un n'est pas le parent ou l'enfant de l'autre.

3. Une nouvelle mesure de contraste

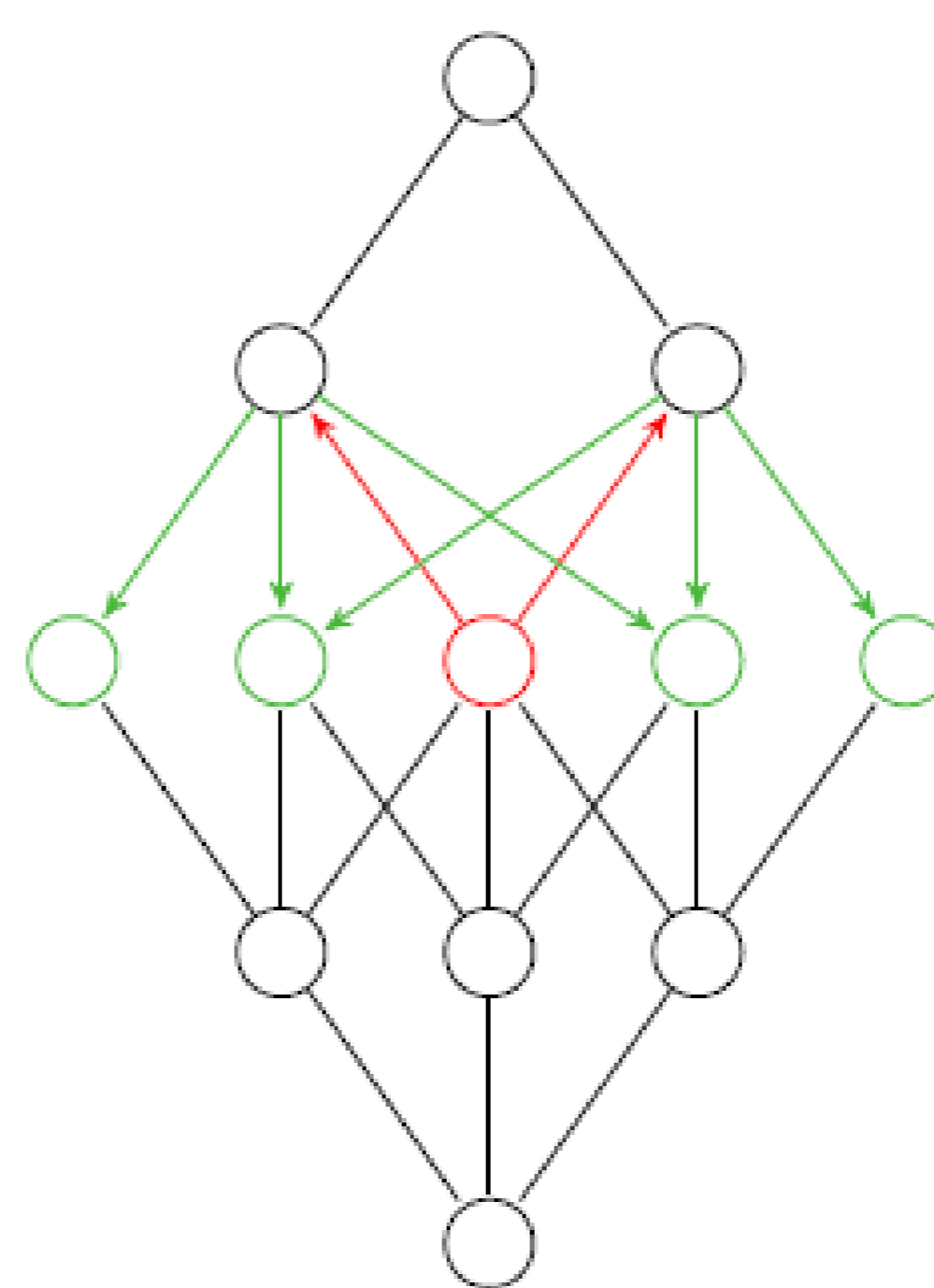


Figure 4: Sibling Misfit Selector

En utilisant le **treillis condensé** et le **taux de croissance des pharmacophores**, on calcul le **Sibling Misfit Selector (SMS)** :

- 1) Sélection d'un nœud du treillis (nœuds rouge).
- 2) Identification des parents (flèches rouges).
- 3) Identifications des frères (nœuds verts).
- 4) Calcul de l'écart-type de la fratrie.
- 5) Sélection du nœud source si il dévie delta fois plus de la moyenne que l'écart-type de sa fratrie.

SMS a été inspiré par la notion de **Activity Cliff** [3]

4. Résultats

Ensemble des molécules : 1,485 ligands de ChEMBL-ABL².
Ensemble des pharmacophores : 112,363 pharmacophores.
Les pharmacophores ont été extraits avec les paramètres suivant :

- 1) Fréquence minimale : 10.
- 2) Ordre : 1 à 7 sommets.

Le SMS a été appliqué sur le treillis condensé avec un seuil minimal de sélection à 2 fois l'écart-type de la fratrie.

	Ordre 1	Ordre 2	Ordre 3	Ordre 4	Ordre 5	Ordre 6	Total
Sélectionné	6	6	208	212	48	12	492
Écarté	0	301	5,181	8,064	1,487	28	15,061
Total	6	307	5,389	8,276	1,535	40	15,553

Table 1: Classes d'équivalence et sélection par SMS

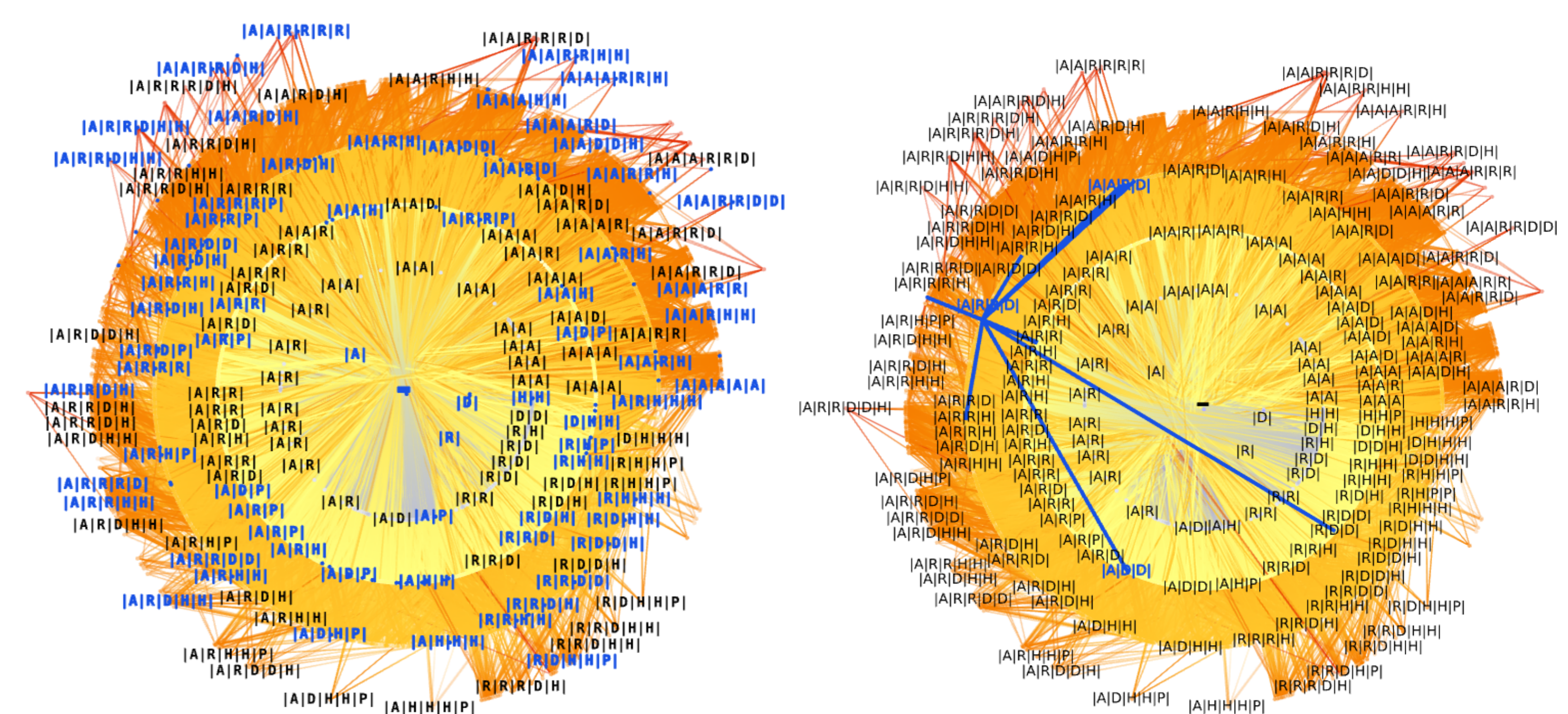


Figure 5: Réseau des classes d'équivalence (gauche) et connections entre les classes d'équivalences (droite)

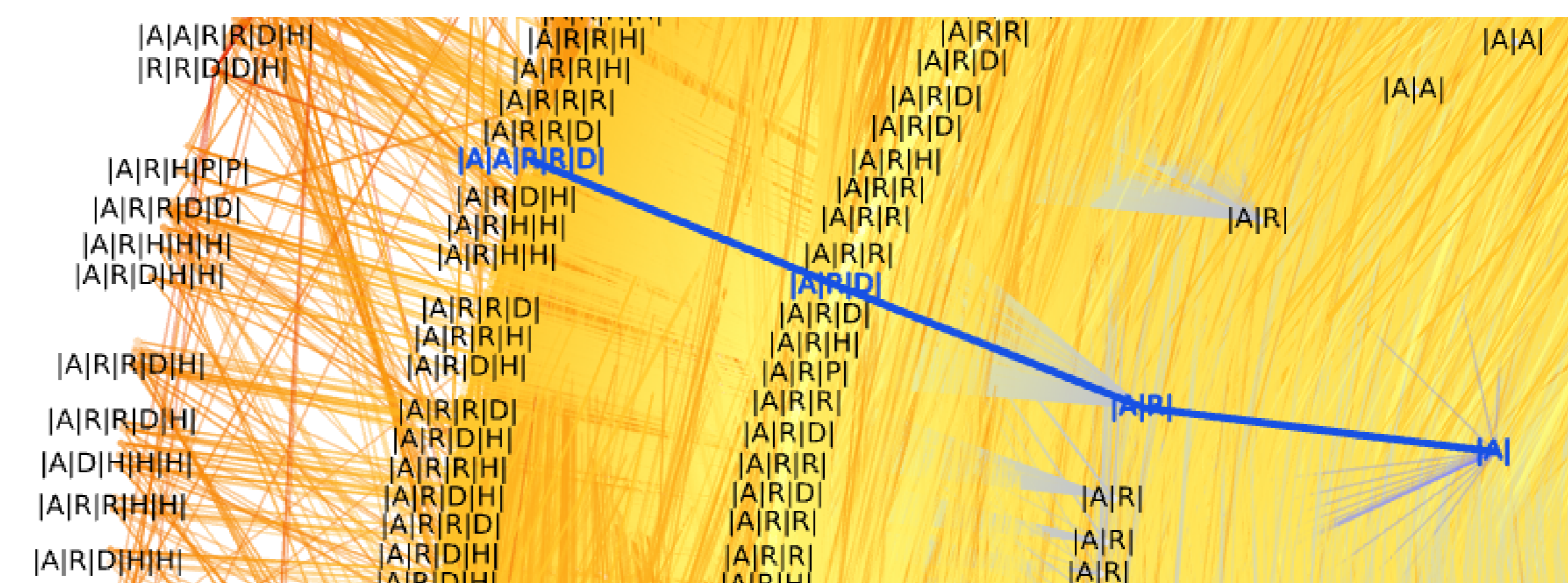


Figure 6: Chemin partant d'une classe simple à une classe complexe avec Tulip^a
^a. <https://tulip.labri.fr/site/>

Références

- [1] Sergei O. KUZNETSOV et Sergei A. OBIEDKOV. "Algorithms for the Construction of Concept Lattices and Their Diagram Graphs". In : sous la dir. de Luc De RAEDT et Arno SIEBES.
- [2] Jean-Philippe MÉTIVIER et al. "The Pharmacophore Network : A Computational Method for Exploring Structure–Activity Relationships from a Large Chemical Data Set". In : *Journal of Medicinal Chemistry* 61.8 (2018), p. 3551-3564.
- [3] Dagmar STUMPF, Huabin HU et Jürgen BAJORATH. "Evolving Concept of Activity Cliffs". In : *ACS Omega* 4.11 (2019), p. 14360-14368. doi : 10.1021/acsomega.9b02221. eprint : <https://doi.org/10.1021/acsomega.9b02221>. URL : <https://doi.org/10.1021/acsomega.9b02221>.

1. <https://involvd.greyc.fr/>
2. <https://www.ebi.ac.uk/chembl/>