

# Modèle de graphe pour l'analyse de dynamiques moléculaires

Ylène Aboulfath

Université Paris-Saclay, Laboratoire DAVID, Versailles, France

ylene.aboulfath@uvsq.fr

**Mots-clés :** *graphe, isomorphisme, chémo-informatique*

## 1 Introduction

Une part importante de la chémo-informatique consiste en la recherche de toutes les structures tridimensionnelles possibles pour une molécule. Chaque structure des atomes d'une molécule est liée à la présence, ou à l'absence, de liaisons hydrogènes. Dans une trajectoire, c.à.d l'évolution de la molécule dans le temps, les liaisons hydrogènes possibles sont connues. On peut donc représenter une trajectoire par une suite de graphes mixtes, dans lesquels les arêtes sont des liaisons communes à tous les graphes tandis que les arcs sont les liaisons hydrogènes. On suppose qu'on se positionne avec une granularité telle que dans deux graphes successifs une seule liaison hydrogène peut apparaître/disparaître de l'un à l'autre.

Des travaux récents [1], montrent qu'une méthodologie pertinente pour analyser la proximité structurelle entre plusieurs molécules est de les considérer non pas au niveau atomique mais au niveau des cycles formés par les liaisons entre les atomes. En effet, pour chaque graphe on obtient un ensemble de cycles dont la forme est liée aux liaisons hydrogènes.

Ainsi, l'hypothèse que l'on souhaite vérifier est l'existence de cycles qui impliquent la structure tridimensionnelle de la molécule de la même façon bien qu'ils changent de forme au fil des graphes. Selon cette hypothèse, il existe des ensembles de cycles qui sans être identiques, ont le même rôle structurel sur la molécule. La définition de ces cycles permettrait la reconnaissance de bassins sur la surface d'énergie potentielle, une problématique forte en chimie.

Notre objectif est de trouver des ensembles de cycles pertinents pour modéliser l'évolution d'une trajectoire. Ce que nous avons défini comme le problème de polymorphisme de cycles, et que nous présentons ici.

## 2 Modèle et Problème

Soit  $\mathcal{G} = \{G_1, \dots, G_m\}$  l'ensemble de graphes qui constituent une trajectoire. Chaque graphe  $G_i = (V, E, A_i) \in \mathcal{G}$  est une structure de la molécule. Soit  $A$  l'union de tous les arcs qui apparaissent dans au moins un graphe de  $\mathcal{G}$ .

Pour chaque graphe  $G_i$  on détermine  $B_i = \{c_1, \dots, c_l\}$  une base de cycles minimale (c.à.d. on peut construire tous les cycles du graphe à partir des éléments de cette base et la longueur totale de ces cycles est minimale).

Dans un graphe, une méthode pour obtenir une base de cycles minimale en temps polynomial est celle de Horton [2]. Nous avons adapté cette méthode, en incorporant des critères sur les liaisons hydrogènes dans la définition de la base afin qu'elle s'intègre mieux dans notre modèle. A partir de  $B_i$ , on définit  $GC_i$  le graphe des cycles correspondant où les sommets sont les cycles de la base, et où il existe une arête entre deux sommets s'ils partagent au moins une arête dans  $G_i$ . Ainsi, on peut construire l'ensemble des cycles  $\mathcal{C}$  de la trajectoire, c.à.d les cycles

qui apparaissent dans au moins une base de cycles  $\mathcal{C} = \bigcup_{i=1}^m B_i$ .

Une partition  $\Pi = \{\pi_1, \dots, \pi_k\}$  de cet ensemble  $\mathcal{C}$  est un polymorphisme de cycles de  $\langle \mathcal{G}, \mathcal{C} \rangle$  si et seulement si les critères suivants sont respectés :

- Dans toute partie  $\pi_i$  tous les cycles doivent partager au moins un sommet extrémité d'un arc. *Expérimentalement, on a constaté que les cycles qui partagent le même type de liaison hydrogène (c.à.d. avec une extrémité commune) peuvent avoir le même rôle.*
- Dans toute partie  $\pi_i$  aucun cycle ne peut apparaître dans la même base de cycles qu'un autre cycle de la même partie.
- Dans toute partie  $\pi_i$  chaque couple de cycles doit avoir des voisinages équivalents, c'est-à-dire :
  - On a deux cycles  $c, c' \in \pi_i$  qui apparaissent respectivement dans  $B_x$  et  $B_y$ .
  - Et, deux cycles  $z, z' \in \pi_j$  qui apparaissent respectivement dans  $B_x$  et  $B_y$ .
  - Alors, il existe une arête  $(c, z)$  si et seulement si il existe une arête  $(c', z')$ .

Notons, que chaque partie de la partition ainsi obtenue est un ensemble de cycles dont la forme peut changer au cours du temps mais ayant la même implication sur la structure de la molécule.

Finalement, on souhaite regrouper le plus de cycles possibles selon les critères du polymorphisme de cycles. On définit donc le problème Min-Polymorphisme qui minimise le nombre de parties de la partition formant un polymorphisme de cycles  $\langle \mathcal{G}, \mathcal{C} \rangle$ .

### 3 Résultats

Nous présenterons nos résultats sur le problème que nous avons défini : Min-Polymorphisme.

Le premier résultat concerne la complexité de ce problème de minimisation qui est NP-difficile, et ce même dans le cas des graphes planaires. Cela peut être montré par réduction polynomiale depuis le problème *Isomorphisme de sous-graphe induit* (ISI), celui-ci ayant été montré comme NP-complet même si les graphes sont "outerplanar"<sup>1</sup> [3].

On proposera ensuite une heuristique développée pour ce problème. Cette heuristique incrémentale se base sur une fonction d'évaluation du polymorphisme d'un ensemble de cycles, prenant en compte le nombre d'atomes qu'ils partagent, en plus des critères de définition du polymorphisme de cycles. L'efficacité de cette méthode a pu être vérifiée par comparaison avec les résultats d'une évaluation totale, coûteuse, mais possible pour certaines trajectoires.

Pour finir, on pourra alors montrer la pertinence de notre approche sur des exemples de trajectoires.

### Références

- [1] S. Nouleho Ilemo, D. Barth, O. David, F. Quessette, M-A. Weisser, D. Watel. *Improving graphs of cycles approach to structural similarity of molecules*. PLoS One. 2019 Dec 27;14(12) :e0226680. doi : 10.1371/journal.pone.0226680. PMID : 31881046.
- [2] J. D. Horton. *A polynomial-time algorithm to find the shortest cycle basis of a graph*. SIAM Journal of Computing, April, 1987.
- [3] M.M. Sysko *The subgraph isomorphism problem for outerplanar graphs*. Theoretical Computer Science 17(1), 91 – 97 (1982)

---

1. Un graphe outerplanar peut être dessiné dans le plan de telle sorte que tous les sommets appartiennent à la face extérieure