Maximiser l'intersection de bases de cycles minimum dans un ensemble de graphes dynamiques.

Ylène Aboulfath[1], Dimitri Watel[2]

[1] Université Paris-Saclay, Laboratoire DAVID, Versailles, France
[2] Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise
ylene.aboulfath@uvsq.fr, dimitri.watel@ensiie.fr

Mots-clés : *graphe dynamique, bases de cycles, complexité, chémo-informatique*

1 Introduction

Une part importante de la chémo-informatique consiste en la recherche de toutes les structures possibles pour une molécule. Chaque structure des atomes d'une molécule est liée à la présence ou non, 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, dans lesquels certaines arêtes sont communes à tous les graphes tandis que d'autres *varient au cours du temps.* La granularité de ces trajectoires est telle que dans deux graphes successifs une seule arête peut apparaître/disparaître, on les qualifie alors de edge-close.
Des travaux récents [2], 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. Pour chaque graphe on obtient alors un ensemble de cycles dont la forme est liée aux liaisons hydrogènes.
Ainsi, l'hypothèse que l'on souhaite confirmer est le lien entre les cycles et la structure tridimensionnelle de la molécule. Pour cela, on va utiliser la base de cycles minimum de chaque graphe, considérée alors comme l'ensemble de cycles pertinents de celui-ci. Dans un graphe $G = (V, E)$, *une base de cycles* $\mathcal{B}_G$ est un ensemble de cycles générateurs de tous les cycles du graphe. Une base est qualifiée de minimum lorsque la somme du poids des cycles qui la composent l'est. Cependant, une base de cycles minimum n'est pas toujours unique, et cela constitue un problème si on souhaite utiliser cette base comme un outil d'analyse.
Notre objectif est de trouver des ensembles de cycles pertinents pour modéliser l'évolution d'une trajectoire. On souhaite non seulement avoir une base de cycles minimum pour chaque graphe mais aussi que ces bases soient les plus "proches". Ainsi, nous avons défini un problème qui vise à maximiser l'intersection de bases de cycles minimum, et que nous présentons ici.

2 Définitions, Modèle et Problèmes

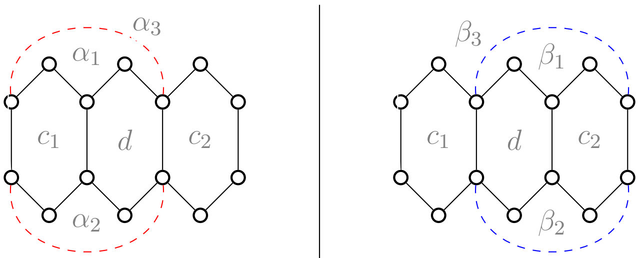
Trouver une base de cycles minimum pour un graphe peut se faire en temps polynomial, par exemple avec l'algorithme de Horton [1] : construire un ensemble de cycles élémentaires contenant une base de cycles minimum ; ordonner cet ensemble par ordre croissant de poids ; puis en prenant les cycles dans l'ordre avec un algorithme glouton ajouter un cycle à la base s'il est indépendant des cycles déjà pris.
On note que le nombre de cycles dans une base de cycles est connu $|\mathcal{B}_G| = |E| - |V| + 1$.

On considère une trajectoire comme une suite de graphes $\mathcal{S} = \mathcal{G}_1, \mathcal{G}_2, ...\mathcal{G}_t$ tels que $\mathcal{G}_i$ est une occurrence d'un graphe de l'ensemble $\mathcal{R} = \{G_1, G_2, ..., G_n\}$ avec $n \leq t$. On définit alors notre problème MMCBI (Maximum Minimum Cycle Basis Intersection) comme suit : A partir de l'ensemble $\mathcal{R} = \{G_1, \ldots, G_k\}$ formé de graphes d'une séquence $\mathcal{S}$, existe-t-il une base de cycle minmum $\mathcal{B}_i$ pour chaque graphe G_i tel qu'on maximise la taille de l'intersection entre ces bases.

3 Résultats

Les premiers résultats concernent la complexité de ce problème qui est connectée à la proximité qui existe entre les graphes. Dans le cas où on peut réordonner la séquence de graphes $\mathcal{S}$ de sorte que, pour tout couple de graphes, $\mathcal{G}_i \subset \mathcal{G}_j$ avec $i < j$, nous avons défini et prouvé un algorithme polynomial qui trouve une solution optimale.
Deux graphes edge-close vérifient nécessairement la propriété ci-dessus. Cependant, avec trois graphes ce n'est plus forcément le cas. On peut avoir une séquence $\mathcal{S} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3\}$ où une arête a est supprimée entre $\mathcal{G}_1$ et $\mathcal{G}_2$, et une arête b ($a \neq b$) est ajoutée entre $\mathcal{G}_2$ et $\mathcal{G}_3$. Alors, même si $\mathcal{G}_2$ est un sous-graphe de $\mathcal{G}_1$ et de $\mathcal{G}_3$, trouver $\mathcal{B}_1$ et $\mathcal{B}_3$ les bases de cycles minimum de $\mathcal{G}_1$ et de $\mathcal{G}_3$ qui résolvent le problème *MMCBI* n'est pas simple. La figure 1 illustre une situation de ce type. On a deux graphes F_1 et F_2 que l'on pourrait apparenté à G_1 et G_3. Le sous-graphe commun G_2 est représenté en traits pleins dans ces figures. F_1 et F_2 présentent chacun le choix entre deux bases de cycles minimum. Les bases de cycles des graphes sont composées des cycles courts dénotés par α et β : α_3 pour F_1 et β_3 pour F_2 de longueur 4, puis α_1 et α_2 pour F_1 (resp. β_1 et β_2) de longueur 5, ainsi que de deux cycles parmi $\{c_1, c_2, d\}$ de longueur 6. Le problème étant que dans la base de cycles de F_1, le cycle d peut être choisi au même titre que le cycle c_1 et de même dans la base de F_2. Cela rend les méthodes gloutonnes, très utilisées lorsqu'on recherche une seule base de cycles insuffisantes. Dans notre exemple, il existe une solution réalisable $\mathcal{B}'_1 \cap \mathcal{B}'_2 = \{d\}$ mais on ne peut pas transformer simplement cette solution en solution optimale.
C'est par ce biais que nous avons pu prouver que le problème était NP-difficile. Une réduction est possible depuis Minimum Monotone Planar 3-SAT. Il faut cependant noter que la réduction ne fonctionne que si le nombre de graphes dans $\mathcal{R}$ est non borné. La complexité paramétrée du problème vis à vis de la taille de $\mathcal{R}$ est ouverte.

Une de nos perspectives à court terme est de s'intéresser au cas où le nombre de graphes de $\mathcal{R}$ est fixe avec l'objectif de trouver un algorithme résolvant le problème en un temps raisonnable.



(a) Un graphe F_1

(b) Un graphe F_2

Bases de cycles min de F_1 :
- $\mathcal{B}_1$: $\alpha_1, \alpha_2, \alpha_3, c_1, c_2$
- $\mathcal{B}'_1$: $\alpha_1, \alpha_2, \alpha_3, d, c_2$

Bases de cycles min de F_2 :
- $\mathcal{B}_2$: $\beta_1, \beta_2, \beta_3, c_1, c_2$
- $\mathcal{B}'_2$: $\beta_1, \beta_2, \beta_3, c_1, d$

Solution optimale = 2 avec $|\mathcal{B}_1 \cap \mathcal{B}_2| = |\{c_1, c_2\}|$

FIG. 1 – Illustration des contraintes de construction d'une base de cycles en temps polynomial.

Références

[1] J. D. Horton. *A polynomial-time algorithm to find the shortest cycle basis of a graph.* SIAM Journal of Computing, April, 1987.

[2] 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.