

UNIVERSIT CLAUDE BERNARD LYON 1

Rapport de stage

Auteur

Eric Cumunel

Tuteur

Arnaud Mary

May 30, 2018

Contents

0.1 Matriels et mthodes :	2
-------------------------------------	---

0.1 Matriels et mthodes :

Matriels

...

Modlisation de la molcule et de son arbre de fragmentation

Soit la formule chimique associe une molcule. D'aprs un arbre de fragmentation, nous pouvons associer chaque noeud un fragment de la molcule. Ce noeud reprsente un ensemble d'atome dont la connexit a t conserve aprs fragmentation : nous faisons pour l'instant le choix de ne pas prendre en compte les potentiels rarrangements. [Note : c'est une hypothse videmment trop simpliste, nous reviendrons dessus plus tard.]

D'aprs les propriets de valence des atomes, il est ainsi possible de modliser un atome comme tant un sommet du graphe, ncessitant d'tre connect par un nombre prcis d'artes correspondant sa valence. Le degr de chaque sommet du graphe devra correspondre exactement la valence de l'atome auquel il est associ.

[Insrer graphe exemple]

Par extension, une formule chimique correspond une squence de degr.

[Insrer graphe exemple]

Chaque noeud de l'arbre de fragmentation apporte ainsi des contraintes de connexit supplmentaires tous les atomes qui le compose. Nous pouvons nous reprsenter cet arbre de fragmentation comme un supergraphe, connect, compos de graphes connects correspondant aux noeuds enfant de la racine de l'arbre, eux-mme compos de sous-graphes connexes associs aux petits-enfants de la racine; et cela jusqu' ce que l'on atteigne les feuilles de l'arbre.

[Big schma d'une patate-matriochka] -> définir feuille, noeud, racine
feuille = premier ensemble de sommet

Modéliser d'une telle manière l'arbre de fragmentation permet de poser le problème visuellement et de déduire les règles de connexion que devront respecter chaque noeud de l'arbre.

Stratégie adoptée :

Rappel : l'objectif est d'énumérer toutes les configurations structurales possible qu'une molécule puisse prendre tout en respectant un arbre de fragmentation.

D'après [papier], il a été démontré qu'à partir de n'importe quel graphe respectant un arbre de fragmentation, il est possible de reconstruire le supergraphe composé de tous les graphes répondant à ces contraintes de séquence de degré grâce à une opération appelée le switch (ou le swap, plusieurs dénominations existent dans la littérature).

Un point sur le switch :

Le switch est tout simplement l'opération consistant, à partir de deux arêtes d'un graphe, à intervertir deux sommets, nous redonnant au final deux nouvelles arêtes.

Deux switches différents peuvent être appliqués sur un jeu de deux arêtes :

[Schéma switch : straight/crossed]

Cette opération a la propriété de ne pas modifier la séquence de degré du graphe dont sont issues les arêtes.

En appliquant toutes les opérations de switch possible au sein d'un graphe (c-à-d, sur chaque paire d'arêtes), nous pouvons énumérer tous les graphes répondant aux contraintes de l'arbre de fragmentation : le supergraphe obtenu se compose en effet de chacun de ces graphes, reliés par l'opération de switch qui nous a permis de passer de l'un à l'autre.

[Schéma du supergraphe]

Plus précisément, ce supergraphe se compose de tous les graphes respectant les contraintes imposées par une séquence de degré : il contient également tous les graphes non-connexes. Cependant, d'après [papier ?], il a été démontré que le supergraphe ne perdait pas sa propriété de connexité lorsque sont retirés les sous-graphes non connexes. Ainsi, quelque soit le graphe (connexe) qui fait partie du supergraphe, il existera toujours un chemin (correspondant à une suite de switches) le reliant à n'importe quel autre graphe connexe.

[schéma ?]

Cela nous permet d'affirmer qu'à partir d'un seul graphe valide, nous pouvons en déduire toutes les solutions existantes pour une séquence de degré.

Il faudra ensuite faire un tri pour ne conserver que les graphes respectant également les conditions de connexité induites par chaque noeud de l'arbre de fragmentation (savoir que chaque ensemble de sommet d'un noeud doit être connexe).

Algorithme de rsolution

Gnratio du premier graphe respectant l'arbre de fragmentation :

Appart :

Soit une molcule compose de n atomes.

Soit $\text{sum}(\text{di})$ la somme de la squence de degr correspondante.

Le nombre d'artes du graphe representant cette molcule est gal $\text{sum}(\text{di})/2$. $\rightarrow \text{nb_e_tot}$

Le nombre d'artes minimum pour obtenir un graphe connexe est gal $(n-1)$. $\rightarrow \text{nb_e_min}$

Combien peut-on construire de graphe connexe non-isomorphe avec seulement $(n-1)$ artes?

Le nombre d'arte a ajouter une fois avoir obtenu ce graphe connexe est de $\text{sum}(\text{di}/2) - (n-1)$. $\rightarrow \text{nb_e_to_add}$

Exemple : C2N3O2H3

$\text{sum}(\text{di}) = 24$

$\text{nb_e_tot} = \text{sum}(\text{di})/2 = 12$

$\text{nb_e_min} = \text{nb_atom} - 1 = 9$

$\text{nb_e_to_add} = \text{nb_e_tot} - \text{nb_e_min} = 3$

?? D'aprs nauty (`geng -c 9 10:10 -u`), 797 graphes simples connexes et non-isomorphes existent pour $\text{nb_e_min} = 9$ artes pour 10 sommets.

Validit de l'arbre de fragmentation et de la squence de degr La procdure suivre pour obtenir ce premier graphe doit tre particulirement rigoureuse si nous voulons qu'elle soit adapte quelque soit la composition de la molcule et son arbre de fragmentation.

Dans un premier temps, prcisons qu'il est possible de vrifier au pralable qu'une solution existe rellement.

Pour cela, ... respect des 3 conditions ...

Si ces conditions sont respectes alors d'aprs [papier], il a t dmontr qu'au moins une solution existe pour ce problme.

Pour trouver ce premier graphe, il est ensuite ncessaire, aprs avoir gnr un sommet par atome, d'ajouter le bon nombre d'arte de telle sorte que chaque sommet atteigne exactement le degr impos par la squence de degr dduite de la formule chimique.

Connexion des feuilles Un des premieres difficults pouvant tre rencontre lors de la gnratio du premier graphe concerne la gestion des sommets ncessitant un grand nombre de degr.

En effet, si la distribution des artes entre les sommets n'est faite correctement, il est possible qu'un sommet se retrouve seul avec un nombre pair de liaison distribuer : or, tant donn que les loops soient interdits, le graphe doit tre considr comme invalide.

Pour avoir la certitude que cela n'arrive pas, une des solutions est de construire un arbre courant, pour chaque premier ensemble d'atome de l'arbre de fragmentation. Cela revient ajouter le minimum d'arte possible pour connecter les sommets de chacune des feuilles. L'objectif est ici de prvenir le problme que peuvent poser les sommets de fort degr tout en assurant la connexit des feuilles du graphe.

Brve description de l'algo en schma :

[Liste des atomes connects et non_connects d'une feuille. Jusqu' ce que tous les atomes soient connects : on prend dans chacune des listes l'atome qui il reste le plus de liaison tablier et on ajoute une arte dans le graphe entre ces deux atomes]

En itrant cette procudre sur toutes les feuilles, nous sommes ensuite en mesure d'tablier recursivement des artes au niveau des noeuds suprieurs de l'arbre de fragmentation sans avoir ajouter des boucles.

Connexion des noeuds suprieurs Une fois les premiers niveaux connects (les feuilles), nous pouvons commencer rpondre aux contraintes de connexit imposes par l'arbre de fragmentation dans les noeuds suprieurs.

Pour les mmes raisons que prcdemment, il est plus prudent de d'abord tablier la bonne connexit des noeuds, avant de commencer combler les sommets en ajoutant des liaisons. La dmarche adopte dans notre cas a t de partir des feuilles de l'arbre, puis de remonter recursivement les noeuds de l'arbre jusqu' la racine, en ajoutant des artes de manire connecter les noeuds de mme parent (toujours au niveau des atomes des noeuds disposant du plus grand nombre de liaison distribuer).

[petit schma]

Cela revient tablier un arbre courant mimant la hirarchie de l'arbre de fragmentation.

Nous obtenons la fin de cette tape un graphe simple, sans boucle, non-dirig, respectant toutes les conditions de connexit entre atomes de l'arbre de fragmentation : la dernire condition respecter pour obtenir un premier graphe valide est de combler les sommets disposant encore de degr distribuer.

Ajout des dernires artes Grce aux precautions prises lors des tapes prcdentes, nous disposons d'un nombre pair de degr libre, et ceux-ci sont disperss sur plusieurs sommets. En suivant le mme protocole que prcdemment, et en tenant compte cette fois de tous ces atomes quelque soit leur noeud, nous pouvons ajouter les dernires artes jusqu' ne plus avoir de degr libre. Le graphe ainsi obtenu est un multigraphe, sans boucle.

Nous obtenons ainsi un premier graphe respectant toutes les contraintes de l'arbre de fragmentation.

Gnration de tous les graphes valides

D'aprs [papier], il est possible d'obtenir tous les graphes connects respectant l'arbre de fragmentation partir d'un seul graphe connect.

Cela revient parcourir le supergraphe en partant du premier graphe en effectuant des oprations de switches successifs,

Optimisation