

# Identification de chaînes de réactions dans des réseaux métaboliques et génomiques pour la comparaison d'espèces

Florent Cabret, Ronan Bocquillon, Emmanuel Néron

Laboratoire LIFAT, Université de Tours, France  
{ronan.bocquillon,emmanuel.neron}@univ-tours.fr

**Mots-clés :** réseaux métaboliques et génomiques, bioinformatique, recherche opérationnelle.

## 1 Introduction

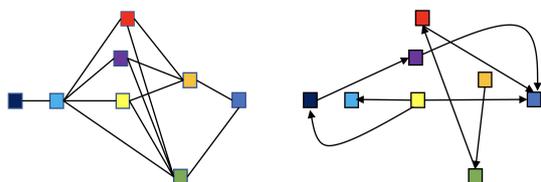
Pour comprendre le fonctionnement des organismes vivants, les biologistes ont commencé par étudier séparément leurs composants biologiques. Cette approche ne suffit pas pour des organismes complexes et les biologistes s'intéressent désormais à l'étude des relations entre ces composants en comparant les réseaux [2].

Dans la continuité de travaux de la littérature [3, 4], nous nous intéressons ici à l'étude conjointe de réseaux métaboliques et génomiques. Nous cherchons à identifier des *marqueurs* d'espèce en vue de leur comparaison. Le réseau métabolique est modélisé par un graphe orienté  $D$  dont les sommets représentent des réactions et dont les arcs indiquent qu'une réaction produit un métabolite substrat d'une autre réaction. La proximité génomique est représentée par un graphe non-orienté  $G$ , dans lequel deux réactions sont reliées par une arête si elles sont catalysées par des enzymes codées par des gènes "proches" (éloignés d'au plus  $\delta_G$  gènes) sur le génome. Notons que les deux graphes partagent le même ensemble de sommets.

Il s'agit alors de chercher, pour une espèce donnée, un plus long chemin, élémentaire ou non, dans le graphe  $D$  (réseau métabolique), dont les sommets induisent un sous-graphe connexe de  $G$  (proximité génomique). Ces plus longs chemins sont ensuite comparés d'une espèce à l'autre pour identifier des motifs présents dans plusieurs espèces.

## 2 Description du problème et travaux de références

Considérons un graphe acyclique orienté  $D$  et un graphe non-orienté  $G$ , construits sur le même ensemble de sommets  $V = \{1, 2, \dots, n\}$ . Un chemin  $DG$ -consistant est un chemin dans  $D$ , tel que le sous-graphe de  $G$  induit par les sommets de ce chemin est connexe. Le problème *One-To-One SkewGraM* consiste à calculer un plus long chemin  $DG$ -consistant.



Une instance du problème *One-To-One SkewGraM* : à gauche  $G$  ; à droite  $D$  ; les sommets orange, vert, rouge et bleu forment un chemin  $DG$ -consistant de longueur 4.

Fertin et al. [3] ont introduit ce problème, ont montré sa NP-difficulté au sens fort, et ont proposé une méthode exacte de type *Branch-and-Bound*. Plus récemment le problème a été étendu à la recherche de trails (chemins simples mais non-nécessairement élémentaires ; les chemins peuvent passer plusieurs fois par un même sommet, mais pas par un même arc), de couverture maximale [4]. Des modèles mathématiques et de programmation par contraintes [1] ont depuis été proposés pour ces deux variantes du problème.

### 3 Prise en compte des objectifs de comparaison d'espèces dès l'identification des chaînes de réactions

Dans la chaîne de traitements proposée par Zaharia et *al.* [4], les trails une fois identifiés, sont utilisés pour la comparaison d'espèces. Pour cela, il faut construire un tableau — dont chaque ligne est une réaction de l'espèce de référence, dont chaque colonne est une espèce cible, et dont les cellules peuvent contenir trois symboles différents (x, o et -) — sur la base duquel une “distance” inter-espèces peut être calculée. Pour chaque espèce cible on construit le sous-ensemble-connexe maximum de  $D$ . On remplit une cellule avec un 'x' si la réaction se trouve à la fois dans l'espèce de référence et dans ce sous-ensemble, un '-' si elle existe dans l'espèce cible mais en dehors de ce sous-ensemble, et un 'o' sinon. Il faut également préciser que ce tableau est à considérer pour chaque distance maximale entre gènes ( $\delta_G$ ) utilisée pour construire  $G$ . Le défaut de cette approche est qu'elle ne considère que très peu le fait qu'une partie du réseau métabolique est conservée entre les deux espèces, c.-à-d. que les réactions forment ou non un trail dans l'espèce cible.

Notre travail cherche à intégrer cette notion de distance entre trails, et donc entre espèces dans la recherche des trails, en considérant que les réactions doivent aussi correspondre à un trail dans l'espèce cible et en intégrant une notion de proximité. On définit cette proximité grâce au problème du *Graph Edit Distance*, qui correspond à trouver un ensemble minimum de transformations (insertions, délétions ou substitutions de sommets et arêtes) au sens d'une fonction de coût sur ces mêmes opérations pour passer d'un graph à un autre. La particularité d'appliquer ce problème entre deux trails est que l'on connaît à l'avance les sommets et arêtes en commun et que les différences (*e.g.* une réaction qui change de nom car elle utilise une autre enzyme mais consomme et produit les mêmes molécules, ou encore une nouvelle réaction qui relie deux réactions communes) peuvent être traitées indépendamment les unes des autres et sont souvent petites. On cherchera aussi à relaxer la distance  $\delta_G$  en lui attribuant une valeur plus grande que dans les autres études, mais en limitant la somme totale  $\Delta_G$  induite par la composante connexe de  $G$ .

### 4 Conclusion et perspectives

De premiers résultats expérimentaux sur cette approche prenant en compte le critère de comparaison d'espèces dès la recherche des trails  $DG$ -consistants seront présentés lors de la conférence et comparés aux résultats obtenus avec les approches actuelles de la littérature.

### Références

- [1] Mohamed Lemine AHMED SIDI et al. « Improved Approaches to Solve the One-To-One SkewGraM Problem ». In : *Computers & Operations Research* 138 (1<sup>er</sup> fév. 2022), p. 105584. ISSN : 0305-0548. DOI : 10.1016/j.cor.2021.105584.
- [2] Frank J. BRUGGEMAN et Hans V. WESTERHOFF. « The Nature of Systems Biology ». In : *Trends in Microbiology* 15.1 (1<sup>er</sup> jan. 2007), p. 45-50. ISSN : 0966-842X, 1878-4380. DOI : 10.1016/j.tim.2006.11.003. pmid : 17113776.
- [3] Guillaume FERTIN, Hafedh MOHAMED BABOU et Irena RUSU. « Algorithms for Subnetwork Mining in Heterogeneous Networks ». In : *Experimental Algorithms*. Sous la dir. de Ralf KLASING. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 2012, p. 184-194. ISBN : 978-3-642-30850-5. DOI : 10.1007/978-3-642-30850-5\_17.
- [4] Alexandra ZAHARIA et al. « CoMetGeNe : Mining Conserved Neighborhood Patterns in Metabolic and Genomic Contexts ». In : *BMC Bioinformatics* 20.1 (10 jan. 2019), p. 19. ISSN : 1471-2105. DOI : 10.1186/s12859-018-2542-2.