



## Exact methods for finding a longest DG-consistent trail in biological networks

---

Mohamed Lemine Ahmed Sidi, Ronan Bocquillon,  
Hafedh Mohamed Babou, Cheikh Dhib,  
Mohamedade Farouk Nanne, Emmanuel Neron and  
Ameur Soukhal

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

February 10, 2020

# Méthodes exactes pour la détermination d'un plus long trail $DG$ -consistant dans des réseaux biologiques

Mohamed Lemine AHMED SIDI<sup>1,2</sup>, Ronan BOCQUILLION<sup>1</sup>,  
Hafedh Mohamed BABOU<sup>2</sup>, Cheikh DHIB<sup>2</sup>, Mohamedade Farouk NANNE<sup>2</sup>,  
Emmanuel NERON<sup>1</sup>, Ameer SOUKHAL<sup>1</sup>

<sup>1</sup> Laboratoire d'Informatique Fondamentale et Appliquée de Tours, LIFAT EA 6300,  
ROOT ERL-CNRS 7002, Polytech Tours 64 av. Jean Portalis, 37200 Tours, France  
mohamedlemine.ahmedsidi@etu.univ-tours.fr,

{ronan.bocquillion,emmanuel.neron,ameur.soukhal}@univ-tours.fr

<sup>2</sup> Unité de Recherche Documents Numériques et Interaction de l'Université de Nouakchott  
Al-Asriya, BP 880, Nouakchott, Mauritanie  
hafedh.mohamed-babou@esp.mr, dhib.cheikh@iscae.mr, farouk@una.mr

**Mots-clés** : *bio-informatique, comparaison de réseaux biologiques, programmation linéaire en nombres entiers, procédure par séparation et évaluation.*

## 1 Introduction

La biologie des systèmes est un domaine récent proposant d'étudier les organismes vivants tels qu'ils se présentent en réalité. Cette approche s'oppose aux précédentes en intégrant différents niveaux d'information (biologiques, physiologiques, biochimiques, *etc.*) pour comprendre le fonctionnement de ces organismes [5]. Cela nécessite des algorithmes de traitement et d'analyse de plus en plus spécialisés et efficaces. De nombreuses approches destinées à la comparaison de réseaux biologiques (homogènes ou hétérogènes) reposent sur des modèles de graphe.

En effet, en bio-informatique, ces réseaux sont souvent modélisés par des graphes. Les sommets sont des composants biologiques et les arêtes représentent leurs interactions. En fonction de la nature de ces réseaux, les graphes correspondants peuvent être orientés ou non-orientés. Notre intérêt porte sur la comparaison de deux réseaux biologiques hétérogènes.

L'objectif principal est d'étudier la relation entre le métabolisme et le génome. Nous avons choisi de nous concentrer sur la détection de réactions voisines catalysées par des produits de gènes voisins, où la notion de voisinage peut être modulée en autorisant que certaines réactions et/ou gènes soient omis. Ces travaux s'inscrivent dans la continuité de ceux entrepris dans [1] et [2], où les auteurs ont identifié un problème d'optimisation combinatoire  $\mathcal{NP}$ -difficile, appelé *One-to-One SkewGraM*, et défini comme suit.

### One-to-One SkewGraM

- **Définition** : Soient  $D = (V, A)$  un graphe orienté acyclique et  $G = (V, E)$  un graphe non-orienté. Un chemin  $P$  dans  $D$  est  $DG$ -consistant ssi  $G[V(P)]$  est *connexe*.
- **Objectif** : Trouver un plus long chemin  $DG$ -consistant

Des travaux récents abordent une version plus générale du *One-to-One SkewGraM*, appelé *SkewGraM* par [1] et [2], où  $D$  peut contenir des cycles. Les auteurs cherchent à calculer des "trails" de réactions catalysées par des produits de gènes voisins [3].

**SkewGraM :**

- **Définition :** Soient  $D = (V, A)$  un graphe orienté cyclique et  $G = (V, E)$  un graphe non-orienté. Un *trail*  $T$  dans  $D$  est un chemin *élémentaire* dans  $D$ . La *couverture* d'un trail est le nombre de sommets *distincts* qu'il traverse.
- **Objectif :** Trouver un trail  $DG$ -consistant de couverture maximale.

À notre connaissance, peu d'études ont été consacrées à la comparaison de réseaux hétérogènes. Dans [1] et [2], pour le problème *One-to-One SkewGraM*, les auteurs identifient quelques cas polynomiaux. Pour résoudre le cas général, une méthode exacte de type branch-and-bound est proposée. Celle-ci permet de calculer le plus long chemin  $DG$ -consistant passant par un arc choisi initialement. Cette méthode a fait l'objet d'une étude dans [4], où les auteurs proposent des améliorations.

Dans [3], les auteurs proposent une méthode exacte pour résoudre le problème *SkewGraM*. L'idée principale de leur approche de résolution est l'énumération de tous les chemins.

Toutes ces méthodes exactes s'avèrent peu efficaces dès que la taille des instances augmentent (elles échouent sur des instances réelles de plus de 50 réactions).

## 2 Approches de résolution exactes

Nous proposons des méthodes de résolution basées la programmation linéaire en nombres entiers pour les deux problèmes ci-dessus (le *One-to-One SkewGraM* et le *SkewGraM*).

Pour le problème *One-to-One SkewGraM*, nous présenterons un nouveau modèle linéaire plus efficace (temps de calcul réduit) que ceux présentés dans [4]. Pour le problème *SkewGraM*, nous présenterons deux modèles linéaires améliorant les résultats de la littérature (en termes de qualité des solutions et de temps de calcul).

Tous ces modèles sont constitués de deux ensembles de contraintes. Le premier correspond à la recherche d'un trail ou chemin dans  $D$ . Le deuxième vérifie la connexité du sous-graphe dans  $G$  induit par les sommets du trail ou du chemin dans  $D$ . Ce deuxième ensemble de contraintes correspond donc à la recherche d'un arbre couvrant.

## 3 Conclusions

Pour le *One-to-One SkewGraM*, un nouveau PLNE a été développé. Les résultats expérimentaux sur des instances générées aléatoirement montrent son efficacité par rapport aux méthodes présentées dans [4]. Concernant *SkewGraM*, deux nouveaux PLNE ont été développés. Ces PLNE ont été testés sur des instances réelles. Les premiers résultats montrent l'efficacité de nos approches par rapport à l'existant.

## Références

- [1] G. Fertin, H. M. Babou, and I. Rusu. Algorithms for subnet-work mining in heterogeneous networks. In Symposium on Experimental Algorithms (SEA), 2012.
- [2] H. M. Babou. Comparaison de réseaux biologiques. Thèse de doctorat en Bio-Informatique, Université de Nantes, 2012.
- [3] A. Zaharia, B. Labedan, C. Froidevaux, and A. Denise. CoMetGeNe : mining conserved neighborhood patterns in metabolic and genomic contexts. BMC Bioinformatics, 2019
- [4] M. L. Ahmed Sidi, R. Bocquillon, H. Mohamed Babou, C. Dhib, M. F. Nanne, E. Néron, and A. Soukhal. Méthodes exactes pour la détermination d'un plus long chemin  $DG$ -consistant dans des réseaux biologiques. Congrès annuel de la ROADEF, société Française de Recherche Opérationnelle et d'Aide à la Décision, Le Havre, France, 2019.
- [5] A. Lesne. Multiscale Analysis of Biological Systems. Acta Biotheorica, 2013.