

Un modèle de programmation par contraintes pour la recherche d'un chemin DG -consistant dans des réseaux biologiques

Mohamed Lemine AHMED SIDI^{1,2,3,4}, Ronan BOCQUILLON¹,
Hafedh MOHAMED BABOU^{2,4}, Cheikh DHIB^{2,3}, Emmanuel NÉRON¹,
Ameur SOUKHAL¹, Mohamedade Farouk NANNE²

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

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

² Unité de Recherche Calcul Scientifiques, Informatique et Data Science de l'Université
de Nouakchott Al-Asriya, Nouakchott, Mauritanie

³ Unité de Recherche Pluridisciplinaire en Informatique et Mathématiques Appliquées de l'Institut
Supérieur de Comptabilité et d'Administration des Entreprises, Nouakchott, Mauritanie

⁴ Unité de Recherche Intelligence Artificielle de l'École Supérieure Polytechnique de Nouakchott,
Nouakchott, Mauritanie

hafedh.mohamed-babou@esp.mr, cheikh.dhib@iscae.mr, mnanne@fst.e-una.mr

Mots-clés : *programmation par contraintes, bio-informatique.*

1 Introduction

L'analyse de réseaux issus de données réelles est un domaine qui a donné lieu à de nombreux travaux ces dernières années. Le concept de "réseau" offre un modèle de représentation pour une grande variété de systèmes, aussi bien naturels que sociaux, dans lesquels un ensemble d'entités homogènes ou hétérogènes interagissent entre elles. Les graphes sont communément utilisés comme support formel pour modéliser la structure de ces réseaux, dont la sémantique peut être diverse. Dans le cas d'un réseau d'interactions biologiques, les composants biologiques (*e.g.* les protéines) sont représentés par des nœuds, et une arête entre deux nœuds représente une possible interaction. Selon la nature de ces réseaux, les graphes correspondants sont orientés ou non-orientés.

Pour comprendre le fonctionnement des organismes vivants, les biologistes ont commencé par étudier individuellement 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 [1].

Il est possible de comparer des réseaux homogènes (c'est-à-dire de même type), ce qui conduit généralement à des problèmes d'alignement de graphes très étudiés dans la littérature, ou de comparer des réseaux hétérogènes. Peu d'études portent sur ce deuxième cas. Nous citons Fertin et *al.* [2], qui ont introduit un problème NP-difficile au sens fort nommé *one-to-one skewgram*, défini comme suit.

Considérons un graphe acyclique orienté D et un graphe non-orienté G , induits 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 dans 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. Les sommets V représentent des protéines, le graphe D un réseau métabolique, et G la proximité des gènes qui produisent ces protéines. On cherche ainsi un plus long chemin métabolique, qui implique des protéines produites par des gènes voisins. De tels chemins sont des marqueurs des espèces et peuvent servir à leur comparaison.

Dans [2], un algorithme nommé AlgoBB a été développé. Une version améliorée de AlgoBB, nommée AlgoBB++, et deux formulations linéaires ont été proposées dans [4, 5].

2 Une nouvelle approche de résolution exacte

Nous présenterons dans cet exposé un nouveau modèle de *programmation par contraintes* pour le problème *one-to-one skewgram*, qui s’avère plus efficace, en pratique, que toutes les méthodes présentées dans [4, 5]. Le modèle complet ne peut tenir dans ces deux pages.

Nous vous présentons néanmoins quelques résultats numériques préliminaires. Nous avons reporté dans le Tableau 1 les résultats obtenus sur cinq instances de type Erdős–Rényi générées aléatoirement, par la meilleure méthode proposée jusqu’à maintenant (nommée PLNE 2) et par la nouvelle (PPC). La durée d’exécution est limitée à 1300 secondes par instance. Les valeurs obtenues sont en gras lorsque la méthode a prouvé leur optimalité dans le temps imparti.

L’algorithme AlgoBB++ évoqué plus haut échoue sur les instances de 460 sommets et plus, PLNE 2 sur les instances de 3000 sommets, PPC sur les instances de 5000 sommets.

TAB. 1 – Résultats numériques préliminaires

Instance	Nombre de sommets	Nombre d’arcs	PLNE 2 (CPLEX)		PPC (Choco)	
			Temps	Longueur du chemin	Temps	Longueur du chemin
1	3000	224736	1252.09	262	432.20	262
2	3000	224210	1300	0	181.57	276
3	3000	224655	1300	2	378.03	276
4	3000	224409	1300	0	479.90	263
5	3000	224973	1300	3	728.34	292

3 Conclusions

Un modèle de programmation par contraintes a été développé pour résoudre le problème *one-to-one skewgram*. Les résultats expérimentaux que nous avons obtenu sur des instances générées aléatoirement et des données réelles montrent son efficacité.

Des travaux récents abordent une version plus générale du problème *one-to-one skewgram*, appelée *skewgram* par [2], où D peut contenir des circuits [3]. Nous sommes donc en train de travailler à l’adaptation de nos modèles à ce nouveau cadre. Nous évaluons nos solutions en termes de qualité en les injectant dans le module dit de “trail grouping”, développé par Zaharia et *al.*, et utilisé par les biologistes pour comparer les espèces. Si le temps nous le permet, nous discuterons des ces perspectives à la fin de notre exposé.

Références

- [1] F. Bruggeman and H. Westerhoff. The nature of systems biology. *Trends in Microbiology*, 15:45–50, 2007.
- [2] G. Fertin, H. M. Babou and I. Rusu. Algorithms for subnetwork mining in heterogeneous networks. In *Symposium on Experimental Algorithms (SEA)*, v. 7276, pp 184–194, 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. 20^e congrès annuel de la ROADEF, Le Havre, 2019.
- [5] 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 trail DG -consistant dans des réseaux biologiques. 21^e congrès annuel de la ROADEF, Montpellier, 2020.