

Quadruplets et réseaux non enracinés de niveau k

Philippe Gambette*

Vincent Berry*

Christophe Paul*

20 septembre 2010

Résumé

Les réseaux phylogénétiques permettent d'enrichir le modèle arboré de l'évolution, en insérant entre les branches de l'arbre du vivant des arêtes qui expriment les échanges de matériel génétique entre espèces co-existantes. Ces échanges impliquent que l'histoire de chaque gène peut toujours être décrite par un arbre, mais que ces arbres peuvent être différents pour deux gènes n'ayant pas suivi les mêmes processus évolutifs.

De nombreuses approches combinatoires ont été conçues pour reconstruire des réseaux phylogénétiques à partir de données extraites de plusieurs arbres de gènes contradictoires, en particulier à partir de *triplets*, c'est-à-dire d'arbres phylogénétiques binaires enracinés à trois feuilles [1, 3, 2]. Toutefois, le processus d'enracinement étant souvent source d'erreurs en reconstruction phylogénétique, il est préférable de l'effectuer en dernière étape.

Nous nous intéressons donc au problème de reconstruction de réseaux phylogénétiques non enracinés à partir de *quadruplets*, qui sont des arbres phylogénétiques à quatre feuilles. Nous montrons comment généraliser au contexte non enraciné le paramètre de *niveau* introduit pour exprimer la complexité des réseaux enracinés, et pour obtenir des algorithmes en temps polynomial pour la reconstruction de réseaux à niveau borné.

Nous montrons tout d'abord qu'il est possible d'obtenir un algorithme en temps polynomial pour déterminer l'ensemble de tous les quadruplets d'un réseau non enraciné, en utilisant un algorithme de résolution du problème 2-DISJOINT PATHS. En revanche, déterminer s'il existe un réseau non enraciné de niveau 1 qui contient un ensemble de quadruplets donné est NP-complet.

Nous étudions alors la restriction de *densité*, où pour tout ensemble de quatre feuilles il existe, parmi les données en entrée, au moins un quadruplet sur ces feuilles. Dans ce cas, nous généralisons l'approche de décomposition du problème en *SN-ensembles* introduite pour les réseaux enracinés [1]. Elle permet de traiter de manière indépendante chacune des composantes biconnexes du réseau à reconstruire.

Nous étudions donc le problème de reconstruction de réseaux simples, c'est-à-dire n'ayant qu'une composante biconnexe non triviale, de niveau 1. Il est possible de résoudre ce problème en temps polynomial quand on connaît tous les quadruplets du réseau, mais la question reste ouverte dans le cas dense. Nous illustrons la difficulté de ce problème, variante de BETWEENNESS, en montrant que plusieurs réseaux simples de niveau 1, et même un nombre exponentiel en fonction du nombre de feuilles, peuvent contenir un même ensemble dense de quadruplets.

Références

- [1] Jesper Jansson and Wing-Kin Sung. Inferring a level-1 phylogenetic network from a dense set of rooted triplets. *Theoretical Computer Science*, 363(1):60–68, 2006. <http://dx.doi.org/10.1016/j.tcs.2006.06.022>.
- [2] Thu-Hien To and Michel Habib. Level- k phylogenetic networks are constructable from a dense triplet set in polynomial time. In *Proceedings of the 20th Annual Symposium on Combinatorial Pattern Matching (CPM'09)*, volume 5577 of *Lecture Notes in Computer Science*, pages 275–288. Springer Verlag, 2009. http://dx.doi.org/10.1007/978-3-642-02441-2_25.
- [3] Leo van Iersel, Judith Keijsper, Steven Kelk, Leen Stougie, Ferry Hagen, and Teun Boekhout. Constructing level-2 phylogenetic networks from triplets. *IEEE/ACM Transactions in Computational Biology and Bioinformatics*, 6(4):667–681, 2009. <http://dx.doi.org/10.1109/TCBB.2009.22>.

* Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM, UMR 5506, CNRS - Université Montpellier II), 161, rue Ada, 34392 Montpellier Cedex 5, France