

# Appariement et fouille dans les graphes de connaissances du Web des données Applications en pharmacogénomique

Pierre Monnin

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France  
pierre.monnin@loria.fr,  
<https://pmonnin.github.io/>

Dans le Web des données, des graphes de connaissances de plus en plus nombreux sont simultanément publiés, édités, et utilisés par des agents humains et logiciels. Cette large adoption rend essentielles les tâches d'*appariement* et de *fouille*. L'*appariement* identifie des unités de connaissances équivalentes, plus spécifiques ou similaires au sein et entre graphes de connaissances. Cette tâche est cruciale car la publication et l'édition parallèles peuvent mener à des graphes de connaissances co-existants et complémentaires. Cependant, l'hétérogénéité inhérente aux graphes de connaissances (*e.g.*, granularité, vocabulaires, ou complétude) rend cette tâche difficile. Motivés par une application en pharmacogénomique, nous proposons deux approches pour appairer des relations  $n$ -aires représentées au sein de graphes de connaissances : une méthode symbolique à base de règles et une méthode numérique basée sur le plongement de graphe. Nous les expérimentons sur PGxLOD, un graphe de connaissances que nous avons construit de manière semi-automatique en intégrant des relations pharmacogénomiques de trois sources du domaine. La tâche de *fouille* permet quant à elle de découvrir de nouvelles unités de connaissances à partir des graphes de connaissances. Leur taille croissante et leur nature combinatoire entraînent des problèmes de passage à l'échelle que nous étudions dans le cadre de la fouille de patrons de chemins. Nous proposons également l'annotation de concepts, une méthode d'amélioration des graphes de connaissances qui étend l'Analyse Formelle de Concepts, un cadre mathématique groupant des entités en fonction de leurs attributs communs. Au cours de tous nos travaux, nous nous sommes particulièrement intéressés à tirer parti des connaissances de domaines formalisées au sein d'ontologies qui peuvent être associées aux graphes de connaissances. Nous montrons notamment que, lorsqu'elles sont prises en compte, ces connaissances permettent de réduire l'impact des problèmes d'hétérogénéité et de passage à l'échelle dans les tâches d'appariement et de fouille.

