

Nom – Prénom Sabeur ARIDHI
Laboratoire de rattachement Laboratoire lorrain de recherche en informatique et ses applications (LORIA)
Intitulé du diplôme HDR HDR Informatique
Titre de l'HDR Découverte de connaissances à partir de grands graphes biologiques

Abstract (français)

Les graphes sont utilisés dans de nombreux domaines différents, allant de la sécurité informatique des réseaux sociaux, à la géographie et à la bioinformatique. Leur polyvalence et leur capacité à modéliser des données complexes en font un outil précieux pour résoudre plusieurs problèmes notamment dans le domaine de l'extraction de connaissances à partir de graphes biologiques qui est notre domaine d'intérêt. Cependant, l'utilisation de graphes peut présenter plusieurs défis tels que la représentation des données, l'évolutivité et le comportement dynamique/temporel de certains graphes réels. Dans le cas particulier des graphes biologiques, des défis liés à la complexité des systèmes biologiques, à l'interprétabilité et à la validation pourraient être soulevés. Sans atteindre tous ces défis, nous présentons dans ce manuscrit plusieurs contributions. Tout d'abord, nous proposons une nouvelle approche basée sur un graphe de protéines qui combine la notion de similarité de domaine avec une technique d'inférence de voisinage de graphe pour l'annotation de la fonction des protéines. Ensuite, nous avons décrit une approche de prédiction de lien explicable dans les graphes de connaissances biologiques pour le repositionnement des médicaments. Enfin, nous présentons nos contributions sur le clustering de graphes distribués et le plongement de graphes de connaissances à grande échelle.

Mots-clés: Annotation fonctionnelle de protéines, graphes de connaissances, prédiction de liens, clustering structurel, plongement de graphes de connaissances.

Abstract (anglais)

Graphs are used in many different fields, ranging from social networks computer security, geography and bioinformatics. Their versatility and their ability to model complex data make them a valuable tool for solving several problems especially in the domain of knowledge extraction from biological graphs which is our domain of interest. However, the use of graphs can present several challenges such as data representation, scalability and the dynamic/temporal behavior of some real world graphs. In the particular case of biological graphs, challenges related to the complexity of biological systems, interpretability and validation could be raised. Without achieving all of these challenges, we present in this manuscript several contributions. First, we propose a novel protein-protein network (PPN) based approach which combines the notion of domain similarity with a graph neighborhood inference technique for protein function annotation. Then, we described an explainable link prediction approach in biological knowledge graphs for drug repositioning. Finally, we present our contributions on distributed graph clustering and large- scale knowledge graph embedding.

Keywords: Protein function annotation, knowledge graphs, link prediction, structural clustering, knowledge graph embedding.