

Appel à projets Challenge 2

Fiche Allocation doctorale et co-encadrement étranger

NB: Avant toute demande d'allocation, compléter une fiche «Action scientifique»

ACTION CONCERNÉE

Intitulé de l'action (et acronyme)	Enumération des Structures Intéressantes : vers des algorithmes performants (ESI)
Porteur de l'action	Nom : Vincent Limouzy/Lhouari Nourine Qualité/Titre : MCF/ Prof Courriel : vincent.limouzy@uca.fr/ lhouari.nourine@uca.fr Laboratoire de rattachement :LIMOS

IDENTIFICATION DU PROJET DE THÈSE

Encadrement	
Directeur de thèse	Nom :Nourine, Lhouari Qualité/Titre : Professeur Courriel : lhouari.nourine@uca.fr Laboratoire : LIMOS Taux d'encadrement : 30%
Co-encadrants nationaux de thèse	Noms : Limouzy, Vincent Qualités/Titres : Maître de Conférences Laboratoires : LIMOS Taux d'encadrement : 50%
Co-encadrant étranger de thèse	Nom : Uno, Takeaki Qualité/Titre : Professeur Laboratoire : National Institute of Informatics (NII) - Tokyo, Japon Taux d'encadrement : 20% Présentation succincte du co-encadrant étranger : NB : Toute obtention d'une allocation doctorale nécessitera un co-encadrement par un chercheur étranger qui s'engagera à venir un mois par an durant les 3 ans de la thèse
Partenaires socioéconomiques impliqués (facultatif)	
Partenaires internationaux impliqués (facultatif)	National Institute of Informatics - Tokyo - Japon
Sujet	
Titre du projet de thèse	Enumération des Structures Intéressantes : vers des algorithmes performants NB : En français et en anglais éventuellement.
Descriptif du projet de thèse (limité à 300 mots)	Cette demande d'allocation doctorale fait partie de l'action ESI (<i>Enumération des Structures Intéressantes : vers des algorithmes performants</i>) proposée en réponse à l'appel à projets 2018 du thème Ressourcement du <i>Challenge 2</i> du programme CAP 20-25 . Ce travail de thèse sera co-encadré par Takeaki UNO, professeur au NII, Tokyo, Japon. Le sujet de thèse est décrit ci-dessous, suivi d'une présentation du professeur Takeaki UNO et de son apport à ce travail de thèse.

Sujet de thèse

L'objectif de la thèse serait l'étude des problèmes de réduction et de complétion des données dans le but de garantir des propriétés de performance des algorithmes. La complétion des données se ramène à ajouter des informations manquantes aux données pour qu'elle devienne intéressante, alors que la réduction consiste à ignorer des informations des données, la tâche consiste alors à déterminer quelles sont les informations inutiles.

Dans cette thèse nous nous limitons aux données graphes. Ce type de données est aujourd'hui omniprésents et constitue le modèle de données de base des systèmes de gestion de données modernes de la mouvance NoSQL. Parmi les graphes intéressants on peut identifier des familles de graphes qui possèdent des propriétés remarquables, comme par exemple les graphes d'intervalles, les graphes cordaux ou les graphes de comparabilité. Pour ces familles de graphes, les problèmes d'optimisations qui d'ordinaire s'avèrent difficiles à résoudre peuvent devenir facile (c.-à-d., disposer d'algorithmes efficaces pour trouver une solution optimale).

Cependant, lorsque l'on traite un problème sur un graphe issu d'un problème de la vie réelle, il est fréquent sinon systématique que le graphe n'exhibe pas les bonnes propriétés décrites ci-dessus.

L'objectif d'une complétion de graphe est alors d'ajouter des arêtes dans le graphe original de manière à ce que le graphe résultant respecte une propriété particulière (par exemple, le graphe d'origine se transforme en un graphe d'intervalle). Le problème de **complétion minimum**, s'avère dans la majorité des cas être un problème difficile (NP-difficile). Par contre dès lors qu'on demande une **complétion minimale** (au sens de l'inclusion), le problème devient souvent plus facile à résoudre. Dans ce contexte, les problèmes de complétions minimales ont été étudiés ces dix dernières années [CPT15,HMP08,HM09,LMP10]. Cependant, les algorithmes proposés se contentent de ne proposer qu'une solution minimale.

Un premier objectif de ce travail de thèse consistera à étudier les problèmes de complétions minimales du point de vue de l'énumération. Dans la mesure où trouver une solution optimale est un problème difficile (i.e. pas d'algorithme efficace pour résoudre le problème), l'objectif sera de pouvoir lister des complétions minimales (au sens de l'inclusion) de manière rapide (en un temps polynomial). En effet afin de considérer une possible complétion et déterminer si cette dernière est satisfaisante, nous devons pouvoir obtenir les solutions successives de manière rapide (i.e. le temps entre l'obtention de deux complétions doit idéalement être borné par un polynôme qui dépend de la taille de la donnée initiale).

En plus de la complétion, le travail de thèse s'intéressera aussi à la réduction minimale d'un graphe pour qu'il devienne *intéressant*. Cette opération, qui consiste à supprimer des arêtes d'un graphe afin de le rendre conforme à une propriété donnée, est semblable à la notion de réduction de la dimension d'une matrice.

Pour certaines propriétés nous savons déjà que les problèmes de complétion et de réduction de graphes sont équivalents. Nous commencerons notre étude par l'analyse de ces propriétés là.

Cette problématique a émergé récemment pour des problèmes de bases de données [CKK17]. Néanmoins les techniques utilisées dans chacun des résultats sont assez différentes. L'objectif de cette action consisterait également à concevoir un framework générique pour unifier la résolution de ce genre de problèmes.

Références bibliographiques

- [CKK17] Nofar Carmeli, Batya Kenig, and Benny Kimelfeld, *Efficiently enumerating minimal triangulations*, Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017, Chicago, IL, USA,
- [CPT15] Christophe Crespelle, Anthony Perez, and Ioan Todinca, *An $O(n^2)$ time algorithm for the minimal permutation completion problem*, Graph-Theoretic Concepts in Computer Science - 41st International Workshop, **WG 2015**, Garching, Germany, June 17-19, 2015, Revised Papers, 2015, pp. 103–115.
- [CT13] Christophe Crespelle and Ioan Todinca, *An $O(n^2)$ -time algorithm for the minimal interval completion problem*, **Theor. Comput. Sci.** 494 (2013), 75–85.
- [HMP08] Pinar Heggernes, Federico Mancini, and Charis Papadopoulos, *Minimal comparability completions of arbitrary graphs*, **Discrete Applied Mathematics** 156 (2008), no. 5, 705–718.
- [HM09] Pinar Heggernes and Federico Mancini, *Minimal split completions*, **Discrete Applied Mathematics** 157 (2009), no. 12, 2659–2669.
- [LMP10] Daniel Lokshantov, Federico Mancini, and Charis Papadopoulos, *Characterizing and computing minimal co-graph completions*, **Discrete Applied Mathematics** 158 (2010), no. 7, 755–764.

Présentation du co-encadrant étranger (c.f., cv joint en annexe)

Ce travail de thèse sera co-encadré par Takeaki Uno, professeur au NII (National Institute of Informatics), Tokyo, Japon et directeur de la division *Principles of Informatics Research* du NII.

Takeaki Uno est un scientifique de tout premier plan, reconnu au niveau international pour ses travaux autour de l'algorithmique et de la fouille de données. Il a reçu ces dernières années plusieurs prix, notamment pour ses travaux sur les algorithmes efficaces de fouille de données.

Le professeur Takeaki Uno est porteur actuellement d'un grand projet national japonais financé par *JST CREST*, sur une technique nouvelle technique de lissage des données appelée : *Data Particlization for Next Generation Data Mining*.

Takeaki Uno est l'un des collaborateurs du LIMOS qui a visité à plusieurs reprises notre laboratoire, ce qui nous a permis de réaliser plusieurs publications de haut niveau. Il a eu aussi des visites des chercheurs du LIMOS à NII en particulier Vincent Limouzy.

Apport du co-encadrement étranger au travail de thèse

Takeaki Uno est un expert des algorithmes d'énumération et de fouille de données. Il a publié environ 42 articles dans des revues classés A et une centaine de conférences. Takeaki Uno a déjà collaboré avec les membres de l'action et cette collaboration est toujours active.

Par ailleurs, Takeaki Uno a développé plusieurs programmes considérés parmi les plus performants dans le domaine de la fouille des données. On peut citer les références suivantes :

- Programs of the algorithms for enumeration problems and data mining problems (<http://research.nii.ac.jp/~uno/codes.htm>)
- Hypergraph Dualization Repository (<http://research.nii.ac.jp/~uno/dualization.html>)

Cette expertise sera extrêmement utile pour le travail de thèse envisagé. Dans ce cadre, le professeur Takeaki Uno effectuera des séjours réguliers au LIMOS pendant la durée de la thèse (un séjour d'un mois/an pendant 3 ans). Par ailleurs, il est prévu des séjours du doctorant et du porteur de l'action au NII dans l'équipe du professeur Takeaki Uno. Ce programme d'échange intensif permettra la mise en place d'une collaboration étroite s'appuyant sur une interaction continue entre l'équipe Clermontoise, incluant le doctorant, et le professeur Takeaki Uno, co-encadrant étranger de la thèse.