

Projet GIRAFON

Protection de la vie privée dans le cadre de l'interrogation, la fouille et l'analyse de grands graphes

LIFO

6 Rue Léonard de Vinci
45067 Orléans, France

LI

64 avenue Jean Portalis
37200 Tours, France

A PROPOS DU PROJET

Le projet GIRAFON "Grands graphes : interrogation, fouille et analyse" est un projet de recherche d'initiative académique financé par la région Centre Val de Loire et mené à bien au sein du Laboratoire d'Informatique Fondamentale d'Orléans (LIFO) et du Laboratoire d'Informatique de Tours (LI). Il a pour objectif la conception d'algorithmes génériques de fouilles de grands graphes prenant en compte la protection de la vie privée des individus.

Contact : cedric.eichler@insa-cvl.fr

DESCRIPTION DU PROJET

Nous vivons désormais entourés de capteurs, nous créons des informations en continu et nous laissons en permanence des traces informatiques de nos activités. Le traitement et l'analyse de ces masses de données, "BigData" ou mégadonnées, offrent des perspectives innombrables et encore largement inexplorées. La maîtrise des mégadonnées a, et aura de plus en plus, un impact très important aux niveaux sociétal, économique et commercial. Des domaines extrêmement divers sont concernés : la santé (épidémiologie, génomique...), les réseaux énergétiques complexes, les villes intelligentes, la prévision et gestion des risques environnementaux, les réseaux sociaux, le linked open data (web sémantique), etc. Malgré la multiplicité et l'hétérogénéité des domaines d'application, les méta-modèles exploités sont relativement homogènes. En effet, dans de nombreux cas, les données et systèmes peuvent être représentées sous forme de graphes.

GIRAFON est un projet de recherche en amont qui vise à la conception d'algorithmes de fouille de données (data-mining) au sein de grands graphes respectant trois principes fondamentaux : ré-utilisabilité, efficacité et respect de la vie privée. Cette présentation portera en particulier sur ce dernier point.

a) Ré-utilisabilité et efficacité: De très nombreux travaux ont été menés dans le contexte de la fouille de grands graphes. Néanmoins, très peu de travaux ont visé à proposer des méthodes et algorithmes génériques, ce qui implique de développer de nouvelles structures de données et algorithmes pour tout nouveau problème posé et reste très coûteux. Le développement de telles méthodes nécessite la spécification d'un langage de requête expressif et riche sur lequel elles pourront s'appuyer. Pour qu'un tel langage soit utilisable, il est nécessaire de concevoir des environnements d'exécution distribués accompagnés de bibliothèques de programmation dédiées.

b) Respect de la vie privée: L'analyse de mégadonnées permet d'acquérir des informations sur des indicateurs généraux et le comportement global des personnes impliquées dans l'étude, mais également des données précises et potentiellement gênantes sur chaque individu. Même dans des études sur des sujets qui peuvent paraître anodins, comme les recherches sur les moteurs de recherches, l'expérience a démontré l'importance de prendre en compte la protection de la vie privée des individus dont les données sont publiées. À cet effet, nous nous proposons d'allier aux algorithmes de GIRAFON les principes d'anonymisation et d'exposition minimale garantissant, sans impacter l'analyse globale, d'une part que les données exploitées ne peuvent être liées à un individu et de l'autre que chaque individu fournit le minimum de données personnelles.

Durant cette présentation, nous introduirons en particulier un nouvel algorithme de calcul de métrique d'influence dans les réseaux sociaux respectant la vie privée de ses membres. Nous étudierons les données personnelles qu'il expose directement, les informations déductibles par attaque et sa robustesse face à des attaques endogènes.