

Extraction et analyse de graphes signés pour détecter la corruption dans les marchés publics

Mots-clés

extraction de données, extraction de graphes, graphes signés, marchés publics, open data, partitionnement de graphes

Encadrants

- Rosa Figueiredo <rosa.figueiredo@univ-avignon.fr>
- Vincent Labatut <vincent.labatut@univ-avignon.fr>

Contexte

L'ouverture massive des données publiques recouvre une importance particulière dans le cadre des marchés publics, et la récente réforme de ces marchés introduit explicitement l'open data dans la commande publique. Ce sont donc dès aujourd'hui une masse importante, et à court terme l'intégralité des données relatives aux marchés publics qui peuvent faire l'objet d'analyses qui jusqu'à présent se heurtaient à l'inexistence ou à la partialité de jeux de données. Que ce soit par les montants en jeu (la commande publique correspond à presque 15 % du PIB, 200 milliards d'euros par an pour les seuls marchés publics), ou par les bénéfices économiques et sociétaux escomptés derrière cet impératif de transparence, le traitement et l'analyse de ces données constituent un enjeu tant académique que sociétal majeur.

Le projet motivant la présente offre de stage s'insère dans les deux axes transversaux du LIA : *Systèmes Complexes* et *Société Numérique*. À l'intersection de l'informatique, des sciences économiques et des sciences juridiques, ce projet vise à collecter, traiter et analyser les données ouvertes relatives aux marchés publics français, afin d'élaborer un outil basé sur les méthodes de graphes signés pour la détection des risques de corruption qui peuvent exister entre acheteurs publics et entreprises. La prise en compte de la nature duale des liens unissant entreprises et donneurs d'ordres (relation contractuelle normale ou corruption) est fondamentale à la bonne modélisation du système étudié, et nécessite d'utiliser des réseaux signés. Il s'agit d'un type de réseau bien moins exploré que les réseaux non-signés classiques, et permettant d'inclure dans un même modèle des relations antagonistes.

L'élaboration d'outils de détection automatique des risques de corruption dans les marchés publics ("red flagging") n'est en soit pas totalement nouvelle, et se développe depuis quelques années déjà [1], notamment à l'échelle européenne [2]. Outre qu'aucun outil n'a à ce jour été adapté au cadre juridique et aux données des marchés publics français, les méthodes appliquées présentent un limite importante. Les approches existantes se focalisent sur des données individuelles, caractérisant de façon indépendante clients et fournisseurs, et ignorent les informations relationnelles, correspondant aux interdépendances et interactions entre ces différents acteurs. De ce fait, les outils produits passent à côté d'un certain nombre de propriétés émergentes, i.e. présentes à une granularité plus élevée que l'acteur isolé.

Travail demandé

Le stagiaire **collectera et structurera l'ensemble des données** issues des marchés publics français, telles que proposées par le *Bulletin Officiel des Annonces des Marchés Publics* (Boamp.fr) (annonces et avis d'attribution)/ Le site propose une API dont le stagiaire devra tirer parti. Les données *a priori* pertinentes à extraire seront identifiées et enrichies à partir à la fois de l'état de l'art académique (analyse juridique, approche économique théorique et empirique) et de l'expertise d'un Professeur en économie, P.-H. Morand (LBNC – UAPV), qui

participe à notre projet. En fonction du déroulement du projet, il est envisageable d'étendre cette tâche à d'autres sources de données, telles que [TED](#) (l'équivalent européen du BOAMP).

Le stagiaire utilisera **les graphes signés afin de modéliser et analyser les réseaux** complexes formés par les entreprises fournisseurs et les collectivités donneurs d'ordres. Il s'agit de graphes dont les liens sont caractérisés par un signe, pouvant être positif ou négatif, afin de représenter des relations antagonistes. Ces graphes seront construits à partir des données obtenues à la première étape. Plusieurs méthodes sont possibles pour effectuer cette opération, que le stagiaire devra mettre en œuvre et comparer.

Une fois les graphes obtenus, ils seront principalement utilisés pour deux tâches. La première est leur **analyse via des outils descriptifs** classiques dans le champ des réseaux complexes (taille, densité, transitivité, centralités diverses...). La seconde vise à **résoudre des problèmes de partitionnement définis sur les graphes signés** en utilisant les méthodes de résolution existantes plus adaptées [3,4] aux réseaux obtenus. Avec les solutions obtenues émergeront des groupes d'acteurs (acheteurs publics et entreprises fournisseurs) susceptibles d'être liés entre eux par des pratiques délictueuses.

Références

- [1] Fazekas, Mihály, et István János Tóth. [New ways to measure institutionalised grand corruption in public procurement](#), U4 Brief, n°9 (2014).
- [2] Ferwerda, Joras, et Ioana Deleanu. [Identifying and Reducing Corruption in Public Procurement in the EU](#). European Commission, OLAF, 2013.
- [3] Figueiredo, Rosa et Yuri Frota. [The maximum balanced subgraph of a signed graph: Applications and solution approaches](#). European Journal of Operational Research, 236(2) : 473-487, 2014.
- [4] Labatut, Vincent. [Generalized Measures for the Evaluation of Community Detection Methods](#), International Journal of Social Network Mining, n°2(2015):44–63.