



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE
MINISTÈRE DE L'ENSEIGNEMENT
SUPÉRIEUR ET DE LA RECHERCHE

PROPOSITION SUJETS DE THESES

CONTRATS DOCTORAUX

2017-2020

Appel ciblé (merci de cocher la case correspondante):

- Contrat doctoral ministériel ED 536
- Contrat doctoral ministériel ED 537
- Contrat doctoral fléché FR Agorantic
- Contrat doctoral Fondation – Pierre Bergé

(Utiliser la fiche type transmise par la SFR Tersys pour le contrat doctoral Fondation Jean-Henri Fabre- Ville d'Avignon)

Directeur de thèse : Pierre-Henri Morand Email : pierre-henri.morand@univ-avignon.fr
EA 3788 LBNC - Laboratoire Biens, Normes, Contrats

Co-directeurs (ADR) : Vincent Labatut Email : vincent.labatut@univ-avignon.fr
Rosa Figueiredo Email : rosa.figueiredo@univ-avignon.fr
EA 4128 LIA - Laboratoire Informatique d'Avignon

Correspondants : Pierre-Henri Morand Tél. : 04 90 16 27 51
Vincent Labatut Tél. : 04 90 84 35 25
Rosa Figueiredo Tél. : 04 90 84 35 73

Titre en français : Outils automatiques et open data pour la détection de corruption dans les marchés publics

Titre en anglais : Automatic tools and open data for corruption detection in public procurements

Mots-clés : Graphes signés, Réseaux complexes, Open data, Détection de corruption, Marchés publics, Optimisation

Co tutelle : Oui - Non

Profil du candidat :

- Détenteur d'un Master d'informatique.
- Maîtrise du français et bon niveau d'anglais.
- Notions de recherche opérationnelle.
- Maîtrise des langages de programmation C++, Java et/ou R.

Domaine / Thématique : informatique appliquée à l'économie

Contexte et enjeux

Pesant aux alentours de 15% des PIB nationaux dans la plupart des pays occidentaux, les marchés publics et leur bonne gestion représentent un enjeu économique majeur. Strictement encadrés par la loi, contrôlés, ils constituent malgré tout, par l'ampleur des sommes en jeu, par la multiplicité des acteurs, et par le secret qui entoure en partie les choix finaux d'attribution des marchés, l'instrument le plus susceptible d'engendrer un usage frauduleux des ressources publiques à grande échelle.

Très tôt, les sciences économiques se sont intéressées à la vulnérabilité de la commande publique aux pratiques de favoritisme ou de corruption. Cependant, si cette littérature a mis en évidence l'optimalité de certaines stratégies déviantes [Arozamena, & Weinschelbaum 2009], leur détection et leur mise à jour restent des enjeux majeurs [Bajari & Summers 2002].

En effet, même si les réglementations encadrent les procédures qui peuvent être utilisées par les adjudicateurs publics, le choix des règles précises, la plus ou moins grande publicité préalable, et la pondération accordée aux différents critères de sélection, rendent difficile de distinguer l'attribution qui reflète le choix de l'offre économiquement la plus avantageuse de celle qui découle de favoritisme, par essence secret. Et cette difficulté de détection renforce en retour la vulnérabilité de la commande publique à ces pratiques.

Cependant, l'évolution récente vers plus de transparence des données publiques (open data) est souvent présentée comme un instrument de normalisation des pratiques publiques. C'est en particulier le cas des données des marchés publics, longtemps peu diffusées ou difficilement exploitables mais dont la publicité et la large diffusion des résultats est aujourd'hui pressentie comme une arme efficace de lutte a priori contre les pratiques frauduleuses [Boehm & Olaya 2006]. Pour autant, la simple publication de données n'est pas en soi suffisante. Encore convient-il de disposer d'instruments permettant d'en extraire des éléments d'attention pertinents, d'autant que la somme des données relevant des marchés publics, à l'échelle d'un pays, est massive et relève de ce que l'on appelle le *Big Data*. Outil implicite de *Sunshine Regulation*, une telle mise en évidence de pratiques suspectes peut également orienter les régulateurs dans leur politique de contrôle [Fazekas, Tóth & King 2013].

Le fond du problème porte sur les interactions entre les entreprises et les donneurs d'ordres impliqués dans les marchés publics, ce qui amène naturellement à une modélisation sous la forme de *réseaux* (au sens de la théorie des graphes). Qui plus est, ces entreprises et donneurs d'ordres peuvent être liés par des contrats relevant aussi bien de *bonnes* que de *mauvaises* pratiques : ceci peut notamment être déterminé grâce aux informations disponibles sur la nature de la relation contractuelle (publicité préalable, nombre de concurrents, choix d'une procédure négocié ou ouverte, nombres de clauses spécifiques etc...). La prise en compte de cette nature duale des liens unissant entreprises et donneurs d'ordres 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.

Objectif

L'objet de la thèse est de mobiliser l'outil des graphes signés afin d'analyser les réseaux complexes formés par les entreprises fournisseurs et les collectivités donneurs d'ordres. Plus précisément, à l'interface des sciences informatiques et des sciences économiques, la thèse

a pour but de résoudre les problèmes de partitionnement de graphes signés liés à l'équilibre structurel des réseaux représentant la commande publique, de manière à faire émerger des groupes d'acteurs (acheteurs publics et entreprises fournisseurs) susceptibles d'être liés entre eux par des pratiques délictueuses.

Méthode

Dans un réseau *signé*, chaque lien est associé à un signe soit *positif*, soit *négatif*. Le concept a été défini par Heider [1946] dans le but de décrire les relations affectives entre des personnes appartenant à un même groupe social. Au cours des dernières décennies, les réseaux signés ont été adoptés comme outil de modélisation par de nombreux chercheurs en sociologie, mais aussi par des chercheurs issus d'autres domaines et travaillant sur des systèmes *polarisés*, i.e. caractérisés par des relations antagonistes.

La notion d'*équilibre structurel* est centrale à l'étude des réseaux signés. Elle stipule qu'un tel réseau est équilibré s'il peut être partitionné en deux sous-groupes (ou plus) dont les membres sont solidaires au sein d'un même groupe, mais mutuellement hostiles quand issus de groupes différents. Cette solidarité et cette hostilité sont respectivement représentées ici par des liens positifs et négatifs.

Déterminer si un réseau est équilibré constitue un problème de partitionnement de graphe, pouvant être résolu via des méthodes d'optimisation. La littérature montre que la notion d'équilibre structurel peut être étendue de différentes façons, amenant autant de variantes du problème de partitionnement, et requérant autant de formulations et de méthodes de résolution. Cependant, lorsqu'il représente un système du monde réel, un réseau est rarement parfaitement équilibré, quelle que soit la façon dont cet équilibre est défini. Quantifier le niveau d'équilibre du réseau constitue alors un défi majeur, qui implique de définir une mesure de l'équilibre puis de résoudre le problème d'optimisation associé.

Malgré l'omniprésence des systèmes réels polarisés, il y a peu d'outils permettant d'étudier les graphes signés. Ceci est principalement dû 1) au fait que la communauté scientifique a concentré ses efforts sur les graphes non-signés, qui sont conceptuellement plus simples ; 2) à un manque de données réelles permettant de concevoir et de valider les méthodes d'analyse ; et 3) au fait que les personnes qui développent ces outils et celles qui seraient susceptibles de les appliquer à des problèmes concrets travaillent de façon très séparée.

Le travail proposé dans ce sujet de thèse vise directement à contribuer au traitement du premier problème identifié, en proposant de nouveaux outils d'analyse des réseaux signés. De plus, la démarche interdisciplinaire adoptée a pour objectif de permettre de résoudre aussi les deux autres problèmes, en travaillant sur des *données réelles* et en attaquant un *problème bien concret* d'économie, dont le traitement est susceptible d'avoir des effets dans la vie réelle.

Le candidat utilisera donc une approche *ascendante*, i.e. partant des données disponibles et des problèmes d'économie posés, afin d'orienter la conception de méthodes d'analyse. Celles-ci seront basées sur des formulations de l'équilibre structurel spécifiquement définies pour correspondre au problème d'identification de pratiques délictueuses dans les marchés publics. Le développement des méthodes de partitionnement correspondantes constituera un des aspects les plus importants de la thèse, notamment dans un contexte Big Data. Enfin, une part non-négligeable du travail sera consacrée au développement de méthodes permettant d'extraire des réseaux signés à partir des données brutes disponibles, ainsi qu'à la définition de modèles aléatoires



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE
MINISTÈRE DE L'ENSEIGNEMENT
SUPÉRIEUR ET DE LA RECHERCHE

permettant de produire des réseaux reproduisant les caractéristiques des données réelles disponibles.

Organisation

On peut distinguer quatre étapes dans le travail proposé : 1) extraire les réseaux signés à partir des données brutes ; 2) identifier des variantes de l'équilibre structurel pertinentes pour le problème d'économie considéré ; 3) formaliser les problèmes de partitionnement de graphe correspondant et proposer des méthodes de résolution efficaces ; et 4) évaluer la qualité de ces méthodes en termes computationnels, mais aussi par rapport à la capacité à identifier les situations délictueuses.

Références bibliographiques

- L. Arozamena & F. Weinschelbaum. *The effect of corruption on bidding behavior in first-price auctions*. *European Economic Review*, 53(6):645-657, 2009.
- P. Bajari & G. Summers. *Detecting collusion in procurement auctions*. *Antitrust Law Journal*, 70(1):143-170, 2002.
- F. Boehm & J. Olaya. *Corruption in public contracting auctions: the role of transparency in bidding processes*. *Annals of Public and Cooperative Economics*, 77(4):431-452, 2006.
- M. Fazekas, I. J. Tóth & L. P. King. *Anatomy of grand corruption: A composite corruption risk index based on objective data*. Corruption Research Center Budapest, Working Paper n° CRCB-WP/2013:02, 2013.
- R. Figueiredo & Y. Frota. *The maximum balanced subgraph of a signed graph: Applications and solution approaches*. *European Journal of Operational Research*, 236:473-487, 2014.
- R. Figueiredo & G. Moura. *Mixed integer programming formulations for clustering problems related to structural balance*. *Social Networks*, 35:639-651, 2013.
- F. Heider. *Attitudes and cognitive organization*. *Journal of Psychology*, 21:107-112, 1946.
- V. Labatut. *Generalized measures for the evaluation of community detection methods*. *International Journal of Social Network Mining*, 2(1):44-63, 2015.
- K. T. Macon, P. J. Mucha & M. A. Porter. *Community structure in the United Nations general assembly*. *Physica A*, 391(1-2):343-361, 2012.
- F. Maréchal & P.-H. Morand. *The public release of information in first-price sealed-bid auctions*. *Review of Economic Design*, 16(4):323-330, 2012.
- I. Mendonça, R. Figueiredo, V. Labatut, & P. Michelon. *Relevance of Negative Links in Graph Partitioning: A Case Study Using Votes From the European Parliament*. In: 2nd European Network Intelligence Conference (ENIC). Karlskrona, SE, pp. 122-129, 2015.
- G. K. Orman, V. Labatut & H. Cherifi. *Comparative evaluation of community detection algorithms: A topological approach*. *Journal of Statistical Mechanics*, 08:P08001, 2012.