

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



DETECTING CORRUPTION COLLUSION AND FRAUD

T E S I S

QUE PARA OBTENER EL TÍTULO DE

MAESTRO EN CIENCIA DE DATOS

PRESENTA

CARLOS EDUARDO PETRICIOLI ARAIZA

ASESOR: DR. ADOLFO JAVIER DE UNÁNUE TISCAREÑO

México, D.F.

2016

*(This page is intentionally left blank)*

“Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada ‘XXXXX’ otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación”.

CARLOS EDUARDO PETRICIOLI ARAIZA

---

FECHA

---

FIRMA

*Aquí van los agradecimientos*

*Aquí va la dedicatoria*

# Contents

---

|          |                            |          |
|----------|----------------------------|----------|
| <b>1</b> | <b>Introduction</b>        | <b>3</b> |
| 1.1      | The World Bank . . . . .   | 3        |
|          | <b>Bibliography</b>        | <b>5</b> |
|          | Other references . . . . . | 5        |

## List of Figures

---

## List of Tables

---



## Web page

---

[corruption.carlospetricioli.com](http://corruption.carlospetricioli.com)

# DETECTING CORRUPTION COLLUSION AND FRAUD<sup>1</sup>

CARLOS PETRICIOLI<sup>2</sup>

## Abstract

The World Bank Group lends billions of dollars each year to fund development projects in its efforts to reduce global poverty. This project helps investigators at the Bank search for patterns of collusion, corruption, and fraud in its contracts data, using models of contract-specific risk. Developing an automated approach to detecting these offenses can help the World Bank efficiently target future investigations.

The Problem Contractors providing goods and services on World Bank projects are typically hired through a competitive bidding process. Occasionally, prospective contractors influence the competitive system by colluding with other contractors, bribing government officials, or otherwise manipulating the bidding process. These offenses have far-reaching effects on the price and quality of contract delivery. The World Bank is committed to detecting instances of collusion, corruption, and fraud in order to maximize its global impact.

Data To develop contract-level risk models, I incorporated data from multiple sources:

- Historical data on over 300,000 major contracts funded by World Bank loans from the past 20 years, including such features as company name, country, sector, and total award amount.
- Annual economic development indicators, collected by the World Bank, for countries and industries within them.
- World Bank investigations data, covering companies and projects investigated for collusion, corruption or fraud in the past. Includes specific allegations and case outcomes.

### Building Tools for Proactive Investigation

To approach the problem, an interactive dashboard was developed for World Bank investigators to track a company's activity across countries, sectors, and time. Using this tool, investigators can: - Track contract awards companies have received, including under different names (e.g. ACME, Inc. vs. ACME Co.) - View a risk score for each World Bank contract, as calculated by our contract risk model - Visualize the immediate neighborhood of the company in its co-award network

---

<sup>1</sup> Last Update: December 10, 2015.

<sup>2</sup>I especially thank to Rayid Ghani (Ph.D. Carnegie Mellon University) and to Eric Rozier (Ph.D. University of Illinois Urbana-Champaign) for their mentorship through out the elaboration of this work. I also appreciate the collaboration and comments of Jeff Alstott (Ph.D. Cambridge), Dylan Fitzpatrick (M.S. Carnegie Mellon), Misha Teplitskiy (Ph.D. University of Chicago), Elizabeth Wiramidjaja (World Bank).

carpetri@gmail.com

XXX.carlospetricioli.com

carlospetricioli.com

**Company Name Disambiguation** Company names are represented by text strings in the data, and a single company may be represented in several very different ways (e.g. ACME Inc. vs. A.C.M.E. Co.) Company names were reconciled by querying each name on Google and comparing their top ten URL results. Names that had at least 7 links in common were considered to be a single company.

**Co-award networks** A company that works with companies or on projects that have been investigated by the World Bank are more likely to be investigated themselves. Companies that are on many projects together may know each other better and have the opportunity to collude.

**Evaluating Contract Risk** This project generates features tracking companies' historical involvement on World Bank projects within specific countries and sectors, as well as co-award network features for each company. We trained a binary classifier separating past contracts that were investigated by the World Bank from those that were not investigated. We evaluated and compared models using precision, recall, and area under ROC curve. A random forest provided the best results across all metrics on a held-out test set.

**Conclusions and Future Work** - Current data collected by World Bank on contracts is sufficient to forecast risk on future contracts. - These risk forecasts can allow World Bank investigators to be more proactive in determining which companies, projects and contracts to examine. - Future analysis will identify separate risk levels for fraud, collusion and corruption.

**Topics:** XXX, XXX .

---

## Chapter 1.

### Introduction

---

The World Bank Group lends billions of dollars each year to fund development projects in its efforts to reduce global poverty. This project helps investigators at the Bank search for patterns of collusion, corruption, and fraud in its contracts data, using models of contract-specific risk. By developing an automated approach to detect these offenses this project can help the World Bank efficiently target future investigations.

#### 1.1 The World Bank

The World Bank was created in 1944 as one institution and has evolved into an association of five development institutions. It is composed by the International Bank for Reconstruction and Development (IBRD), the International Development Association, the International Finance Corporation (IFC), the Multilateral Guarantee Agency (MIGA), and the International Centre for the Settlement of Investment Disputes (ICSID). In its interior it was formed by a large group of engineers and financial analysts who work from Washington, D.C. Now, they have a very heterogenous and distict staff that includes economist, public policy experts, sector experts and social scientists and about one third of the workers has spread to the country officies

[Web Page: *History of the World Bank*] Bivand, Pebesma, and Gómez-Rubio, 2011

Reconstruction remains an important part of our work. However, at today's World Bank, poverty reduction through an inclusive and sustainable globalization remains the overarching goal of our work.

Contractors providing goods and services on World Bank projects are typically hired through a competitive bidding process. Occasionally, prospective contractors influence the competitive system by colluding with other contractors, bribing government officials, or

otherwise manipulating the bidding process. These offenses have far-reaching effects on the price and quality of contract delivery. The World Bank is committed to detecting instances of collusion, corruption, and fraud in order to maximize its global impact.

# Bibliography

---

## Books

Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gómez-Rubio (2011). *Applied Spatial Data Analysis with R*. Springer, Baltimore, MD, USA.

## Other references

The World Bank (2015). *History of the World Bank*. <http://www.worldbank.org/en/about/history>. (Visited on 12/09/2015).