
Detecting Anomalous Business Ownership with Graph Convolutional Neural Networks

Project Proposal

Dominic Thorn

Contents

Notes for the Reviewer	3
Introduction	3
Subject Overview	3
Project Title	3
Aims, Objectives and Research Questions	3
NOTES	3
Aims	4
Literature Review	4
Methods	4
NOTES	4
Data	4
Research Instruments and Tools	4
Ethical Considerations	4
Anticipated Outcomes	4
Project Plan	4
Roadmap	4
Data Acquisition	5
Risks and Challenges	5

Notes for the Reviewer

A key focus in my position as Lead Data Scientist at Quantexa is the application of graph analytics for the detection of fraud. Traditional Machine Learning techniques struggle to identify anomalous entities in business networks, as their main giveaway is their relationship to neighbouring entities.

Through this project, I hope to demonstrate the effectiveness of Graph Convolutional Neural Network (GCN) models for identifying suspicious actors in business ownership networks.

Introduction

Subject Overview

In October of 2021, The International Consortium of Investigative Journalists (ICIJ) revealed the findings of their ‘Pandora Papers’ investigation, implicating hundreds of politicians, public officials, and businesses in efforts to conceal the true ownership of companies and assets around the world (ICIJ 2021). The intentions behind this secrecy range from legitimate privacy concerns to criminal activities, including money laundering, tax evasion, and fraud. According to a study by the European Commission, an estimated total of USD 7.8 trillion was held offshore in 2016. The share of this attributed to the European Union (EU) was USD 1.6 trillion, which corresponds to an estimated tax revenue loss to the EU of EUR 46 billion (“Estimating International Tax Evasion by Individuals” 2019).

Project Title

The proposed title for this project is “Detecting Anomalous Business Ownership with Graph Convolutional Neural Networks”.

Aims, Objectives and Research Questions

NOTES

- reasons for study
 - existing studies do not test GCN on business ownership graphs
 - important for detecting fraud, specifically money laundering
 - traditional methods do not take into account contextual data

Aims

The aim of this project is to assess the performance of Graph Convolutional Neural Network (GCN) models in identifying anomalous entities in a business ownership graph.

Literature Review

Methods

NOTES

- split graph into weakly connected components (define term)
- select random node(s) from outside of the connected component as target
- impossible for traditional methods to identify as anomalous as features are indistinguishable from others
- can attempt traditional anomaly detection techniques on individual connected component as a baseline
 - random forest
 - gradient boosted tree
 - K Nearest Neighbours
 - logistic regression
 - GraphGym
 - Tuned GCN (Optuna or similar)

Data

Research Instruments and Tools

Ethical Considerations

Anticipated Outcomes

Project Plan

Roadmap

- Data acquisition

- Data understanding
- Data preparation
- Feature engineering
- Preprocessing
- Modeling
- Evaluation

Data Acquisition

Risks and Challenges

“Estimating International Tax Evasion by Individuals.” 2019. European Commission. https://ec.europa.eu/taxation_customs/system/files/2019-10/2019-taxation-papers-76.pdf.

ICIJ. 2021. “Offshore Havens and Hidden Riches of World Leaders and Billionaires Exposed in Unprecedented Leak - ICIJ.” <https://www.icij.org/investigations/pandora-papers/global-investigation-tax-havens-offshore/>.