Mapping the laundromat: A network analysis of money laundering in the United Kingdom

By

Ian Goodrich

Submitted to

Central European University

School of Public Policy

In partial fulfilment of the requirements for the degree of

Master of Arts in Public Policy (one year)

Supervisor: Mihály Fazekas

Budapest, Hungary

2019

Authorship Declaration

I, Ian Goodrich, hereby declare that I am the sole author of this thesis. To the best of my knowledge this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted as part of the requirements of any other academic degree or non-degree program, in English or in any other language. This is a true copy of the thesis, including final revisions.

March 4, 2019

Acknowledgements

I'm writing this having returned to university after over a decade out in the wilderness of work. It's been a life-changing experience, and one that I've thoroughly enjoyed, but there's no chance I'd have done it without the faith others have shown in me.

First and foremost, I must thank my beautiful and brilliant wife and partner, Dion. For many years, she has provided the love, support and encouragement which has enabled everything I've done. She was the first to suggest I return to university, she packed up her life to be with me in Budapest, and she has been consistently kind, caring and patient as I've completed my studies. Thank you, baby, I love you.

If without Dion, this couldn't have happened, then without Dustin Gilbreath, it wouldn't have happened. Both the decision to return to study and the choice of CEU can be traced back to a conversation one night in a seedy Tbilisi bar with Dustin – so it's his fault. His passion for his field is infectious, his generosity boundless, and his insights invaluable. I'm also particularly grateful for his explanation of sample weighting.

I am also indebted to Sam Leon of Global Witness, whose dataset is very much the backbone of this study; my brother Steve, who got me interested in money laundering and who has been kind enough to produce a lot of useful literature on the subject; and to Mihaly Fazekas for reining in some of my wilder ideas.

Honourable mentions to Till and Hans, for their support to my initial CEU application, and for generally being top blokes; my Dad, who has always encouraged this sort of nonsense; my Mum, who I wish could have read this; and my other brothers, Neil and Nic, because it would be utterly unfair to thank Steve without mentioning them.

Also, to Merlin, who's a very good dog, sometimes.

Abstract

Money laundering enables crime, saps public resources, and undermines the state in rich and poor countries alike. Investigations into money laundering, once traditionally the domain of government, financial institutions and international bodies, are increasingly being undertaken by journalists, NGOs and academics. Using the case study of the United Kingdom, this study explores the use of network analysis for the study money laundering using open data. Using a custom-built Python module, this study analyses money laundering risk in over 2,000 corporate networks and 130,000 companies from the UK Companies House registry.

At the company level, it finds that high risk companies may be younger and more likely to use the Limited Liability Partnership structure. At a network level, it provides tentative evidence that companies used for money laundering may cluster with other such companies, and that network size and triangular graph structures may have potential for use in risk prediction using open data.

Word Count

The core component of this document (Chapters 1-7) contains 13,139 words, excluding footnotes and figures.

Software and Source Code

This study made use of a variety of software tools to acquire, process and analyse the data presented herein. Work was primarily undertaken in Python, making extensive use of the NetworkX, Pandas, StatsModels and chpy modules, with the Gephi and Visio programmes used for visualisation.

Data used for this study includes Companies House bulk data products (see 4.2), a dataset provided by Global Witness (3.4), and data sampled using the chpy module (4.1). Taken together these datasets exceed 5 gigabytes of disk space, and as such are not submitted with this document.

Data cleaning and analysis source-code and output is provided as an appendix to this document (Appendix I). Source code for the chpy module is available on GitHub.[1] Sampling source code and all other data used for the purpose of this study are available upon request from the author, who may be contacted by email at goodrichian@gmail.com.

[1] Goodrich, *Chpy: Build Networks from the Companies House API*. https://github.com/specialprocedures/chpy

Table of Contents

Accronyms

Accronym	Meaning
4MLD	4th EU Money Laundering Directive
AML	Anti-Money Laundering
API	Application Programming Interface
CTF	Counter-Terrorism Financing
FATF	Financial Action Task Force
GDP	Gross Domestic Product
ICIJ	International Consortium of Investigative Journalists
LLP	Limited Liability Partnership
LTD	Limited Company
NGO	Non-Governmental Organisation
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PSC	Person of Significant Control
SLP	Scottish Limited Liability Partnership
UK	United Kingdom
UNDOC	UN Office on Drugs and Crime
US	United States
USD	United States Dollars

Figures

Tables

1 Introduction

1.1 Civil society and the fight against money laundering

Money laundering is the process making “dirty” money “clean”. From officials engaged in embezzlement and bribery to international criminal and terrorist networks, money that has been obtained through illegal means must be “laundered” before it may be used in the open economy.

Criminal activity is constrained by its ability to make its proceeds available for legitimate expenditure: high-end property, investments and luxury goods may only be made available to illegitimate actors once the source of their funds is concealed. Laundering thus enables fraud, theft and trafficking; smuggling, sanctions-busting and arms dealing. It supports the theft of public funds by corrupt officials, and facilitates the operations of criminal and terrorist groups, increasing global security risks. As such, Anti Money Laundering (AML) measures are of a critical concern not only to policy-makers, but also to financial institutions, journalists, non-governmental organisations (NGO) and researchers.

Recent high-profile scandals have underscored the complicity of financial institutions in laundering activities, with the fall-out of such cases occasionally highlighting a reluctance on the part of policy-makers to enforce accountability.[2] In this context, civil society organisations have increasingly played a critical role in AML activity, by highlighting regulatory weaknesses, identifying high-risk cases and advocating for strengthened oversight.[3]

[2] Neate, “HSBC Escaped US Money-Laundering Charges after Osborne’s Intervention.”
[3] Global Witness, “The Companies We Keep”; Cowdock, “Hiding in Plain Sight,” November 2017; Cartin and Higgins, “Offshore in the UK.”

The role of United Kingdom (UK) companies in money laundering schemes has been highlighted by numerous high-profile investigations in recent years.[4] The UK has become attractive as a destination for dirty money, given the relative ease with which a company can be established, perceived reputational benefits of a UK address, proximity to high-value property, and crucially, the existence of legal loop-holes that enable the anonymisation of company ownership.[5]

Traditional AML actors – such as banks and law enforcement agencies – are able to leverage vast swathes of proprietary data, including transaction records and customer profiles, when investigating financial crime. A considerable body of scholarly work has been undertaken to support these efforts,[6] however, such approaches are not available to non-traditional actors (e.g. journalists and NGOs) for whom access to data is limited.

Civil society actors engaged in research on money laundering are increasingly making use of open data to effectively scrutinise companies and advocate for stronger enforcement mechanisms.[7] Scholarly research has the potential to enhance and support these efforts, with this study seeking to test new lines of enquiry using large-scale public data.

Network science[8] presents opportunities for the detection and study of money laundering without access to privileged data. Scope for network approaches to detection has

---

[4] Harding, "What Are the Panama Papers? A Guide to History's Biggest Data Leak"; Rankin, "European Parliament Calls for Investigation into 'Azerbaijani Laundromat'"; Garside, "Paradise Papers Leak Reveals Secrets of the World Elite's Hidden Wealth."

[5] Sharman, "Shopping for Anonymous Shell Companies."

6 Khan et al., "A Bayesian Approach for Suspicious Financial Activity Reporting"; Ngai et al., "The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature"; Chung et al., "Fighting Cybercrime: A Review and the Taiwan Experience"; Senator et al., "The FinCEN Artificial Intelligence System: Identifying Potential Money Laundering from Reports of Large Cash Transactions."

[7] Global Witness, "The Companies We Keep."

8 Euler, "Solutio Problematis Ad Geometriam Situs Pertinentis (The Seven Bridges of Konigsberg)"; Erdos, "Graph Theory and Probability"; Barabási, "Linked: The New Science of Networks."

been examined since the 1990s,[9] with a growing body of work employing such methods.[10] These approaches, however, also employ non-public data, leaving a gap in the literature that this thesis will seek to address.

1.2 Research overview

This study assesses money laundering risk in UK companies at the company and network level, seeking to identify characteristics of laundering networks and develop new tools for civil society AML investigation. It performs an empirical analysis of a sample of 131,174 companies and over 1,900[11] networks to test company and network attributes against a risk-flag dataset provided by the NGO, Global Witness.[12] Networks are constructed using a custom-built Python module[13] which utilises data from the Application Programming Interface (API) of Companies House, the UK company registrar.

Three hypotheses are tested, relating to company attributes, the distribution of money laundering risk across networks, and the potential for graph properties of networks to be used as indicators of money laundering risk. Firstly, this study examines the extent to which company age and type (i.e. legal structure) influence money laundering risk at both the company and network level. Secondly, it addresses the relationship between the risk of a "root company" – the company from which a network is built – and other entities in its

9 Sparrow, "The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects."
10 Fronzetti Colladon and Remondi, "Using Social Network Analysis to Prevent Money Laundering"; Drezewski, Sepielak, and Filipkowski, "The Application of Social Network Analysis Algorithms in a System Supporting Money Laundering Detection."
11 Networks may be sampled in ways that yield larger and smaller networks, a phenomenum described in this study as "depth" (see 4.3). This study samples 1,960 smaller, "shallower" corporate networks, of which a sub-sample of 83 "deeper" networks is taken.
12 Leon, *Analysis and Code to Accompany The Companies We Keep Briefing on the State of the UK's Register of Persons of Significant Control.*
13 Goodrich, *Chpy: Build Networks from the Companies House API*.

network. Finally, the study examines the role of graph size, density, bipartiteness, and the presence of triangular structures in money laundering risk.

Chapter 2 provides a background to the study, defining money laundering, examining the role of shell companies, and examining the nature of the problem in the United Kingdom. The chapter concludes with a background on corporate structures as networks, setting the stage for later empirical analysis.

Chapter 3 outlines the study's methodology in full, addressing choices relating to case selection, clarifying units of observation, and making explicit terminology. This chapter also provides a detailed description of the study's research questions, an overview of the primary dependent variable, and a discussion of independent variables at company and network levels.

Chapter 4 covers data collection, including a discussion of the operation of the Python module built for the purposes of the study. This chapter also discusses the sampling process, network "depth", and provides summary statistics for key variables.

Investigations into money laundering are fraught with methodological challenges, and as such, chapter 5 touches upon the study's limitations. It addresses the strengths and weaknesses of methodological choices, including risk-scoring as a dependent variable, the use of an API for bulk data collection, and the over-all quality of data used.

Chapter 6 presents the study's findings, outlining empirical tests of the three core hypotheses outlined above. It provides discussion of the implications of the tests for each hypothesis and identifies potential areas for further investigation. Finally, Chapter 7 briefly concludes with a discussion of the study's implications for future research.

2 Background

2.1 Money laundering

The Financial Action Task Force (FATF), the intergovernmental body coordinating international efforts against financial crime, describe money laundering as "…the processing of […] criminal proceeds to disguise their illegal origin."[14] The scale of the problem is by its nature difficult to measure, with the United Nations Office on Drugs and Crime estimating that as much as 2.7 percent of global GDP, or 1.6 trillion USD may have been "available" for money laundering in 2009.[15]

The money laundering process is often described using a three-step model, in which illicit funds are subject to *placement*, the process of integrating cash into the financial system; *layering*, through which money is moved through a series of bank accounts obscuring its origin; and *integration*, whereby funds are used in the "legitimate" market to make investments, and purchase goods and services. This process is reflected in the typologies of global institutions,[16] prosecution guidelines,[17] and in much of the academic literature.[18]

Estimations of the impact of money laundering are diverse, with negative implications highlighted for crime, inequality, financial stability, development and corruption, and property markets in financial centres.[19]

---

[14] Financial Action Task Force (FATF), "Money Laundering."
[15] UNDOC, "Estimating Illicit Financial Flows Resulting from Drug Trafficking and Other Transnational Organized Crimes."
[16] United Nations Office on Drugs and Crime (UNDOC), "Money-Laundering Cycle"; Financial Action Task Force (FATF), "Money Laundering."
[17] The Crown Prosecution Service, "Proceeds Of Crime Act 2002 Part 7 - Money Laundering Offences."
[18] Buchanan, "Money Laundering—a Global Obstacle"; Schneider and Windischbauer, "Money Laundering: Some Facts."
[19] Unger et al., "The Amounts and the Effects of Money Laundering"; Chaikin and Sharman, *Corruption and Money Laundering: A Symbiotic Relationship*; Aluko and Bagheri, "The Impact of Money Laundering on Economic and Financial Stability and on Political Development in Developing Countries"; Cowdock, "Faulty Towers"; Hendriyetty and Grewal, "Macroeconomics of Money Laundering"; Cowdock, "Kept in the Dark"; Cooley, Heathershaw, and Sharman, "Laundering Cash, White Washing Reputations."

These succinct definitions, large figures, straightforward steps, and intuitive impacts belie considerable diversity in assumptions, practices and processes between which scholars must distinguish prior to undertaking research. The global anti-money laundering regime as we understand it today is in reality very young, having risen to prominence in the context of drug trafficking in the 1980s. In this context, many of the understandings outlined above are contested, and there is a strong and growing body of critical research into the field.[20]

Money laundering is, if nothing else, a highly diverse practice, with one study by Irwin et. al. identifying as many as 300 typologies promulgated by AML and Counter-Terrorist Financing (CTF) bodies in the 1996-2009 period.[21] The study highlights a dazzling variety of techniques and approaches, including (to name but a few): "smurfing" and structuring, currency smuggling, gambling and casinos, fictitious sales and purchases, fake invoicing, underground banking, bank cheques and bank drafts, and – an approach of particular interest to this study – the use of shell companies.[22] Publications by the FATF reflect a similar diversity in approaches, with reports produced on non-profits,[23] diamonds,[24] hawala,[25] cryptocurrencies,[26] and football,[27] amongst many others. There is even scholarly work on money laundering in virtual worlds, such as World of Warcraft.[28]

[20] van Duyne, Harvey, and Gelemerova, *The Critical Handbook of Money Laundering*.
[21] Samantha Maitland Irwin, Raymond Choo, and Liu, "An Analysis of Money Laundering and Terrorism Financing Typologies," 94.
[22] Samantha Maitland Irwin, Raymond Choo, and Liu, 94–97.
[23] Financial Action Task Force (FATF), "Risk of Terrorist Abuse in Non-Profit Organisations."
[24] Financial Action Task Force (FATF), "Money Laundering and Terrorist Financing through Trade in Diamonds."
[25] Financial Action Task Force (FATF), "The Role of Hawala and Other Similar Service Providers in Money Laundering and Terrorist Financing."
[26] Financial Action Task Force (FATF), "Virtual Currencies."
[27] Financial Action Task Force (FATF), "Money Laundering through the Football Sector."
[28] Irwin et al., "Are the Financial Transactions Conducted inside Virtual Environments Truly Anonymous? An Experimental Research from an Australian Perspective."

For the purposes of refining the scope of this study, the clearest, most useful (and most colourful) disambiguation can be found in “What went wrong with money laundering law”, by Peter Alldridge.[29] Alldridge encourages readers to distinguish between the “elaborate, sophisticated, glamorous, and vague” understanding preferred by international organisations such as FATF, which emphasises the role of the international financial system; and money laundering’s “concrete, quotidian, and easily comprehensible” manifestation – exemplified the front businesses used by the characters in the HBO TV series, Breaking Bad.[30]

Despite a respect for more critical avenues of scholarship and a sincere appreciation of Alldridge’s somewhat mocking distinction, this study is firmly situated with a world of money laundering that is “elaborate, sophisticated, glamorous, and vague”. Despite considerable scope for investigation of the dynamics of small-scale laundering processes, this study is primarily concerned with high-volume, transnational money-laundering, which uses UK shell companies to obscure a variety of ill-gotten gains.

2.2 Shell companies

Shell companies are described by the World Bank as “a non-operational company—that is, a legal entity that has no independent operations, significant assets, ongoing business activities, or employees”.[31] As the authors of the report in which this definition is presented note, the term is not unambiguous, with key actors often holding multiple simultaneous definitions[32] or choosing to avoid defining the term altogether.[33]

---

[29] Alldridge, *What Went Wrong with Money Laundering Law?*
[30] Alldridge, 4.
[31] de Willebois et al., *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*, 34.
[32] The Organisation for Economic Co-operation and Development (OECD), “Glossary of Tax Terms”; Organisation for Economic Co-operation and Development (OECD), *Behind the Corporate Veil*, 17.
[33] Financial Action Task Force (FATF), “Glossary.”

There are legitimate and legal reasons for establishing a shell company. Legitimately, a shell company may serve as a basis for a neutral jurisdiction for arbitration in a joint venture, provide political stability and access to global financial centres for companies based in volatile markets, and establish structures which underlie securitisation processes and mutual funds.[34] Notoriously, but regrettably often legally, shell companies are also frequently used by individuals and companies to minimise their tax bill.[35]

Whilst the use of shell companies is widespread, and in many cases may be based in sound, legitimate business practices, their use is somewhat harder to justify when deployed primarily for the purpose of disguising their operator; and there is overwhelming evidence of the centrality of such entities in cases of financial crime, corruption and money laundering:

The use of shell companies to obscure the source of illicit funds has been highlighted in recent cases involving corruption in Azerbaijan, Malaysia, and Russia,[36] industrial-scale tax evasion,[37] and even during the Mueller inquiry into interference into the 2016 United States (US) election.[38] The importance to policy-makers is underscored by the substantial sums of money involved: totalling multiple billions of dollars in the largest cases.[39] These significant financial resources, whether obtained through corruption, procurement fraud or tax evasion, represent vast sums of money lost to the public purse – diverted from public services in rich and developing nations alike.

[34] Burns and McConvill, "An Unstoppable Force: The Offshore World in a Modern Global Economy."
[35] Reuters, "Google Shifted $23bn to Tax Haven Bermuda in 2017, Filing Shows"; Drucker and Bowers, "After a Tax Crackdown, Apple Found a New Shelter for Its Profits."
[36] Harding, Barr, and Nagapetyants, "Everything You Need to Know about the Azerbaijani Laundromat"; Ramesh, "1MDB: The Inside Story of the World's Biggest Financial Scandal"; Blum, Obermaier, and Obermayer, "Panama Papers - Putin's Rich Friends."
[37] United States of America vs. Wegelin & Co., Michael Berlinka, Urs Frei and Roger Keller.
[38] McIntire, "After Campaign Exit, Manafort Borrowed From Businesses With Trump Ties."
[39] Ramesh, "1MDB: The Inside Story of the World's Biggest Financial Scandal."

2.3 The role of the United Kingdom

It is perhaps a result of their tax-avoiding property that in the popular imagination, shell companies are frequently associated with off-shore tax havens. These perceptions have been reinforced by recent high-profile scandals,[40] but the reality of their operation is more complex. Regulatory frameworks in major financial centres, including the United States and United Kingdom, have enabled the rapid and cheap bulk creation of corporate vehicles – often with levels of anonymity exceeding those afforded by "traditional" secrecy jurisdictions.[41]

There are strong incentives for money laundering schemes to be established within the UK, companies may be established quickly and easily, and there is strong evidence to suggest systemic weaknesses in oversight,[42] furthermore, the UK provides access to global markets, and a veneer of respectability offered by operations out of a global financial capital.[43]

2.4 Companies as networks

Companies do not exist in isolation. Requirements vary between company type and jurisdiction, but all companies in all jurisdictions are required to be registered against at least one other entity – a director, secretary or member, for example. In the United Kingdom, for most categories of company, at least one of these entities must be a natural person – through which legal accountability for the movement of funds can be traced. This noted, particular structures – notably Limited Liability Partnerships (LLP)[44] – allow the creation of corporate

[40] Garside, "Paradise Papers Leak Reveals Secrets of the World Elite's Hidden Wealth"; Harding, "What Are the Panama Papers? A Guide to History's Biggest Data Leak."
[41] Sharman, "Shopping for Anonymous Shell Companies," 135; Kasperkevic, "Forget Panama: It's Easier to Hide Your Money in the US than Almost Anywhere."
[42] Findley, Nielson, and Sharman, *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*; Bullough, "Offshore in Central London."
[43] Sharman, "Shopping for Anonymous Shell Companies."
[44] Companies House, "Set up and Run a Limited Liability Partnership (LLP)."

vehicles which are registered entirely against other corporate vehicles; a property which has been reported to have been abused to disguise beneficial ownership in cases of money laundering.[45] Moreover, the ability of the LLP structure to contain no natural persons has been utilised to create circular ownership structures, whereby multiple companies interlock to create a closed loop of ownership.[46]

The properties described above may be utilised to identify and better understand shell companies used for money laundering. Using methodologies from the fields of graph theory and network science, it is possible for researchers to construct networks from the officer and ownership relationships between companies. The use of graph tools to investigate corporate relationships dates back to the 1970s,[47] and has been utilised by numerous scholars in the study of corporate politics,[48] knowledge transfer,[49] and more recently and ambitiously, power structures in the global financial system.[50]

The process by which criminals and regulators compete to find and close loop-holes through which money can be laundered has been described as a "game of cat and mouse"[51], in which network analysis is a growing tool. The scope for network approaches to money laundering detection has been examined since the 1990s,[52] with a growing body of work scholarly work employing such methods.[53] Today, network approaches to AML are mainstream within the corporate sector, with a wide range of products on the market which

---

[45] Cowdock, "Hiding in Plain Sight," November 2017.
[46] Global Witness, "The Companies We Keep."
[47] Sonquist and Koenig, "Interlocking Directorates in the Top US Corporations: A Graph Theory Approach"; Fennema and Schijf, "Analysing Interlocking Directorates: Theory and Methods."
[48] Neustadtl and Clawson, "Corporate Political Groupings: Does Ideology Unify Business Political Behavior?"
[49] O'Hagan and Green, "Corporate Knowledge Transfer via Interlocking Directorates: A Network Analysis Approach."
[50] Vitali, Glattfelder, and Battiston, "The Network of Global Corporate Control."
[51] Ryder, "The Financial Services Authority and Money Laundering."
52 Sparrow, "The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects."
53 Fronzetti Colladon and Remondi, "Using Social Network Analysis to Prevent Money Laundering"; Drezewski, Sepielak, and Filipkowski, "The Application of Social Network Analysis Algorithms in a System Supporting Money Laundering Detection."

support companies and financial institutions in the detection and analysis of suspicious activity.[54]

Non-profit and media organisations are also increasingly actors in AML, NGOs have traditionally played a role from a policy perspective, but alongside journalists are also increasingly engaged in conducting large scale data analysis as part of investigative research. Tools designed for financial institutions presuppose large, real-time, privileged data, such as a bank may hold on its customers. Non-profit organisations and journalists on the other hand, may instead utilise open data [55] or increasingly leaks[56] when investigating money laundering and corruption. In this context, there is scope for new techniques and approaches which combine open data with scholarly insight, which may support non-profit and media actors in their investigations.

[54] Devaux, "Reinforcing AML Analysis Systems with Graph Technologies"; Ramachandran, "How Link Analysis Can Help In Anti-Money Laundering Investigations"; Xu, "Fighting Money Laundering with Real-Time Deep Link Analytics"; Lee, "The Use of Link Analysis Is Vital for AML Investigators."
[55] Global Witness, "The Companies We Keep"; Cowdock, "Hiding in Plain Sight," November 2017; Transparency International UK and Thomson Reuters Solutions, "London Property."
[56] Obermayer and Obermaier, *The Panama Papers*.

3 Methodology

3.1 Case Selection

This study examines the role of UK companies in money laundering. Whilst recognising that the challenge is by its very nature highly international, and that this phenomenon could be examined with a focus on a wide range of jurisdictions – from Switzerland[57] to the Seychelles;[58] or from Denmark[59] to Delaware[60] – there are compelling reasons for the choice of the United Kingdom as a case study.

Firstly, as noted above there are numerous reasons, including regulatory weakness, access to markets and global reputation, why the UK is an attractive market for money laundering.

Secondly, the UK and its dependencies are currently the site of a significant regulatory shift with regard to company formation and registration. Alongside the implementation of major European directives with regard to money laundering,[61] the country has recently introduced the requirement for a register of significant control.[62] Whilst studies by NGOs have shown flaws in its implementation,[63] the availability of Person of Significant Control (PSC) data presents unique opportunities for analysis.

57 "Switzerland Must Urgently Do More to Tackle Corruption, the OECD Says."
58 Shaer, Hudson, and Williams, "Sun and Shadows."
59 Chopping and Rubenfeld, "SEC Joins List of Authorities Probing Money Laundering at Danske Bank."
60 Kasperkevic, "Forget Panama: It's Easier to Hide Your Money in the US than Almost Anywhere."
61 Directive (EU) 2015/849 of the European Parliament and of the Council on the prevention of the use of the financial system for the purposes of money laundering (4MLD).
62 Baroness Neville-Rolfe and Department for Business, Innovation & Skills, "'People with Significant Control' Register Comes into Force."
63 Global Witness, "The Companies We Keep."

Thirdly, linked to the above, but more broadly, the UK's commitment to open data provides a large – if also flawed, see 5.1.3 – dataset through which to conduct analysis.

Fourth and finally, the author is obliged to note that they themselves are a British citizen. Furthermore, in the interests of transparency, it should also be noted that a member of the author's family has also conducted research into money laundering through UK companies in their role at Transparency International.

Whilst personal factors have undoubtedly influenced case selection, the author feels that the prominence of the UK in global financial markets strengthens the study's *relevance*, ongoing reforms contribute to its *timeliness*, data availability to its *feasibility*, and that personal connection is merely a matter of *interest.*

3.2 Definitions

3.2.1 Levels of Analysis

This study operates at two levels of analysis: the individual company and the corporate network. At the individual level, the unit of analysis is the company, with variables derived from Companies House data (see 3.5.1). At the network level, the unit of analysis is the corporate network, with variables derived from aggregated company level data, network derived attributes (e.g. the proportion of natural persons within a network), and graph attributes (see 3.5.2).

3.2.2 Corporate Networks

For the purpose of the study, the corporate network is defined as the multidigraph network structure built by linking edges (vertices) and nodes. Edge relationships are defined as officer appointments and PSC relationships. Nodes consist of companies, company officers (e.g. directors, secretaries), or PSCs. Companies may be registered in the United Kingdom, or in other jurisdictions; officers and persons of significant control may be either natural persons or companies themselves. Edge relationships are directional – a person, for example may be the director of a company, but not vice-versa. The graph is also a multigraph, in that parallel edges may exist – a person, for example may be both director and PSC of a company. Networks are constructed using chpy, a custom Python tool developed for the purpose of this study (see 4.1). An example of a structure that might be formed from this process is provided in Figure 1below.

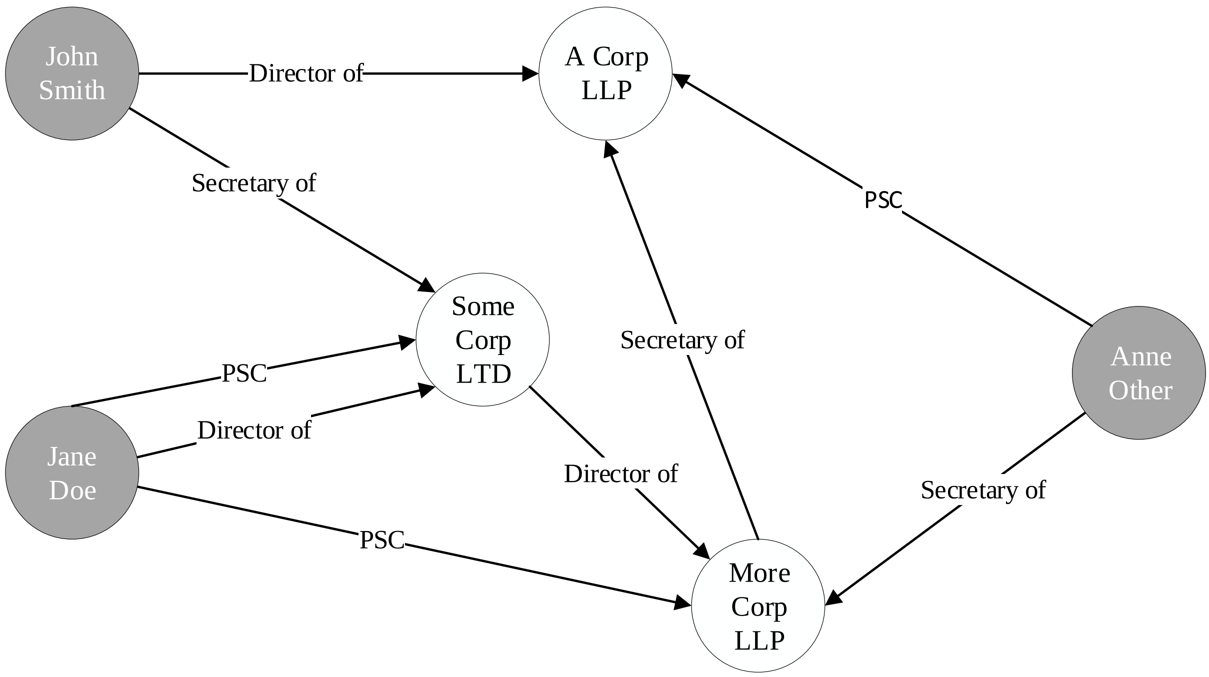

Figure 1: A toy corporate network

3.3 Research Questions

This study analyses money laundering risk through three hypotheses at two levels of observation. Initially, at *company level*, it undertakes an assessment of the risk score of individual companies drawn from networks within the sample; secondly, risk will be examined at *network level*, whereby networks constructed under the study will be examined using their network properties, aggregate properties of their constituent parts, and the properties of their root node. This two-tier analysis enables the examination of the extent to which relationships scale from the company to network level, which provides an important check on the robustness of findings.

3.3.1 H_1: Company age and type are strong predictors of money laundering risk at both company and network levels

This study considers that shell companies established for the purposes of large-scale money laundering are predominately temporary and artificial entities, established for the sole purpose of concealing illegal activity. Within this context, we can anticipate companies engaged in such activity to be comparatively short-lived. The ease with which companies may be created[64] means that companies may be established for the purpose of a limited number of transfers and then left dormant before being automatically closed.[65] They may be established within a constrained time period in which a feature of the regulatory system facilitates laundering.[66] Their operators may also choose to limit the duration of their activities to reduce administrative burden and risk of detection.

Numerous reports have also drawn attention to the scope for abuse of the UK limited liability partnership (LLP) model.[67] UK limited partnerships, and in particular Scottish Limited Partnerships (SLP) have featured prominently in a number of high-profile money laundering cases.[68] Furthermore, investigators have noted an explosion in the registration of such companies in the period following the year 2007 linking this growth to rising popularity

[64] Sharman, "Shopping for Anonymous Shell Companies"; Cowdock, "Hiding in Plain Sight," November 9, 2017.
[65] Findley, Nielson, and Sharman, *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*.
[66] Cartin and Higgins, "Offshore in the UK."
[67] Cowdock, "Hiding in Plain Sight," November 9, 2017; Global Witness, "The Companies We Keep."
[68] Cartin and Higgins, "Offshore in the UK"; "Ukrainian Mercenaries Are Using Scottish 'Tax Haven' Firm as Front"; "United Nations Blacklists Scottish Firms after International Aid Fraud"; Leask, "Scots Shell Firms Play Key Role in Latin America's Bribery 'Mega Scandal'"; Barr, "The Scottish Firms That Let Money Flow from Azerbaijan to the UK."

of UK corporate vehicles with suspicious financial flows from the Former Soviet Union (FSU).[69]

Given the above, this study anticipates that both age and company type will be significant factors in any analysis of risk at both company and network level. At the company level, this hypothesis will be tested through a Poisson (count) regression assessing individual company risk (see 3.4) as the dependent variable, in addition to company age and type as independent variables, alongside relevant control variables, including jurisdiction, company status and a compliance measure. At the network level, this analysis will be aggregated across sampled networks, with the arithmetic mean of the previously discussed variables being examined in a simple Ordinary Least Squares (OLS) regression. Finally, the analysis at both levels will be compared as a test of the robustness of the findings.

3.3.2 H_2: Money laundering risk within a corporate network is evenly distributed

This paper considers companies established for the purpose of money laundering as synthetic, with little or no engagement with the broader economy. Most high-risk companies are thus unlikely to possess substantial relationships with low-risk companies, and vice-versa. As such, corporate networks formed from a low-risk company are anticipated to be comprised of other low risk companies; similarly, networks formed from a high-risk company are anticipated to be comprised of other high-risk companies. This hypothesis will be examined at the network level through the relationship between the risk scores of n root company – the company from which a network is formed – and the average risk score of its constituent companies. That is to say, for a randomly selected company – the root company

[69] Cartin and Higgins, "Offshore in the UK," 6.

identified during sampling – the average risk score of its company network (excluding its own score) will be significantly related to its own.

3.3.3 H_3: Graph properties of a network (number of nodes, bipartite status, density and number of triangles) predict money laundering risk

When establishing a shell company, its operators (or more often their service providers)[70] will make conscious decisions as how best to avoid detection. Particular company types may be used – as discussed above, particular jurisdictions may be used or certain loopholes exploited. What may not be immediately obvious to someone seeking to establish companies as a vehicle for illicit funds is the geometry of their network. This study hypothesises that shapes may be formed from networks associated with money laundering that are distinct from other corporate networks. Network science provides a range of tools and measures for analysing networks, which have the potential to be instrumentalised in the identification and prevention of money laundering. Specifically, this study will examine measures of graph size (nodes), density, "bipartiteness", and the occurrence of triangles – circular ownership structures (see 3.5.2.1) – which may predict risk and provide new insights into laundering networks.

3.4 Primary Dependent Variable: Risk Scoring

3.4.1 Risk and Incidence

The study is concerned with how money laundering risk is distributed across such networks, and examines the extent to which graph structure differs between low and high-risk networks. The decision to examine *risk* (as opposed to *incidence*) of a given company or network being used for the purposes of money laundering results from practical constraints

[70] Sharman, "Shopping for Anonymous Shell Companies."

relating to data availability. Money laundering, as a criminal activity, is by its nature illicit, and moreover, as a practice it seeks to further *conceal* criminal activity. As such, identification of companies and networks engaged in the practice is challenging, with no centralised record of successful prosecutions available to the public. Whilst a number of high-profile cases have been reported on (see 2.2), these cases – whilst often linking to numerous companies – typically relate to a very limited number of networks. In this context, the number of network level observations available to researchers are highly constrained, given first a paucity of data points, and secondly a high likelihood of overlap within the available cases.

To address this challenge, this study utilised risk scoring undertaken by the non-profit organisation Global Witness in its 2018 report "The Companies We Keep".[71] The Global Witness study used open data from Companies House to highlight risk through a "red flag" system, whereby the presence of given risk factors was marked by an automated process across all companies in the Companies House database. A full rationale behind each flag is available within the Global Witness report. A summary description – alongside a count of all flags within the data provided by Global Witness – is provided below.

[71] Global Witness, "The Companies We Keep."

Table 1: Risk scoring flags from Global Witness

	Flag	Count
1	Company was registered at a company factory or mailbox address	208,572
2	Company's PSCs function as PSCs for a large number of other companies	9,199
3	Company is controlled via a trust	143,939
4	Officers or PSCs are based in secrecy jurisdictions e.g. British Virgin Islands	140,409
5	Company officers or PSCs are politicians	390
6	Company frequently changes its name	416
7	PSCs of company are disqualified directors	345
8	Company shares a PSC, officer or registered postcode with a company suspected of having been involved in money laundering	NA
	Total:	503,270

3.4.2 Limitations of the dataset

The final flag in the Global Witness analysis (flag 8) was regrettably unavailable for this study, as it was derived separately from other flags in a computationally intensive process, and not provided to this author. The absence of this flag from the analysis is unfortunate, given that it – more so than any other – connects a company directly to the practice of money laundering.

Also absent from the dataset are a number of high-profile confirmed cases, for example Metastar Invest LLP,[72] which – as a dissolved company – was not included in the Companies House bulk data product from which the Global Witness analysis was drawn.

As noted in the report, no one flag provides any indication of wrong-doing. The corporate vehicle behind Glastonbury Festival,[73] for example – an entity which the author sincerely hopes has not engaged in serious financial crime – has changed its name no less than ten times since 2005.

[72] Cartin and Higgins, "Offshore in the UK."
[73] Companies House, "Glastonbury Festival Events Limited."

3.4.3 Benefits, alternatives and use

Despite the above shortcomings, the data has properties that are beneficial in large-scale quantitative analysis. It is relatively open (available on request from Global Witness), objective and provides a straightforward approach to sampling. Alternative approaches, including a “manual” search for confirmed cases and freedom of information requests were considered. The manual search was discounted, in part for the reasons outlined above (see 3.4.1), but also given time constraints in the study. A freedom of information request was considered, however the presence of an acceptable alternative and the risk of receiving inadequate data precluded this option for this study.

This study uses the total number of flags for a given company within the Global Witness analysis as a proxy for risk of financial crime. Flags are analysed unweighted,[74] given the arbitrary nature of any such weighting scheme, however certain transformations and selections have been made for different levels of analysis. Where the company is the unit of observation, the number of flags possessed by a company is utilised (see 6.1.1); additionally, where company attributes constitute the key independent variables, and aggregation is undertaken to the corporate network level, the arithmetic mean of flags within a network is used (see 6.1.2). Where network attributes are of interest, a key independent variable is the number of flags possessed by a root company, in this case the arithmetic mean of a network is also taken as the dependent variable, with the root company’s risk score excluded from the calculation (see 6.2).

[74] Sampling weights have been applied, see 4.4.

3.5 Independent Variables

3.5.1 Company level

H_1 examines money laundering risk at the level of the individual company, utilising the large ($n = 131{,}174$) combined company dataset derived from network level sampling (see Figure 9 in 4.4). At this level, the study examines basic company attributes to test assumptions in the literature around the nature of money laundering in the United Kingdom. Specifically, the variables examined at this level are: company age, company type, company status, and the extent to which companies comply with requirements around filing (delinquency).

3.5.1.1 Company Age

As outlined in 3.3.1 there are numerous practical reasons why a company established for money laundering may only exist for a limited period of time. Additionally, money laundering in the UK is often framed as a relatively recent phenomenon, and indeed, the issue has only received international attention since the 1980s. Much work by non-governmental organisations on money laundering in the UK has a focus on the risk from countries of the former Soviet Union (FSU),[75] a concern mirrored in the narrative found in UK government investigations into the issue.[76] This focus has a temporal aspect, as by necessity, capital flight from the FSU is a phenomenon of the past 30 years. To better understand these aspects of risk, this study uses the age of companies as a key variable of interest and control within testing of all hypotheses.

[75] Cartin and Higgins, "Offshore in the UK"; Cowdock, "Faulty Towers"; Cowdock, "Kept in the Dark."
[76] Foreign Affairs Committee, "Moscow's Gold."

3.5.1.2 Company Type and Jurisdiction

Given the properties of LLPs described in 2.4, much attention has been given to this company form by NGO investigations. Scottish Limited Partnerships (SLP) have been singled out by Transparency International and others as of particular risk, given a now-addressed[77] exemption to PSC requirements. This study uses dummies for company type (LTD, LLP, private limited by guarantee, and other) and jurisdiction (UK, England/Wales, Scotland, Wales, and Northern Ireland) as variables of interest and control respectively across company and network level regressions.

The structure of the data provided on company jurisdiction in the Companies House dataset is not consistent –Wales for example, appears both alone and with England, and a separate category exists for the United Kingdom as a whole. These variables were considered as too important to omit at design phase, but given inconsistencies in the way jurisdiction is presented jurisdiction has been removed in the study's final analyses (see 6.1.1).

3.5.1.3 Delinquency and Company Status

Companies which are engaged in criminal activity may be careful in ensuring compliance with rules around timely filing, being cautious of detection. Equally, particularly where a vehicle is established for the purposes of a single transaction, they may also be negligent in filing. Companies House provides data on whether a given company's accounts (i), annual returns (ii) or confirmation statement (iii) are overdue; as well as whether a company has any charges (iv) made against it or an insolvency history (v). These five Boolean variables have been combined unweighted into a single "delinquency" variable, measured between 0 and 1 – none and all of the above, respectively.

[77] Department for Business, Energy & Industrial Strategy, "Limited Partnerships: Reform of Limited Partnership Law."

Company status (i.e. whether a company is active or not) is also taken as a control variable. Given discrepancies between the Company's House bulk data product and its API which prohibit easy mapping between inconsistently-provided company status types, this variable is simply coded 1 for active companies and 0 for all other possible statuses.

3.5.2 Network level

Analysis at the network level examines three major categories of variables: specifically graph metrics (arising from graph structure), network metrics (arising from attributes of the network), and control variables (aggregated from the company level).

3.5.2.1 Graph Metrics

3.5.2.1.1 Size (number of nodes)

Networks of companies engaged in money laundering are expected to be dependent on complexity to mask the identity of their controlling entities, and as such, the size of is an important factor in any analysis. The use of shell companies (see 2.2) is expected to inflate the overall size of a network, as additional non-productive entities through which funds are channelled are added. The use of shell companies is not unique to corporate structures involved in financial crime (see again, 2.2) This noted, given the complexity of such arrangements, and the associated administrative overheads, large, elaborate corporate structures are not anticipated to be utilised by the majority of UK companies, and as such network size is expected to be a positive predictor of risk.

The simplest measure of a network's size is the total number of its nodes (n) and edges (m) – indicating how many entities and relationships exist within a network. Within all samples under this study, strong collinearity is found between the number of nodes and edges in each network graph (see Figure 2 below), presenting challenges for the regression models used in Chapter 6.[78] To address this issue, a decision has been taken to examine the number of nodes (as opposed to edges) as a measure of overall graph size.

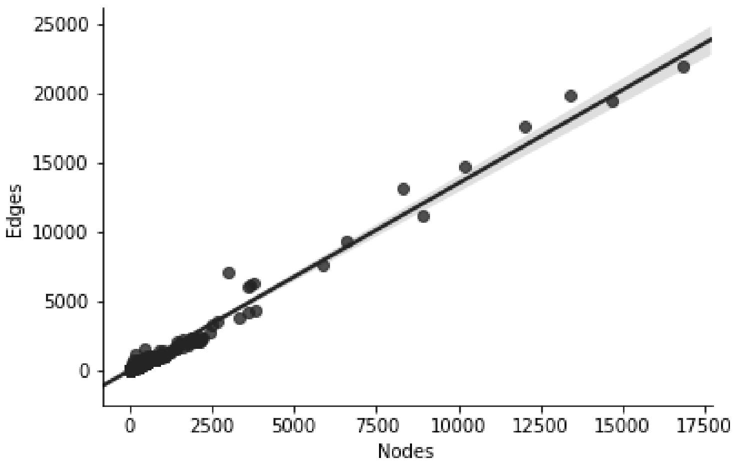

Figure 2: Collinearity between number of nodes and edges in network sample

3.5.2.1.2 Density

By omitting edges from the analysis, it becomes prudent to examine another measure of the extent of relationships within a graph. Density is a measure of the overall connectedness of a graph. A graph in which all nodes are connected to each other node is described as connected, similarly a graph in which no nodes are connected is described as disconnected, with the continuum between the two described as dense to sparse.[79] In the context of corporate interlock networks and money laundering, it is challenging to make a prediction regarding the relationship between density and risk. As described above, this study

[78] David A. Belsley, *Regression Diagnostics - Identifying Influential Data and Sources of Collinearity*, vol. 564, chap. 3.
[79] Barabási and Pósfai, *Network Science*, 61–63.

anticipates that high-risk networks may be larger, however the extent to which node connectivity may relate to risk is less intuitive.

Nonetheless, given that the metric expresses the relationship between nodes and a maximal number of edges, and given strong collinearity between edges and nodes – it will play an important part in this analysis as a control variable, ensuring variation within this relationship is modelled. An understanding of density within the context of high-risk networks may provide predictive insights, and open new avenues of research into the phenomenon.

The density of a directed graph is defined as follows, and provides a number scaled between 0 (disconnected) and 1 (connected).

$$d = \frac{m}{n(n-1)}$$

3.5.2.1.3 Bipartite Graphs

At the most fundamental level, graphs produced by chpy consist of two forms of node: individual nodes and corporate nodes, referring to natural persons and companies respectively. A bipartite graph may be divided uniquely into two sets of nodes U and V, whereby edges only exist *between* nodes of sets U and V respectively, and not within. In the illustration below, for example, the bipartite graph may be divided into two distinct groups so as described above, the unipartite graph, in contrast may not.

Figure 3: Bipartite and unipartite graphs

When applied to company networks, this property of a graph may be instructive in identifying more complex structures. A bipartite graph, whereby individuals are only linked to companies and vice-versa indicates a simple company structure, whereby companies are only linked to individuals. Where a graph is not bipartite, this indicates that within the graph at least one company is owned by another company, which may indicate that a network contains shell companies.

This study uses a 2-colour colouring algorithm built into the *NetworkX* Python module[80] to produce a binary value 0-1 to indicate the presence of a bipartite graph.

3.5.2.1.4 Triangles

A well-documented property of some money laundering networks is the presence of a circular ownership structure,[81] described below by the journalist, Oliver Bullough in his excellent account of transnational corruption, “Moneyland”:

[80] Hagberg, Swart, and S Chult, “Exploring Network Structure, Dynamics, and Function Using NetworkX.”
[81] Global Witness, “The Companies We Keep.”

"In February 2004, for example, Formations House created three companies: Corporate Nominees, Legal Nominees and Professional Nominees. The second company owned the other two, while itself being owned by the first company. The third company was secretary of the other two, while its own secretary was the first company. The second company was director of the other two, while its own director was the first company. It is hard to appreciate the curious symmetry of this arrangement unless you draw it out on paper, but it is marvellous, a real connoisseur's trick."[82]

To save the reader's pen and paper, a diagram presenting a representation of such a scheme is produced below in Figure 4.

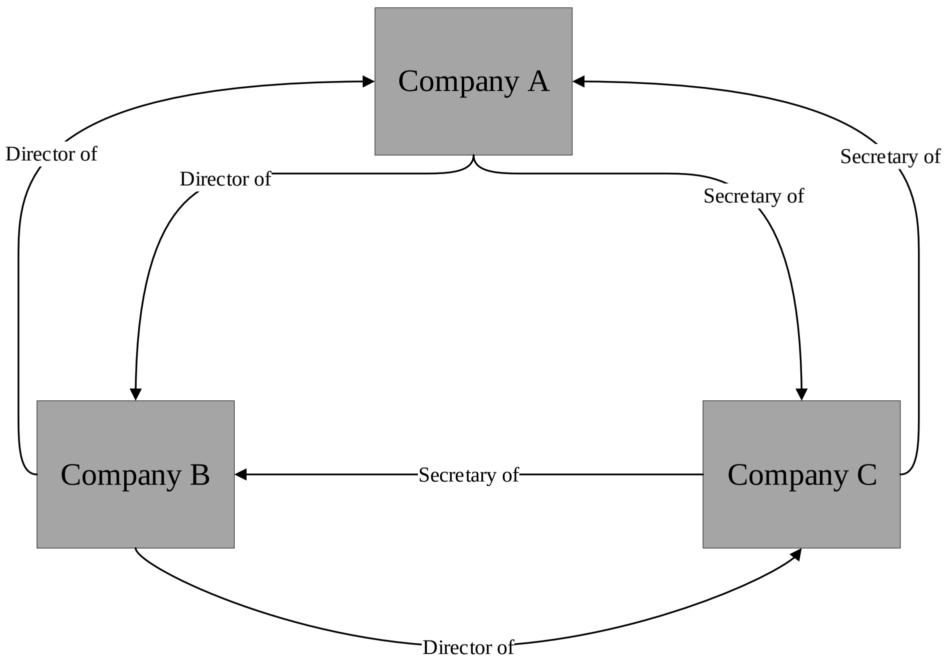

Figure 4: A circular ownership structure

[82] Bullough, *Moneyland*.

The identification of such structures is the sort of task that network analysis excels at. This study utilises a triangle detection algorithm within the NetworkX Python module[83] to count triangles within sampled networks. This process takes place within the corporate sub-graph (i.e. excluding natural persons), highlighting only companies with circular relationships. The presence of triangles is naturally anticipated to be highly correlated with money laundering risk.

3.5.2.2 Network metrics

3.5.2.2.1 Root company risk score

Principally utilised as a test for H_2 (3.3.2) this study utilises the risk score of a root company in relation to its constituent network. Network risk is aggregated excluding the score of the root company, see 3.4.3. As outlined in 3.3.2, root company risk is anticipated to have a strong relationship with overall risk within its network.

3.5.2.2.2 Proportion of Natural Person Nodes

Companies used for the purposes of money laundering seek to minimise links with natural persons, this study anticipates that networks comprised of fewer natural persons will have higher risk. There is, however, evidence that weak controls on company registration may allow for the registration of fictitious individuals as PSC.[84] As such, the presence of natural persons may not necessarily be as powerful indicator as one might expect. This variable is included in all network level regressions (6.2) as a control.

[83] Hagberg, Swart, and S Chult, "Exploring Network Structure, Dynamics, and Function Using NetworkX."
[84] Global Witness, "The Companies We Keep."

3.5.2.3 Aggregated Network Variables

Analysis at the network level examines company age, type, jurisdiction, delinquency and company status as described in 3.5.1 (aggregated by arithmetic mean). Aggregated network variables are used to assess the impact of company attributes on risk scoring at network level 3.5.2 and as controls (age and delinquency only) during graph and network metric regressions (6.2).

4 Data Collection

This study utilises data from Companies House, the UK company registrar. The choice of Companies House as an information source is predicated on the factors of openness, sufficiency and relevance. A key consideration during data source definition was the extent to which this study's methodology could be replicated, and its insights applied by other researchers and policy actors (e.g. NGOs). In this context the availability of a free, openly accessible Application Programming Interface (API) at Companies House, without restrictions on registration or monthly API calls made the data source a clear preference. Data from the Companies House API was deemed as sufficient for the purpose of the study (most importantly in network construction). Furthermore, its nature as a public entity provides scope for policy-level engagement. Commercially available datasets (e.g. Orbis)[85] were immediately discounted given its prohibitive cost and limited options for replication; the researcher did gain access to the OpenCorporates API,[86] however, given the presence of rate limiting, and a degree of path-dependency resulting from earlier work with the Companies House API, it has not been utilised within this particular study.

Companies House offers two categories of data products of relevance to the study, bulk download of all current company information, and its API. Bulk data presents considerable opportunities for analysis, allowing for examination of the entirety of the UK corporate register; having been used to great effect by international NGOs, including Global Witness.[87] Officer data, upon which network construction is dependent, is however not publicly advertised as available,[88] leading to a decision in the design phase of the study to

[85] Bureau van Dijk, "Orbis - Comparable Company Data."
[86] OpenCorporates, "The Open Database Of The Corporate World."
[87] Global Witness, "The Companies We Keep."
[88] Companies House, "About Our Services."

utilise the API, working with a sample of networks contained therein. At a later stage in the research process, however, the author discovered that bulk data may be made available upon request from Companies House – by which point time constraints prohibited the redesign and scale-up of the study.

The above notwithstanding, and whilst noting that results of a sample-based study will ultimately be less conclusive than a whole population study, the use of the API has two key advantages. Firstly, the scale of a full, bulk record-based study is non-trivial, and highly ambitious for the limited scope of this project, particularly in a context in which substantial data manipulation is required. Secondly, the use of the API has supported the development of an open-source tool (see below) which the author hopes may be used by researchers and NGOs in future studies.

4.1 Network Construction with chpy

As described above, the study examines networks built from the linkages between a root company and its associated officers and PSCs. Networks were constructed using the Companies House API, access to which was facilitated and automated through a Python module developed by the researcher for the purposes of this study. The module, entitled chpy, queries the Companies House API, starting from a root company, constructing a network structure from its corporate relationships – a process which can be iterated through an arbitrary number of times (see below). The tool is publicly available on GitHub[89] and PyPi,[90] and open-sourced under the permissive MIT licence. Networks constructed for the purposes of this study were built using version 0.1.1 of the tool during the period January 29 to February 12, 2019.

[89] Goodrich, *Chpy: Build Networks from the Companies House API*.
[90] Goodrich.

Beginning from a user-input company number (root company), the chpy tool makes API calls to Companies House requesting a company profile, company officers and persons of significant control for that company. Where officers are companies that are based in the UK, the company profile is called and added to a list for analysis in future iterations. Then a search is performed for all entities (officers, PSCs) to companies for which they are officers. The structure of the Companies House API regrettably limits this search to officer appointments, with no analogous process possible for PSCs – that is, that given an individual name, it is not possible to search for companies for which they are a PSC. This process may then be repeated, with each company identified above in turn taking the place of the root company. See below for a visual representation of the operation of the chpy tool.

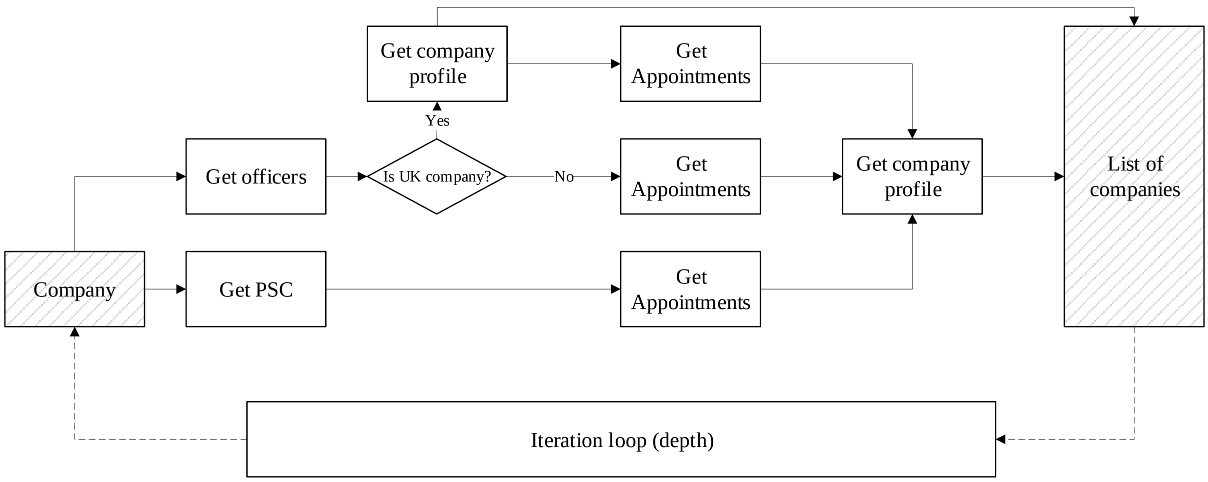

Figure 5: A visual representation of the operation of chpy

The chpy tool returns four categories of objects: specifically, a *NetworkX* multidigraph; an edge table, containing relationship (appointment and PSC) data, including source and target nodes; a list of companies sampled, and a gexf file for visualisation.

4.2 Sampling Strategy

Analysis at the level of the company network presents unique challenges for sampling: whilst a wealth of data is available on individual companies through platforms such as Companies House, networks are not presented as unitary entities in any public dataset and must be constructed. To address this problem, sampling has been undertaken at a company level, from which networks are constructed outwards utilising chpy as described in Figure 6 below.

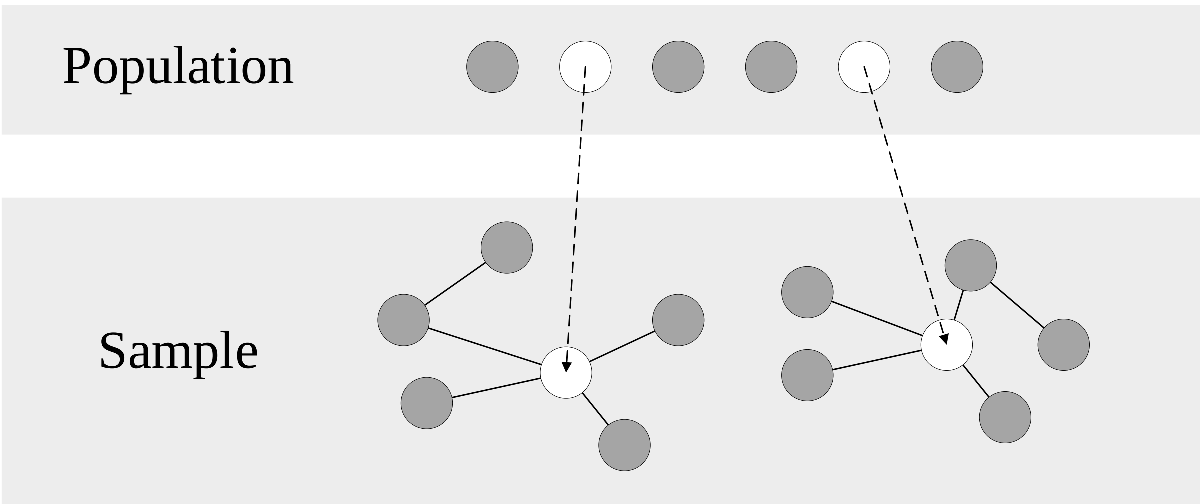

Figure 6: Network construction from a sample

The study is, however, fortunate in having a near complete population from which to sample. The sampling frame for the study was provided by Global Witness in the form of data resulting from the analysis undertaken for The Companies We Keep report (see 3.4). Whilst not current – the analysis from which the dataset was drawn utilised March 2018 data[91] – the sampling frame provides a near complete population of companies registered in the UK, alongside flags indicating risks of money laundering, which were used for the purposes of stratification.

CEU eTD Collection

[91] Leon, *Analysis and Code to Accompany The Companies We Keep Briefing on the State of the UK's Register of Persons of Significant Control.*

4.3 Depth and Sampling Constraints

The chpy tool used to construct networks for analysis may be run at multiple *depths* – iterations over the same data-acquisition loop for each item within its node list. At $depth = 1$, chpy draws data on its root company, its officers, PSCs and the companies to which they are appointed; at $depth = 2$, this process is repeated for each of the companies identified at $depth = 1$ (see Figure 5). Networks constructed at $depth = 2$ are considerably larger and more complex than those constructed at $depth = 1$, and thus take significantly longer to sample – frequently well in excess of one hour per network, equivalent to the time spent sampling nearly 2,000 companies at $depth = 1$.

Additionally, whilst every effort has been taken to ensure the robust operation of the chpy tool, minor bugs persist, as do problems with the Companies House API – which is still in Beta. The risk of encountering such errors grows with the number of API calls made – a substantially greater challenge at $depth = 2$. The management of errors during the sampling process was highly disruptive, with the process restarted more than once for debugging, limiting the time available for sampling. This phenomenon also resulted in a degree of sampling loss, whereby companies selected from the sampling frame were skipped during the implementation of sampling due to unresolvable errors, see Table 2 below for details. Additionally, at a smaller scale, and often as a consequence of data errors in the API, analysis of the generated network has identified missing nodes and edges in a small minority of cases – the absence of which may bias findings.

Table 2: Sample size and data loss

Sample	Companies Sampled	Total Networks Downloaded	Loss
Depth 1	2,000	1,960	40
Depth 2	100	83	17
Total	2,100	2,043	57

The above notwithstanding, from a network analysis perspective, depth 2 graphs are particularly valuable, as their size allows for the formation of more complex network structures thus allowing for greater differentiation between graphs. This phenomena is illustrated in Figure 7 and Figure 8 below, which shows visualisations and a limited number of summary statistics for graphs generated from the documented laundering company, Metastar Invest LLP[92] at depths 1 and 2 respectively.

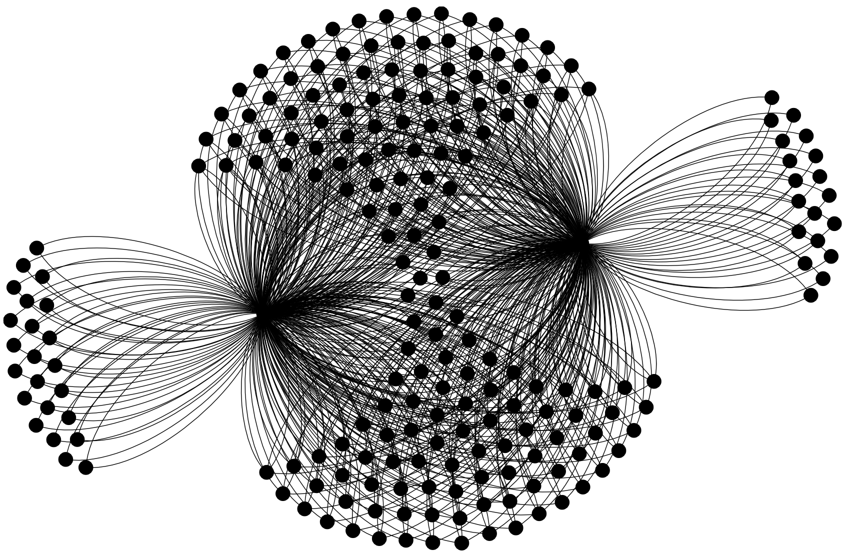

Figure 7: The Metastar Invest LLP network (depth = 1)

At $depth = 1$, the network is small and self-contained, and formed symmetrically around two major Belize-based hubs – Advanced Developments Limited and Corporate Solutions Limited. All nodes within this network are companies (not natural persons), and no PSC relationships are found.

[92] Cartin and Higgins, "Offshore in the UK."

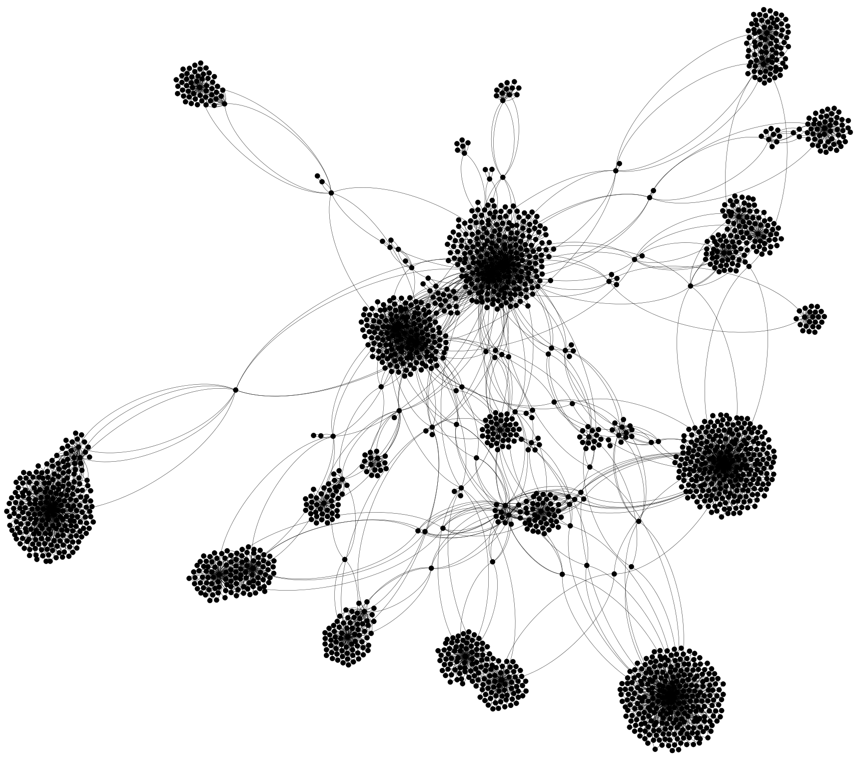

Figure 8: The Metastar Invest LLP network (depth = 2)

The deeper iteration at $depth\ =\ 2$ provides a dramatically different picture of the network. Many more hubs (high degree nodes) emerge as further entities are found to be connected to the low-degree companies identified at $depth\ =\ 1$ – there are in fact 18 nodes in this network with degree 50 or higher. The network also now contains both companies and natural persons, and is linked by appointment and PSC relationships.

Table 3: Example networks at depth = 1 and depth = 2

	$Depth\ =\ 1$	$Depth\ =\ 2$
Time to download	5min 20s	1h 23min 38s
Nodes	212	2,086
Edges	378	2,732
Percent natural persons	0%	12.6%
PSCs	0	68

4.4 Sampling Methodology

Sampling attempted to balance the need for a large sample which captures a broad range of risk, with the requirements of a multi-level analysis (company and network) and time constraints in the light of the factors outlined above. To this end, two data acquisition processes were undertaken, at depths 1 and 2 respectively.

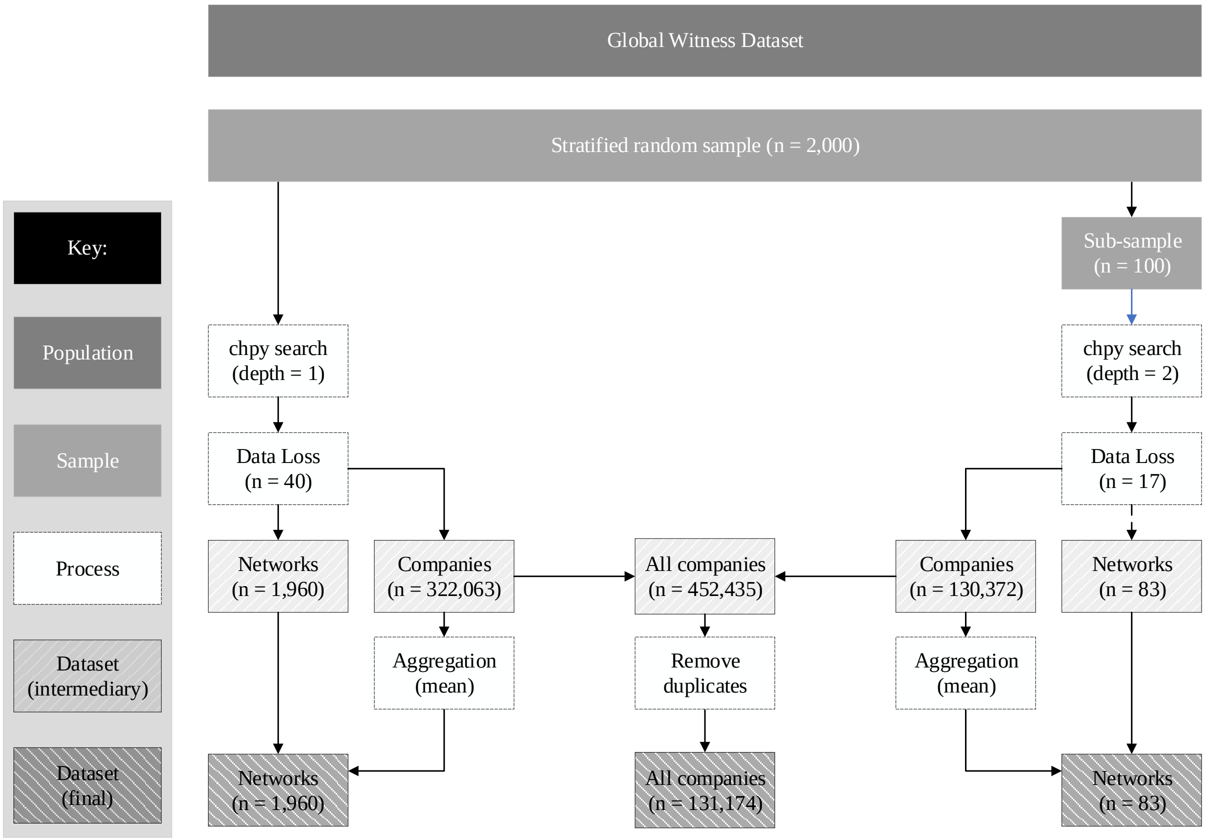

Figure 9: Sampling strategy

An initial sample ($n = 2{,}000$), stratified and oversampled on the basis of the number of flags was taken randomly from the Global Witness dataset. From this, a smaller sub-sample similarly stratified and oversampled ($n = 100$) was taken for the more time-consuming $depth = 2$ analysis. Oversampling was undertaken to ensure representation of adequate levels of risk across the sampling frame, with weighting undertaken against the Company's House bulk data product on the basis of risk (number of flags), company type and status (active or other). Numerous companies in the Companies House bulk data product were not included in the Global Witness dataset, and thus had no flags. In such cases, for the purposes of weighting, flags have been imputed as the mode value (i.e. 0 flags).

Data collection was conducted against the large sample and smaller sub-sample using chpy at $depth = 1$ and $depth = 2$ respectively. The results of each sample were used to generate two data tables (four in total) one comprised of company data and the second of network data (graph attributes). Both sets of company data were then merged with duplicates removed to form a single data set for analysis at the individual company level. These datasets were also aggregated and merged with their respective network data.

The final results of the sampling are three data sets: one containing company data for companies identified during both depth 1 and depth 2 sampling (used to test H_1, see 6.1.1) and two network datasets containing graph attributes and aggregated company attributes (used for H_{1-3} see 6.1.2, 6.2.1 and 6.2.2).

4.5 Summary Statistics

Presented below are top-line summary statistics for key variables in the results outlined in 6.

Table 4: Summary statistics -- all companies

	mean	std	min	max
Age	12.24	13.51	0.02	155.81
Number of risk flags	0.21	0.53	0.00	4.00
LLP	0.03	0.16	0.00	1.00
Active	0.63	0.48	0.00	1.00
Delinquency	0.07	0.14	0.00	1.00
Number of observations:				

Table 5: Summary statistics -- networks (depth = 1)

	mean	std	min	max
Age	5.99	6.21	0.58	49.58
Average number of flags	0.99	0.97	0.00	3.00
Average flags (excluding root company)	0.73	0.98	0.00	3.00
Root company flags	1.52	1.14	0.00	4.00
LLP	0.02	0.09	0.00	1.00
Active	0.71	0.32	0.00	1.00
Delinquency	0.03	0.06	0.00	1.00
Percent Human	0.22	0.25	0.00	0.95
Nodes	168.23	320.68	2.00	2451.00
Edges	190.82	359.72	1.00	2659.00
Bipartite	0.92	0.27	0.00	1.00
Density	0.24	0.36	0.00	1.50
Triangles	0.70	5.84	0.00	140.00
Number of observations:				

Table 6: Summary statistics -- networks (depth = 2)

	mean	std	min	max
Age	6.70	5.01	0.97	23.49
Average number of flags	0.79	0.87	0.00	3.00
Average flags (excluding root company)	0.48	0.79	0.00	2.31
Root company flags	1.39	1.21	0.00	4.00
LLP	0.01	0.02	0.00	0.11
Active	0.65	0.31	0.00	1.00
Delinquency	0.03	0.05	0.00	0.20
Percent Human	0.28	0.23	0.00	0.83
Nodes	1778.63	3521.28	2.00	16844.00
Edges	2505.12	4904.96	2.00	21897.00
Bipartite	0.65	0.48	0.00	1.00
Density	0.27	0.40	0.00	1.17
Triangles	57.67	186.53	0.00	1332.00
Number of observations:				

5 Limitations

This study is limited simultaneously by a surfeit and paucity of data. As discussed in 0, whilst the availability of data on company networks within the UK is in theory complete – i.e. that at the least officer data is available for all companies – there are several practical limitations to analysis at scale. In the context of this study, methodological choices have been made that have exacerbated some of these limitations; with consequences for sample size and thus the reliability of findings.

Conversely, in some cases, restrictions result from the absence of reliable data and are external to methodological choices. These restrictions may only be resolved by action on the part of regulatory authorities or through coordinated effort by external actors (e.g. OpenCorporates). This section will discuss the limitations encountered in the design and implementation of this study, specifically issues arising from the use of proxy risk scoring for analysis of money laundering networks, the challenges of API use for bulk data collection at scale; and the absence of reliable data on corporate relationships.

5.1 Limiting factors within this study

5.1.1 Risk Scoring

The limitations and advantages of the primary dependent variable – flags from the Global Witness dataset – are discussed in depth under 3.4.2 and 3.4.3, and will not be returned to in depth here. Briefly however, this study examines risk as opposed to incidence of money laundering, and can thus only provide tentative conclusions at best, as to the nature of the problem. The study presupposes the efficacy of the Global Witness methodology, and risks both false positives (e.g. Glastonbury Festival)[93] and false negatives – where identified,

[93] Companies House, "Glastonbury Festival Events Limited."

but dormant, laundering vehicles (e.g. Metastar Invest)[94] are excluded from the bulk product from which the Global Witness study was run. Any future work in this area should focus on either identifying a new sample from which to draw companies of interest, or in refining the Global Witness methodology to address its above-identified limitations.

5.1.2 API use for bulk data collection

As discussed under 4.1, the use of an API for complex bulk data collection places considerable limitations on the volume of data that can be accessed. As illustrated in Table 3, networks at $depth = 2$ are large, requiring substantial time (frequently in excess of 1.5 hours) to sample. The process at $depth = 2$ undertaken for this study, which sampled 100 networks (83 after sampling loss), took in excess of one week. In this context, whilst recognising strong potential as a tool for investigators, the chpy module as it stands is perhaps ill-suited to building company networks at scale – an issue which has severely limited the sample size available for this study.

Moreover, errors resulting from API issues and problems with the chpy code may also have led to graph components being omitted, undermining the reliability of data collected. This noted, there is no reason to suspect that such errors may be biased towards particular company attributes, and as such this limitation has largely been treated by this study as "noise".

Future work towards this area of research could include addressing inefficiencies and improving error-handling within the chpy module. This noted, hard limitations imposed by Companies House rate limiting and a necessarily large number of API calls for larger

---

[94] Cartin and Higgins, "Offshore in the UK."

networks will remain. Ultimately, an approach which better utilises bulk data products, including the elusive bulk officer data (see 4), would likely prove more efficient.

5.1.3 Data quality and network construction from search results

Linked to the above, limitations in the way company data is made available to the public present non-trivial data processing challenges – not least of which being the identification of unique entities. This problem was addressed by chpy (as described below), but is not without its own limitations, and may require substantial modification to scale to a bulk data product.

Companies House and other repositories that are reliant on their data (e.g. OpenCorporates) depend on user-submitted information, which is subject to serious challenges regarding data quality. This issue is challenging for data collection, but also leaves strong potential for abuse.[95]

Officers registered in Companies House do, in theory, possess unique identification codes, which can be used to query aspects of the Companies House API. In practice, however, entities across the database frequently possess multiple identification codes. In the case of larger entities, particularly professional nominee vehicles, the number of codes per entity can run into the hundreds, with single codes often existing per appointment. This phenomenon exists for both companies within the laundering data set, but also for those within the random sample, presenting challenges for the establishment of relationships (i.e. edges) within the constructed networks.

[95] Bullough, "Offshore in Central London"; Global Witness, "The Companies We Keep."

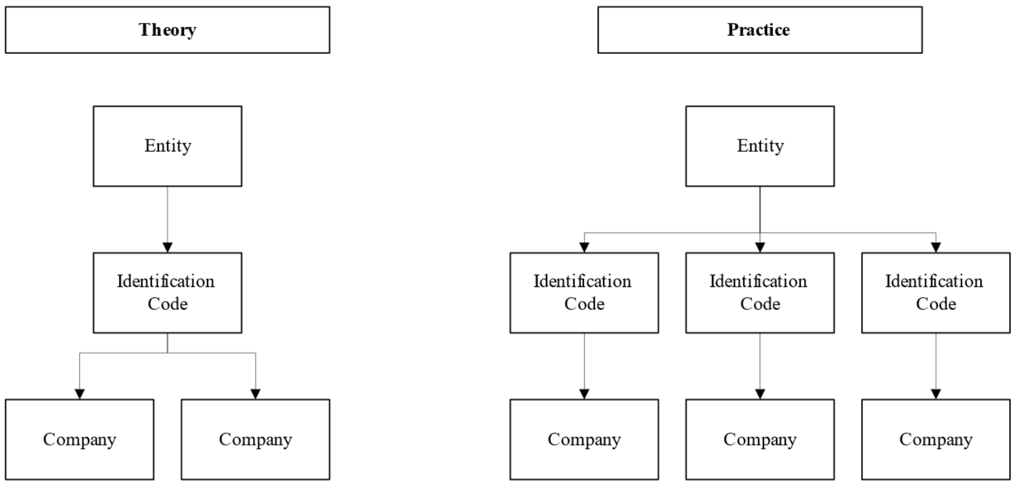

Figure 10: Challenges of officer relationship attribution in the Companies House API

For each officer record, a list of all appointments is provided, however given the above structure, and where an entity has multiple identification codes, these links do not return all the entity's relationships. In many cases, the only relationship returned will be that to the referring company.

In the absence of reliable relationship data, this study has built edges by matching entities from the root company's officer list with results from the API's officer search function. The approach entailed the use of fuzzy string matching through the FuzzyWuzzy module[96] which uses Levenshtein Distance [97] to calculate the differences between strings, specifically a given node and search results' name and address. Precautions were taken to mitigate against false positives, in that the chyp module requires *both* a strong (>90) fuzzy match of an entity and search result's name *and either* an exact date of birth match *or* a strong fuzzy match between addresses – the latter being provided in a highly inconsistent format.

[96] *FuzzyWuzzy: Fuzzy String Matching in Python.*
[97] Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals."

5.2 Overall impact of limitations

Whilst the challenges outlined above undoubtedly limit the strength of the claims that can be made by this study, there are, inevitably, methodological issues posed by the illicit nature of the subject under observation. Whilst the risk-scored approach to the dependent variable presents challenges, such issues would undoubtedly be replaced by others should a different approach be taken – any dataset of laundering companies will inevitably be subject to some form of sampling bias. A prosecution list – were one to be found – would only include those caught; and journalism and NGO work on the subject has a skew towards money from the FSU and developing countries (see 3.5.1.1), which could potentially lead to under-sampling other forms of the phenomenon.

Issues arising from API use have undoubtedly restricted sampling at higher depths, and this study has been careful to frame conclusions drawn from low n samples as tentative. As also noted above, there is most certainly scope for further investigation with larger samples, and it is hoped that this study presents avenues into further work in this regard.

Data quality is in part a programming issue, but to a much greater extent a policy issue. Recommendations on improvements to the Companies House database have featured prominently in key reports[98] with a roadmap in place outlining updates to the service.[99] Again, data quality issues are perhaps to some extent unavoidable in any study, this notwithstanding it is hoped that this study provides a glimpse of what is possible with what is possible with such a large, sophisticated and open dataset.

[98] Global Witness, "The Companies We Keep"; Cartin and Higgins, "Offshore in the UK."
[99] Companies House, "Our Strategy 2017 to 2020."

6 Results

6.1 Company Attributes

6.1.1 Company level

This analysis utilises company level data to test the impact of company age and type on money laundering risk. The dataset used is drawn from the companies identified in both $depth = 1$ and $depth = 2$ samples, merged and deduplicated as per Figure 9 ($n = 131{,}174$). This study considers the number of flags possessed by a company as count data, i.e. data that is positive, integer, ordered and arising from counting – as opposed to an arbitrary scale in which relative ranking is of primary importance.

Given the above, a Poisson regression is utilised with the number of risk flags possessed by a company as the dependent variable and age, company type, jurisdiction and delinquency as the independent variables (see 3.5.1). Variables are weighted on their probability of sampling to compensate for oversampling of high-risk companies. As this study seeks primarily to understand the impact of and relationships between variables – as opposed to making predictions on the basis of the data – variables have been min-max normalised to allow for straightforward comparison.

Table 7 below presents a staged addition of variables to the Poisson regression model, with Table 8 presenting the full outcomes of the final model (5). Omitted dummy variables against which other dummies are referenced are limited company (company type), inactive/other (company status) and United Kingdom (jurisdiction).

Table 7: Poisson regression with staged variable addition: number of risk flags and normalised and weighted company attributes at the company level

Independent Variable: Number of risk flags per company (num_flags)					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Const.	-2.0558***	-2.0559***	-1.9862***	-0.4271***	-0.4272***
	(0.0103)	(0.0107)	(0.0114)	(0.0157)	(0.0157)
Age	-21.9508***	-22.2048***	-13.3477***	-10.0191***	-9.3115***
	(0.7160)	(0.7243)	(0.8685)	(0.8738)	(0.8925)
LLP†		2.5301***	2.5805***	0.5644***	0.5777***
		(0.0906)	(0.0915)	(0.1056)	(0.1057)
Private†		-0.8873***	-0.9647***	-3.2098***	-3.2178***
		(0.0769)	(0.0771)	(0.0836)	(0.0837)
Other†		0.2555	0.9686***	-132.0637***	-132.0142***
		(0.3364)	(0.3720)	(1.5282)	(1.5276)
Active†			-1.5633***	131.0038***	130.9417***
			(0.1054)	(1.4618)	(1.4616)
England/Wales†				-140.2776***	-140.1238***
				(1.5122)	(1.5125)
N. Ireland†				-50.7655***	-50.6967***
				(0.6818)	(0.6820)
Wales†				-51.0378***	-50.9931***
				(1.5903)	(1.5866)
Scotland†				-141.7573***	-141.5846***
				(1.5485)	(1.5489)
Delinquency					-0.9473***
					(0.2821)
Model:	Poisson	Poisson	Poisson	Poisson	Poisson
Observations:	131,174	131,174	131,174	131,174	131,174
Pseudo R^2	0.015	0.023	0.026	0.155	0.155

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Given the large size of the dataset, it is perhaps unsurprising to see highly significant p-values for all variables. What is perhaps more surprising is the relative strengths (and at times directions) of the coefficients. Whilst age is negative, indicating that older companies can be anticipated to have lower risk scores, after normalisation, its coefficient is very small indeed in comparison to other variables – in many cases by multiple orders of magnitude. Similarly, the LLP dummy, which given the abundance of literature on the vehicle's use for money laundering should intuitively play a strong role in any empirical model, possesses an even smaller coefficient than that of age.

Table 8: Poisson regression, full output: number of risk flags and normalised and weighted company attributes at the company level

Independent Variable: Number of risk flags per company (num_flags)						
	coef	**stderr**	**z**	**P>|z|**	**[0.025**	**0.975]**
Const.	-0.4272***	0.016	-27.179	0.000	-0.458	-0.396
Age	-9.3115***	0.892	-10.434	0.000	-11.061	-7.562
LLP†	0.5777***	0.106	5.466	0.000	0.371	0.785
Private†	-3.2178***	0.084	-38.467	0.000	-3.382	-3.054
Other†	-132.014***	1.528	-86.418	0.000	-135.008	-129.02
Active†	130.9417***	1.462	89.588	0.000	128.077	133.806
England/Wales†	-140.124***	1.512	-92.646	0.000	-143.088	-137.159
N. Ireland†	-50.6967***	0.682	-74.334	0.000	-52.033	-49.36
Wales†	-50.9931***	1.587	-32.14	0.000	-54.103	-47.883
Scotland†	-141.585***	1.549	-91.41	0.000	-144.62	-138.549
Delinquency	-0.9473***	0.282	-3.358	0.001	-1.5	-0.394
Model:			Poisson			
Observations:			131,174			
Pseudo R^2			0.155			

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As noted in 3.5.1.2, the author has limited confidence in the jurisdiction variables, given that many overlap and are unclear. Examining Table 8 above, a dramatic increase in coefficients is observed during stage (4) as the jurisdiction dummies are added. These coefficients are surprisingly large, particularly in the context of a normalised regression, a phenomenon that may indicate a problem with the underlying variables.

The Poisson model is then run a final time, omitting these variables producing results which align more closely with expectations. In the final model, again, all variables are highly significant, with age reporting the highest normalised coefficient. Further, the LLP dummy, whilst still a degree of magnitude smaller than age predicts, as anticipated, increased risk of money laundering. Also of note is that the model predicts that inactive and paperwork-tardy companies may be of lower risk.

The above notwithstanding, however, the exclusion of the jurisdiction dummies has had a severe, detrimental impact on the model's pseudo-R^2 value, which has fallen to 0.0267, indicating that the model has very weak overall explanatory power, leaving limited confidence that the model provides sufficient insight to reject the null hypothesis.

Table 9: Poisson regression, full output: number of risk flags and normalised and weighted company attributes -- excluding jurisdiction dummies

Independent Variable: Number of risk flags per company (num_flags)						
	coef	**stderr**	**z**	**P>|z|**	**[0.025**	**0.975]**
Const.	-1.9818***	0.011	-174.125	0.000	-2.004	-1.959
Age	-11.9967***	0.884	-13.578	0.000	-13.728	-10.265
LLP†	2.601***	0.092	28.377	0.000	2.421	2.781
Private†	-0.9838***	0.077	-12.741	0.000	-1.135	-0.832
Other†	1.0151**	0.372	2.725	0.006	0.285	1.745
Active†	-1.45***	0.106	-13.687	0.000	-1.658	-1.242
Delinquency	-1.6536***	0.264	-6.274	0.000	-2.17	-1.137
Model:				Poisson		
Observations:				131,174		
Pseudo R^2				0.027		

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.1.2 Network level

An advantage of an analysis which examines both companies individually and as groups of related entities is that tests can be applied at both the individual and aggregated level. Intuitively, one may assume that findings at the individual company level may also be reflected at the level of the corporate network. To examine this dynamic, and also test the robustness of earlier findings, this study replicates the above regression (Table 9) using the arithmetic mean of each variable aggregated by root company as described in 3.5.

A number of modifications are made to the model for transposition to the network level. The dependent variable – the number of flags – is also averaged across each network's constituent parts, and in doing so becomes a continuous variable, for which a simple Ordinary Least Squares regression (OLS) is more appropriate. The model is run separately against both $depth = 1$ and $depth = 2$ datasets, avoiding distortion resulting from duplicate root companies and enabling an initial comparison of the robustness of findings across each depth. The number of observations is dramatically reduced ($n = 1960$ where $depth = 1$ and $n = 83$ where $depth = 2$), and normalisation and weighting are applied as before. As the arithmetic mean of values within a network, dummy variables now represent the proportion of companies within a network with a given characteristic (e.g. the percent of nodes that are an LLP). Furthermore, the troublesome jurisdiction dummies have also been removed.

Table 10: OLS regression network and company level comparison: number of risk flags and normalised and weighted company attributes -- excluding jurisdiction dummies

Independent Variable: Average risk flags per network (avg_flags)			
	$\boldsymbol{depth = 1}$	$\boldsymbol{depth = 2}$	**Company level**
Const.	0.3249***	0.4766***	-1.9818***
	(0.0119)	(0.0441)	(0.0114)
Age	-0.0691	-0.6317**	-11.9967***
	(0.1166)	(0.2765)	(0.8836)
LLP†	0.5491***	-0.2704	2.6010***
	(0.1527)	(0.1687)	(0.0917)
Private†	0.6830***	0.2998	-0.9838***
	(0.1858)	(0.2769)	(0.0772)
Other†	-0.4056***	-0.1175	1.0151***
	(0.0761)	(0.1313)	(0.3725)
Active†	0.2189***	0.0452	-1.4500***
	(0.0467)	(0.1460)	(0.1059)
Delinquency	-0.6248***	-0.1395	-1.6536***
	(0.1236)	(0.2176)	(0.2636)
Model:	OLS	OLS	Poisson
Observations:	1,960	83	131,174
R^2:	0.047	0.319	0.027‡
Adjusted R^2:	0.044	0.265	

† Dummy variable, ‡ Pseudo R^2, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comparing the results of regressions at depth = 1 and depth = 2, a sharp contrast is observed. In the larger sample (depth = 1), significant results are found for all variables *excluding* age, which has so-far been a highly significant, large coefficient beta across previous tests. The overall fit of the model is however poor, with an adjusted R^2 of only 0.044. The smaller test is a mirror image of its larger twin: its only significant variable *is* age, and the overall fit (adjusted R^2) is considerably stronger. Both weak significance for most variables and stronger overall fit could result from the low number of observations at $depth = 2$. The small sample size could limit the likelihood of significant results, and the comparatively strong R^2 may simply be an artefact of better "luck" from fewer residuals.

What is instructive is a comparison of the key variables of observation across all three regression models. When doing so, it can be noted that where significant, the main variables of interest (i.e. company age and LLP status) are in the same direction and often amongst the largest coefficients – particularly so in the case of age. The company type dummies as a whole, however, are highly unstable with large swings in the direction and significance of coefficients between models.

Regrettably, and as described above, each model possesses major flaws which prohibit the drawing of any firm conclusions. There is also considerable instability between results across different levels of aggregation and depth of search. This noted, challenges with aggregation may result from a) dramatic shifts in sample size, and b) poor data on jurisdiction – both of which being challenges that may be addressed in further studies. Furthermore, relative consistency of age, LLP status and delinquency as predictors does suggest there is room for additional exploration in of these variables.

We may thus tentatively conclude that the – admittedly limited – results shown suggest that companies showing risk of money laundering are typically younger than others, are more likely to use the LLP structure, and may be more likely to maintain their filings in a timely manner to avoid detection.

Age and delinquency will continue to be used as control variables in further regressions within this study. Given the instability of results across type dummies, the LLP variable will not be used in this manner.

6.2 Network Derived Attributes

At the network level, this study analyses risk through a simple OLS regression. It examines mean network risk as the dependent variable, alongside three categories of independent variables, specifically graph metrics, derived from each network's structure (number of nodes, a bipartite dummy, graph density, and triangles within the corporate sub-graph); network metrics, derived from attributes within the network (root company flags, the proportion of natural persons in the network) and control variables (age and delinquency). For regressions at the network level, the dependent is the arithmetic mean of risk scores within a network, excluding that of the root company – a measure to prevent autocorrelation.

Presented below are the results of these tests for networks structures sampled at $depth\ =\ 1$ and $depth\ =\ 2$, followed by analysis and implications for the study's hypotheses. As above, all variables have been weighted to adjust for choices made at the sampling stage, with the independent variables normalised to allow for easy comparison between regressors.

6.2.1 Depth 1

At $depth\ =\ 1$, a model based on graph metrics alone provides a very weak overall goodness of fit ($Adjusted\ R^2\ =\ 0.063$), with this measure improving dramatically following the addition of network variables ($Adjusted\ R^2\ =\ 0.341$). It is noteworthy that the addition of control variables has little impact on the overall goodness of fit. Network variables are highly significant throughout stages 2 and 3, with root company flags showing the largest coefficient.

Table 11: OLS regression, staged variable addition at depth = 1: average risk flags (excluding root), normalised and weighted network attributes and man company level controls

Independent Variable: Average risk flags per network, excluding root company (avg_flags_minus_root)			
	Graph Variables	Network Variable	Controls
Const.	0.2254***	0.1284***	0.1277***
	(0.0104)	(0.0094)	(0.0094)
Nodes	0.0388	-0.2364***	-0.2534***
	(0.0648)	(0.0563)	(0.0566)
Bipartite†	0.1378***	-0.0508	-0.0680
	(0.0392)	(0.0420)	(0.0451)
Density	-0.8392***	-0.2202***	-0.1531
	(0.0812)	(0.0802)	(0.0844)
Triangles	0.4259*	-0.0564	-0.1351
	(0.2188)	(0.1854)	(0.1887)
Root company flags		1.2454***	1.2336***
		(0.0451)	(0.0450)
Percent human		-0.1743**	-0.2303***
		(0.0679)	(0.0696)
Age			0.3200***
			(0.0797)
Delinquency			-0.3107***
			(0.0887)
Model:	OLS	OLS	OLS
Observations:	1,960	1,960	1,960
R^2:	0.065	0.343	0.349
Adjusted R^2:	0.063	0.341	0.346

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Whilst narrowly missing a significance star, and just straddling the positive-negative divide, network density shows some inconclusive promise as a predictor. As noted above, the variable which stands out strongest is the number of root company flags, which shows a very strong relationship with the dependent variable, adding weight to H_3. The number of natural persons also shows high levels of significance, although at the lowest co-efficient of all independent variables, this effect is not as strong as might have been anticipated – shell companies after all are supposed to limit the engagement of humans. Age and delinquency continue to be significant, although interestingly, with the addition of graph and network variables, age has performed an about-face, and now predicts positively – i.e. that older companies are higher risk.

Table 12: OLS regression, final output at depth = 1: average risk flags (excluding root), normalised and weighted network attributes and man company level controls

Independent Variable: Average risk flags per network, excluding root company (avg_flags_minus_root)						
	coef	**stderr**	**z**	**P>|z|**	**[0.025**	**0.975]**
Const.	0.1277***	0.009	13.593	0.000	0.109	0.146
Nodes	-0.2534***	0.057	-4.479	0.000	-0.364	-0.142
Bipartite	-0.068	0.045	-1.508	0.132	-0.156	0.020
Density	-0.1531	0.084	-1.815	0.070	-0.319	0.012
Triangles	-0.1351	0.189	-0.716	0.474	-0.505	0.235
Root company flags	1.2336***	0.045	27.407	0.000	1.145	1.322
Percent Human	-0.2303***	0.070	-3.308	0.001	-0.367	-0.094
Age	0.320***	0.080	4.015	0.000	0.164	0.476
Delinquency	-0.3107***	0.089	-3.503	0.000	-0.485	-0.137
Model:			OLS			
Observations:			1,960			
R^2:			0.349			
Adjusted R^2:			0.346			

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.2.2 Depth 2

At $depth = 1$, graph variables performed poorly. As described in 4.3, this study undertook sampling at $depth = 2$ to enable more complex (and thus differentiable) structures to emerge. The results of regressions at $depth = 2$ provide tentative support for this assumption, although given the very low number of observations at this depth ($n = 83$), a healthy degree of caution is required when drawing conclusions based on these results.

Nonetheless, the Adjusted R^2 is acceptable from the outset, demonstrating a relatively strong goodness of fit before network and control variables are included. Nodes remain significant and negative, and as such provide evidence to suggest that high-risk networks may actually be *smaller* than low risk networks, *ceteris paribus*. Encouraging for H_3 are the results of the triangle variable, which are both significant and in possession of the largest co-efficient within the (normalised) model. That this variable might indicate circular ownership structures is intuitive (see 3.5.2.1.4), and as such its strong performance was anticipated. The contrast between the results for this variable at $depth = 1$ and $depth = 2$ are instructive, suggesting that larger networks are required before such structures are made visible.

Table 13: OLS regression, staged variable addition at depth = 2: average risk flags (excluding root), normalised and weighted network attributes and man company level controls

Independent Variable: Average risk flags per network, excluding root company (avg_flags_minus_root)			
	(1)	**(2)**	**(3)**
Const.	0.2784***	0.2256***	0.2460***
	(0.0365)	(0.0430)	(0.0480)
Nodes	-0.4600***	-0.5129***	-0.4368**
	(0.1502)	(0.1383)	(0.1517)
Bipartite†	-0.1095	0.0672	0.0809
	(0.1243)	(0.1314)	(0.1376)
Density	-0.3443**	-0.1256	-0.1971
	(0.1536)	(0.1611)	(0.1702)
Triangles	0.4480**	0.4999**	0.5320**
	(0.2052)	(0.1913)	(0.1939)
Root company flags		0.3346***	0.2941**
		(0.1049)	(0.1103)
Percent human		-0.3633**	-0.2999
		(0.1573)	(0.1682)
Age			-0.0912
			(0.1548)
Delinquency			-0.1275
			(0.1798)
Model:	OLS	OLS	OLS
Observations:	83	83	83
R^2:	0.223	0.363	0.378
Adjusted R^2:	0.183	0.313	0.310

† Dummy variable, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The number of flags possessed by a root company continues to be a predictor of risk, lending additional weight to hypothesis H_2. Interestingly, both age and delinquency do not provide significant results in this model – an issue that might be resolved with more observations (see 5.1.2). Ultimately, whilst appearing to provide greater insight into larger networks, this regression is perhaps too small for conclusive results, it does however demonstrate scope for further analysis at greater depth.

Table 14: OLS regression, final output at depth = 2: average risk flags (excluding root), normalised and weighted network attributes and man company level controls

Independent Variable: Average risk flags per network, excluding root company (avg_flags_minus_root)						
	coef	**stderr**	**z**	**P>|z|**	**[0.025**	**0.975]**
Const.	0.246***	0.048	5.129	0.000	0.15	0.342
Nodes	-0.4368**	0.152	-2.879	0.005	-0.739	-0.134
Bipartite	0.0809	0.138	0.588	0.558	-0.193	0.355
Density	-0.1971	0.17	-1.158	0.251	-0.536	0.142
Triangles	0.532**	0.194	2.744	0.008	0.146	0.918
Root company flags	0.2941**	0.11	2.667	0.009	0.074	0.514
Percent Human	-0.2999	0.168	-1.783	0.079	-0.635	0.035
Age	-0.0912	0.155	-0.589	0.558	-0.400	0.217
Delinquency	-0.1275	0.18	-0.709	0.481	-0.486	0.231
Model:				OLS		
Observations:				83		
R^2:				0.378		
Adjusted R^2:				0.310		

† Dummy variable, * p < 0.05, ** p < 0.01, *** p < 0.001

6.3 Implications for Hypotheses

6.3.1 H_1: Company age and type are strong predictors of money laundering risk at both company and network levels

At the individual company level of analysis (3.5.1), large n regressions with data from over 130,000 companies provides strong support for H_1. These findings are corroborated by aggregated analysis at the network level (6.1.2), but only where graph and network property-derived variables are not included as controls (6.2.1 and 6.2.2). Age appears to have a strong influence, possessing consistently larger coefficients than other variables across all company-attribute tests (6.1). The LLP company type, whilst significant and a positive predictor of money laundering risk, is not as powerful an indicator as numerous reports suggest (see 3.5.1.2). Finally, the data available from Companies House on company jurisdiction is inconsistent and unclear, making it difficult to use within the analysis. As such, this study fully accepts the component of this hypothesis which addresses age, and more tentatively accepts the company type component, noting that further study into this dynamic may be required, and that exclusive focus on the LLP structure may be unwise.

6.3.2 H_2: Money laundering risk within a corporate network is evenly distributed

There is strong evidence from regressions at both $depth = 1$ and $depth = 2$ that a relationship exists between the risk of a root company and that of its network. Where a large sample is available (6.2.1), it is a very strong predictor, with a coefficient around four times the next most influential factor. This noted, there are other ways through which risk distribution in a network may be measured – an interesting area of study could allow for the inclusion, for example, of measures of centrality. Furthermore – as will be repeated below – the small sample at $depth = 2$ leaves any conclusions at the network level tentative. This study can thus provide evidence which supports this hypothesis, however given its narrow definition and the need for larger samples at greater depth, only preliminary conclusions can be drawn.

6.3.3 H_3: Graph properties of a network (number of nodes, bipartite status, density and number of triangles) predict money laundering risk.

Findings towards this hypothesis are greatly influenced by the depth at which sampling has been undertaken, and thus suffer considerably from the small sample size of the $depth = 2$ sample. As noted in 4.3, sampling depth has a dramatic impact on the size and shape of corporate networks constructed using the chpy tool, and it may only be at $depth = 2$ and beyond that patterns may manifest to an extent that allows empirical investigation. This notwithstanding, there is tentative support for the role of triangular structures in the identification of money laundering risk (6.2.2) although other network properties (notably bipartite status and density) do not appear significant given the limited data at hand. A counter-intuitive, but significant finding is that high-risk networks appear to have fewer nodes than those with lower risk. This could be the result of a number of factors. Professional nominee vehicles, for example, are heavily represented in the data, and often have relationships with hundreds and thousands of companies. Such companies play a key role in the corporate structures of small and large companies of all types, and may dramatically inflate network size where companies within a network use their services. Similarly, large, national and multinational corporations may also have highly complex corporate structures. The British high-street opticians, Specsavers, which was sampled during data collection, seems to register a new company for each of its branches, and appears to have 1,073 name-matching records in the company's house database.[100]

[100] Companies House, "Search Results: Specsavers."

Whilst given the limited evidence, it is not possible to accept the hypothesis, what this study finds encouraging is that there does appear to be evidence that a combination of graph properties and network-derived attributes may have predictive power for money laundering investigation. There is certainly scope for further research into a broader number of graph properties with a larger $depth = 2$ sample.

7 Conclusions

This study situated the efforts to combat money laundering within a domain that encompasses civil society organisations and journalists, alongside more traditional actors such as financial institutions and law enforcement agencies. It has identified the role of shell companies in large-scale laundering schemes and has drawn on evidence from the UK to better understand the problem. It has tested the use of network analysis as a tool for the study of money laundering, using openly available publicly available data.

It finds evidence that money laundering through British companies may indeed be a more recent issue and suggests that a more detailed examination of the temporal aspect of this phenomenon should be key to further studies. It also provides limited evidence that the LLP company structure has been abused to obscure ownership, but cautions that given this variable's limited effect size, money laundering risk is by no means exclusive to this particular structure.

The study has shown that there is a strong link between the risk of a root company and its constituent network. This is a step towards better understanding how money laundering companies interact with those in the "real economy", however the limited scope of this study has constrained further analysis on this part. There is certainly space for further examination of risk across networks, including assessing the role centrality plays in its distribution.

Finally, and most gratifyingly, the study has also tentatively demonstrated the scope for the use of network science tools in investigation of money laundering using open, company registry data. These tools allow the drawing of initial conclusions around network geometry and risk. Counter to initial expectations, high-risk networks may not be as large as low-risk networks, *ceteris paribus*; the presence of triangular structures may also be easily be identified and appears to bear a relationship to risk.

There is however much scope for further investigation in this regard. Seeking to avoid over-fitting and to test variables with a strong, intuitive link to money laundering risk, only a very limited number of graph properties were tested. Furthermore, the study was only able to analyse a relatively small sample of higher-depth networks, against a dependent variable which was not without limitations. Further research with larger samples and alternative dependent variables may yet provide deeper insights. With regard to exploration of additional graph properties, this problem may also be better suited to the domain of machine learning, where predictive power is prioritised over hypothesis testing.

Bibliography

Alldridge, Peter. *What Went Wrong with Money Laundering Law?* Palgrave Pivot. London, United Kingdom: Palgrave-Macmillan, 2016.

Aluko, Ayodeji, and Mahmood Bagheri. “The Impact of Money Laundering on Economic and Financial Stability and on Political Development in Developing Countries: The Case of Nigeria.” *Journal of Money Laundering Control* 15, no. 4 (October 5, 2012): 442–57. https://doi.org/10.1108/13685201211266024.

Barabási, Albert-László, and Márton Pósfai. *Network Science*. Cambridge, United Kingdom: Cambridge University Press, 2016.

Baroness Neville-Rolfe, and Department for Business, Innovation & Skills. “‘People with Significant Control’ Register Comes into Force.” GOV.UK, April 6, 2016. https://www.gov.uk/government/news/people-with-significant-control-register-comes-into-force.

Barr, Caelainn. “The Scottish Firms That Let Money Flow from Azerbaijan to the UK.” *The Guardian*, September 4, 2017. https://www.theguardian.com/world/2017/sep/04/the-scottish-firms-that-let-money-flow-from-azerbaijan-to-the-uk.

Blum, Petra, Frederik Obermaier, and Bastian Obermayer. “Panama Papers - Putin’s Rich Friends.” *Süddeutsche Zeitung*, 2016. https://panamapapers.sueddeutsche.de/articles/56fec05fa1bb8d3c3495adf8/.

Buchanan, Bonnie. “Money Laundering—a Global Obstacle.” *Research in International Business and Finance* 18, no. 1 (2004): 115–27.

Bullough, Oliver. *Moneyland: Why Thieves & Crooks Now Rule the World & How to Take It Back*. London, United Kingdom: Profile Books, 2018.

———. “Offshore in Central London: The Curious Case of 29 Harley Street.” *The Guardian*, April 19, 2016. https://www.theguardian.com/business/2016/apr/19/offshore-central-london-curious-case-29-harley-street.

Bureau van Dijk. “Orbis - Comparable Company Data,” 2019. https://www.bvdinfo.com/en-gb/our-products/data/international/orbis.

Burns, Michael J, and James McConvill. “An Unstoppable Force: The Offshore World in a Modern Global Economy.” *Hastings Business Law Journal* 7 (2) (205AD): 19.

Cartin, Gemma, and Ross Higgins. “Offshore In the UK: Analysing the Use of Scottish Limited Partnerships in Corruption and Money Laundering.” London: Transparency International UK and Bellingcat, June 2017. http://www.transparency.org.uk/publications/offshore-in-the-uk/.

Chaikin, David, and Jason Sharman. *Corruption and Money Laundering: A Symbiotic Relationship*. Palgrave Series on Asian Governance. New York: Palgrave-Macmillan, 2009.

Chapman, Ben. “Switzerland Must Urgently Do More to Tackle Corruption, the OECD Says.” *The Independent*, March 27, 2018. https://www.independent.co.uk/news/business/news/switzerland-bribery-corruption-money-laundering-oecd-report-must-urgently-do-more-whistleblowers-a8276686.html.

Chopping, Dominic, and Samuel Rubenfeld. “SEC Joins List of Authorities Probing Money Laundering at Danske Bank.” *Wall Street Journal*, February 21, 2019, sec. Markets. https://www.wsj.com/articles/sec-joins-list-of-authorities-probing-money-laundering-at-danske-bank-11550746375.

Companies House. “About Our Services.” Companies House, 2019. https://www.gov.uk/government/organisations/companies-house/about-our-services.
———. “Glastonbury Festival Events Limited.” Companies House, 2019. https://beta.companieshouse.gov.uk/company/04348175.
———. “Our Strategy 2017 to 2020.” London: Companies House, 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/609467/Companies_House_Strategy_2017-2020.pdf.
———. “Search Results: Specsavers.” Companies House, January 3, 2019. https://beta.companieshouse.gov.uk/search/companies?q=specsavers.
———. “Set up and Run a Limited Liability Partnership (LLP).” GOV.UK. Accessed February 26, 2019. https://www.gov.uk/guidance/set-up-and-run-a-limited-liability-partnership-llp.
Cooley, Alexander, John Heathershaw, and J.C. Sharman. “Laundering Cash, White Washing Reputations.” *Journal of Democracy* 29, no. 1 (2018): 39–53. https://doi.org/10.1353/jod.2018.0003.
Cowdock, Ben. “Faulty Towers: Understanding the Impact of Overseas Corruption on the London Property Market.” London, United Kingdom: Transparency International UK, March 2017. https://www.transparency.org.uk/publications/faulty-towers-understanding-the-impact-of-overseas-corruption-on-the-london-property-market/.
———. “Hiding in Plain Sight. How UK Companies Are Used to Launder the Proceeds of Corruption.” London: Transparency International UK, November 2017. https://www.transparency.org.uk/publications/hiding-in-plain-sight/.
———. “Hiding in Plain Sight: How UK Companies Are Used to Launder the Proceeds of Corruption.” London, United Kingdom: Transparency International UK, November 9, 2017. https://www.transparency.org.uk/publications/hiding-in-plain-sight/.
———. “Kept in the Dark: An Analysis of New Home Purchases at 375 Kensington High Street.” London, United Kingdom: Transparency International UK, August 2017. https://www.transparency.org.uk/publications/kept-in-the-dark/.
Crown Prosecution Service. “Proceeds Of Crime Act 2002 Part 7 - Money Laundering Offences.” Legal Guidance. London: The Crown Prosecution Service, March 2018. https://www.cps.gov.uk/legal-guidance/proceeds-crime-act-2002-part-7-money-laundering-offences.
David A. Belsley, Roy E. Welsch, Edwin Kuh. *Regression Diagnostics - Identifying Influential Data and Sources of Collinearity*. Reprint. Vol. 564. Wiley Series in Probability and Statistics. Wiley-Interscience, 2004. http://gen.lib.rus.ec/book/index.php?md5=D22E329AA60DA104BD0334F94BC7E279.
Department for Business, Energy & Industrial Strategy. “Limited Partnerships: Reform of Limited Partnership Law.” Consultation outcome. London, December 10, 2018. https://www.gov.uk/government/consultations/limited-partnerships-reform-of-limited-partnership-law.
Devaux, Elise. “Reinforcing AML Analysis Systems with Graph Technologies.” Linkurious, December 5, 2016. https://linkurio.us/blog/reinforcing-aml-graph-technologies/.
Drucker, Jesse, and Simon Bowers. “After a Tax Crackdown, Apple Found a New Shelter for Its Profits.” *The New York Times*, November 6, 2017, sec. World. https://www.nytimes.com/2017/11/06/world/apple-taxes-jersey.html.
Duyne, Petrus C. van, Jackie H. Harvey, and Liliya Y. Gelemerova. *The Critical Handbook of Money Laundering: Policy, Analysis and Myths*. London: Palgrave Macmillan UK, 2018. https://doi.org/10.1057/978-1-137-52398-3.

European Union. Directive (EU) 2015/849 of the European Parliament and of the Council on the prevention of the use of the financial system for the purposes of money laundering (4MLD), 2015/849 L 141/73 § (2015). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:JOL_2015_141_R_0003.

Fennema, Meindert, and Huibert Schijf. "Analysing Interlocking Directorates: Theory and Methods." *Social Networks* 1, no. 4 (1978): 297–332.

Financial Action Task Force (FATF). "Glossary," 2019. http://www.fatf-gafi.org/glossary/.

———. "Money Laundering," 2019. http://www.fatf-gafi.org/faq/moneylaundering/.

———. "Money Laundering and Terrorist Financing through Trade in Diamonds." Paris: FATF, October 2013. http://www.fatf-gafi.org/publications/methodsandtrends/documents/ml-tf-through-trade-in-diamonds.html.

———. "Money Laundering through the Football Sector." Paris: FATF, July 2009. http://www.fatf-gafi.org/publications/methodsandtrends/documents/moneylaunderingthroughthefootballsector.html.

———. "Risk of Terrorist Abuse in Non-Profit Organisations." Paris: FATF, 2014. http://www.fatf-gafi.org/publications/methodsandtrends/documents/risk-terrorist-abuse-non-profits.html.

———. "The Role of Hawala and Other Similar Service Providers in Money Laundering and Terrorist Financing." Paris: FATF, October 2013. http://www.fatf-gafi.org/publications/methodsandtrends/documents/role-hawalas-in-ml-tf.html.

———. "Virtual Currencies: Key Definitions and Potential AML/CFT Risks." Paris: FATF, June 2014. http://www.fatf-gafi.org/publications/methodsandtrends/documents/virtual-currency-definitions-aml-cft-risk.html.

Findley, Michael G, Daniel L Nielson, and Jason Campbell Sharman. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. Vol. 128. Cambridge University Press, 2014.

Foreign Affairs Committee. "Moscow's Gold." London: House of Commons, May 21, 2018. https://publications.parliament.uk/pa/cm201719/cmselect/cmfaff/932/932.pdf.

Garside, Juliette. "Paradise Papers Leak Reveals Secrets of the World Elite's Hidden Wealth." *The Guardian,* November 5, 2017. https://www.theguardian.com/news/2017/nov/05/paradise-papers-leak-reveals-secrets-of-world-elites-hiddenwealth.

Global Witness. "The Companies We Keep." London: Global Witness, July 2018. https://www.globalwitness.org/en/campaigns/corruption-and-money-laundering/anonymous-company-owners/companies-we-keep/.

Goodrich, Ian. *Chpy: Build Networks from the Companies House API* (version 0.1.1). Python, 2019. https://github.com/specialprocedures/chpy.

Hagberg, Aric, Pieter Swart, and Daniel S Chult. "Exploring Network Structure, Dynamics, and Function Using NetworkX." Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

Harding, Luke. "What Are the Panama Papers? A Guide to History's Biggest Data Leak." *The Guardian*, April 3, 2016. https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers.

Harding, Luke, Caelainn Barr, and Dina Nagapetyants. "Everything You Need to Know about the Azerbaijani Laundromat." *The Guardian*, September 4, 2017. https://www.theguardian.com/world/2017/sep/04/everything-you-need-to-know-about-the-azerbaijani-laundromat.

Hendriyetty, Nella, and Bhajan S. Grewal. “Macroeconomics of Money Laundering: Effects and Measurements.” *Journal of Financial Crime* 24, no. 1 (January 3, 2017): 65–81. https://doi.org/10.1108/JFC-01-2016-0004.

Irwin, Angela SM, Jill Slay, Kim-Kwang Raymond Choo, and Lin Liu. “Are the Financial Transactions Conducted inside Virtual Environments Truly Anonymous? An Experimental Research from an Australian Perspective.” *Journal of Money Laundering Control* 16, no. 1 (2012): 6–40.

Kasperkevic, Jana. “Forget Panama: It’s Easier to Hide Your Money in the US than Almost Anywhere.” *The Guardian*, April 6, 2016. https://www.theguardian.com/us-news/2016/apr/06/panama-papers-us-tax-havens-delaware.

Leask, David. “Scots Shell Firms Play Key Role in Latin America’s Bribery ‘Mega Scandal.’” *The Herald*. February 3, 2018. https://www.heraldscotland.com/news/15917473.scots-shell-firms-play-key-role-in-global-web-of-bribery/.

———. “Ukrainian Mercenaries Are Using Scottish ‘Tax Haven’ Firm as Front.” *The Herald*. February 16, 2018. https://www.heraldscotland.com/news/15998656.ukrainian-mercenaries-are-using-scottish-tax-haven-firm-as-front/.

Leask, David, and Richard Smith “United Nations Blacklists Scottish Firms after International Aid Fraud.” *The Herald*. July 28, 2018. https://www.heraldscotland.com/news/16383510.united-nations-blacklists-scottish-firms-after-international-aid-fraud/.

Lee, Scott. “The Use of Link Analysis Is Vital for AML Investigators,” February 25, 2016. https://datawalk.com/2016/02/25/the-use-of-link-analysis-is-vital-for-aml-investigators/.

Leon, Sam. *Analysis and Code to Accompany “The Companies We Keep.”* Python. 2018. Reprint, Global Witness, 2018. https://github.com/Global-Witness/the-companies-we-keep.

Levenshtein, Vladimir I. “Binary Codes Capable of Correcting Deletions, Insertions, and Reversals,” 10:707–10, 1966.

McIntire, Mike. “After Campaign Exit, Manafort Borrowed From Businesses With Trump Ties.” *The New York Times*, January 20, 2018, sec. U.S. https://www.nytimes.com/2017/04/12/us/politics/paul-manafort-donald-trump.html.

Neate, Rupert. “HSBC Escaped US Money-Laundering Charges after Osborne’s Intervention.” *The Guardian*, November 26, 2018. https://www.theguardian.com/business/2016/jul/11/hsbc-us-money-laundering-george-osborne-report.

Neustadtl, Alan, and Dan Clawson. “Corporate Political Groupings: Does Ideology Unify Business Political Behavior?” *American Sociological Review*, 1988, 172–190.

Obermayer, Bastian, and Frederik Obermaier. *The Panama Papers: Breaking the Story of How the Rich and Powerful Hide Their Money*. Revised edition. London: Oneworld, 2017.

O’Hagan, Sean B, and Milford B Green. “Corporate Knowledge Transfer via Interlocking Directorates: A Network Analysis Approach.” *Geoforum* 35, no. 1 (2004): 127–139.

OpenCorporates. “The Open Database Of The Corporate World,” 2019. https://opencorporates.com/.

Organisation for Economic Co-operation and Development (OECD), ed. *Behind the Corporate Veil: Using Corporate Entities for Illicit Purposes*. Paris: OECD, 2001.

———. “Glossary of Tax Terms,” 2019. http://www.oecd.org/ctp/glossaryoftaxterms.htm.

Ramachandran, Ranjith. “How Link Analysis Can Help In Anti-Money Laundering Investigations.” ACA Compliance Group, December 7, 2017. https://www.acacompliancegroup.com/blog/how-link-analysis-can-help-anti-money-laundering-investigations.

Ramesh, Randeep. “1MDB: The Inside Story of the World’s Biggest Financial Scandal.” *The Guardian*, July 28, 2016. https://www.theguardian.com/world/2016/jul/28/1mdb-inside-story-worlds-biggest-financial-scandal-malaysia.

Rankin, Jennifer. “European Parliament Calls for Investigation into ‘Azerbaijani Laundromat.’” *The Guardian*, September 13, 2017. https://www.theguardian.com/world/2017/sep/13/european-parliament-calls-for-investigation-into-azerbaijani-laundromat.

Reuters. “Google Shifted $23bn to Tax Haven Bermuda in 2017, Filing Shows.” *The Guardian*, January 3, 2019. https://www.theguardian.com/technology/2019/jan/03/google-tax-haven-bermuda-netherlands.

Ryder, Nicholas. “The Financial Services Authority and Money Laundering.” *The Cambridge Law Journal* 67, no. 03 (November 2008): 635. https://doi.org/10.1017/S0008197308000706.

Samantha Maitland Irwin, Angela, Kim-Kwang Raymond Choo, and Lin Liu. “An Analysis of Money Laundering and Terrorism Financing Typologies.” *Journal of Money Laundering Control* 15, no. 1 (December 30, 2011): 85–111. https://doi.org/10.1108/13685201211194745.

Schneider, Friedrich, and Ursula Windischbauer. “Money Laundering: Some Facts.” *European Journal of Law and Economics* 26, no. 3 (2008): 387–404.

SeatGeek. *FuzzyWuzzy: Fuzzy String Matching in Python.* (version 0.16.0). Python. 2011. Reprint, SeatGeek, 2019. https://github.com/seatgeek/fuzzywuzzy.

Sharman, J. C. “Shopping for Anonymous Shell Companies: An Audit Study of Anonymity and Crime in the International Financial System.” *Journal of Economic Perspectives* 24, no. 4 (November 2010): 127–40. https://doi.org/10.1257/jep.24.4.127.

Sonquist, John A, and Thomas Koenig. “Interlocking Directorates in the Top US Corporations: A Graph Theory Approach.” *Insurgent Sociologist* 5, no. 3 (1975): 196–229.

Shaer, Matthew, Michael Hudson, and Margot Williams. “Sun and Shadows: How an Island Paradise Became a Haven for Dirty Money.” *International Consortium of Investigative Journalists (ICIJ)* (blog). Accessed February 23, 2019. https://www.icij.org/investigations/offshore/sun-and-shadows-how-island-paradise-became-haven-dirty-money/.

Transparency International UK, and Thomson Reuters Solutions. “London Property: A Top Destination for Money Launderers.” London, United Kingdom: Transparency International UK, December 2016. https://www.transparency.org.uk/publications/london-property-tr-ti-uk/.

UNDOC. “Estimating Illicit Financial Flows Resulting from Drug Trafficking and Other Transnational Organized Crimes.” Vienna, October 2011. https://www.unodc.org/documents/data-and-analysis/Studies/Illicit_financial_flows_2011_web.pdf.

Unger, Brigitte, Melissa Siegel, Joras Ferwerda, Wouter de Kruijf, Madalina Busuioic, Kristen Wokke, and Greg Rawlings. “The Amounts and the Effects of Money Laundering.” *Report for the Ministry of Finance* 16 (2006).

United Nations Office on Drugs and Crime (UNDOC). “Money-Laundering Cycle,” 2019. https://www.unodc.org/unodc/en/money-laundering/laundrycycle.html.

United States of America vs. Wegelin & Co., Michael Berlinka, Urs Frei and Roger Keller, No. 81 12 Cr. 02 (JSR) (United States District Court Southern District of New York 2011).

Vitali, Stefania, James B. Glattfelder, and Stefano Battiston. "The Network of Global Corporate Control." *PLoS ONE* 6, no. 10 (2011): 1–6. https://doi.org/10.1371/journal.pone.0025995.

Willebois, Emile van der Does de, Emily M Halter, Robert A Harrison, Ji Won Park, and J C Sharman. *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*, 2011. https://doi.org/10.1596/978-0-8213-8894-5.

Xu, Yu. "Fighting Money Laundering with Real-Time Deep Link Analytics." *TigerGraph* (blog), February 9, 2018. https://www.tigergraph.com/2018/02/09/fighting-money-laundering-with-real-time-deep-link-analytics/.