Network Topologies of Intermediaries in the Offshore World

Oscar Julius Adserballe
Student ID: S160855

Copenhagen Business School

Supervisor: Rasmus Corlin Christensen

May 21, 2025

Contents

Abstract **1**

1 Introduction & Motivation **2**

1.1 Introduction . . . 2
1.2 Tax Avoidance at the Top of the Income Distribution . . . 3
1.3 Limitations of Traditional Demand-Side Models . . . 3
1.4 The Supply-side: Intermediaries as Gatekeepers . . . 4
1.5 Research Gap: Understanding the *Network Structure* to Inform Intermediary Regulation . . . 6
1.6 RQ: What role do offshore intermediaries play in networks of high-end tax avoidance? . . . 7
1.7 Roadmap of the Thesis. . . . 7

2 Theory **8**

2.1 A Note on Philosophy of Science and Methodological Approach . . . 8
2.2 Conceptual foundations . . . 8
2.2.1 Global Wealth Chains and the Role of Intermediaries . . . 8
2.2.2 Weaponised Interdependence . . . 10
2.2.3 Network Theory as a Lens for Understanding Illicit networks . . . 11
2.2.4 A Typology of Intermediaries and Their Role . . . 12
2.2.5 Secrecy Strategies: Financial Instruments and Legal Innovations . 13
2.3 Propositions . . . 14

3 Data and Methodology **15**

3.1 The ICIJ Offshore Leaks Database . . . 15
3.1.1 The Challenges of Obtaining and Interpreting Offshore Data . . . 15
3.1.2 Overview of the ICIJ Offshore Leaks Database . . . 16
3.1.3 Entities . . . 18
3.1.4 Intermediaries and Feature Engineering . . . 19
3.2 Other Data Sources . . . 20
3.3 Other Data Sources . . . 20
3.4 Agentic AI for Intermediary Classification . . . 22
3.5 Analytical and Statistical Techniques Applied . . . 23

3.5.1 Concepts from Network Analysis . . . 23
3.5.2 Entropy . . . 25
3.5.3 Association Analysis . . . 25
3.5.4 Multiple Hypothesis Testing . . . 26
3.5.5 Testing Significance of Results . . . 26
3.6 Use of LLMs in the Broader Paper . . . 27

4 Empirical Analysis **28**
4.1 Overview of the Dataset . . . 28
4.2 Geographical Specialisation . . . 30
4.3 Functional Specialisation of Intermediaries . . . 40

5 Discussion **46**
5.1 Three Propositions . . . 46
5.1.1 Proposition 1: Specialisation of Intermediaries . . . 46
5.1.2 Proposition 2: Duality of Intermediary Focus - Local Anchors, Global Reach . . . 47
5.1.3 Proposition 3: Structural Centrality of Microstates in Intermediation Network . . . 47
5.2 Implications for Regulation . . . 48
5.3 Limitations . . . 48
5.4 Future Research . . . 48

6 Conclusion **49**

A Appendix **50**
A.1 Country Network Centrality and Associations . . . 50
A.2 Jurisdiction Network Centrality and Associations . . . 52
A.3 Classification of Intermediaries . . . 54

List of Figures

3.1 Overview of Entity Incorporations Over Time from ICIJ Data 17

3.2 Agent Setup for Intermediary Classification 22

4.1 Geographical Concentration of Entities, Intermediaries, and Officers. Bar charts illustrate the percentage of total entities incorporated in the top 15 jurisdictions, entities with activity linked to the top 15 countries, intermediaries based in the top 15 countries, and officers associated with the top 15 countries. 29

4.2 Degree Distribution of Intermediaries and Model Fits. The plot shows the probability density function (PDF) of intermediary degrees on a log-log scale. The empirical data (blue line) is compared against a fitted power-law distribution (red dashed line, $\alpha \approx 2.08$) and a fitted log-normal distribution (purple dash-dot line). 31

4.3 31

4.4 Client and Incorporation Jurisdiction Heatmap for Intermediaries in Top 6-10 Countries (TWN, CHN, PAN, IDN, GTM). Panels follow the same format as Figure 4.3. 32

4.5 Each panel shows the distribution of client entity countries and incorporation jurisdictions for intermediaries based in the specified country. H_c denotes country entropy and H_j denotes jurisdiction entropy for that group of intermediaries. 32

4.6 Distribution of Normalized Entropy for Client Countries vs. Incorporation Jurisdictions, Aggregated at the Country Level of Intermediaries. The plot shows that intermediaries, when grouped by their country of operation, tend to use a more diverse set of incorporation jurisdictions (H_j, mean = 0.48) than the diversity observed in their clients' countries of activity (H_c, mean = 0.23). 33

4.7 Client and Incorporation Jurisdiction Heatmap for Cyprus-based Intermediaries ($H_c = 0.55, H_j = 0.53$). The heatmap shows that 12% of entities serviced by Cypriot intermediaries have links to Russia; a notable concentration. 33

4.8 Distribution of the Number of Countries Linked to Entities Served per Intermediary. The distribution is heavily skewed, with most intermediaries serving entities linked to only one or a few countries - even high-degree intermediaries. 34
4.9 Filtered Network of Co-Served Countries, Coloured by Louvain Community (left legend) and Regime Type (right legend). Edges shown have support ≥ 0.008 and lift ≥ 1.5. Node size can be proportional to degree or another centrality measure. 35
4.10 Distribution of the Number of Distinct Jurisdictions Used per Intermediary. Most intermediaries use only one or two jurisdictions for incorporation, but a tail of intermediaries uses a broader portfolio. 38
4.11 Filtered Network of Co-Used Jurisdictions, Coloured by Predominant Legal Technology (left legend) and Louvain Community (right legend). Edges shown have support ≥ 0.008 and lift ≥ 1.5. 39
4.12 Filtering Process of the Enriched Random Sample for Functional Classification. This flowchart illustrates the steps taken to arrive at the final set of intermediaries with high-confidence functional classifications used in this section's analysis. 41
4.13 Cumulative Distribution Function (CDF) of Degrees by Intermediary Classification (Random Sample). The plot shows distinct degree profiles, with Tax Experts and Investment Advisors generally having lower degrees than Administrators and Legal Experts. 42
4.14 Distribution of Intermediary Classifications: Random Sample vs. Top $\approx$1.5% by Degree. The random sample shows a more diverse mix of functional types, while the top-degree sample is heavily dominated by Administrators. 43
4.15 Average Entropy Measures and Bearer Instrument Usage by Intermediary Classification (Random Sample). Error bars could represent standard errors if available. 44

List of Tables

4.1 Summary Statistics for the Full Country Co-Service Network 34
4.2 Summary Statistics for the Full Jurisdiction Co-Usage Network 38

A.1 Top 10 Countries by Betweenness Centrality in the Full Co-Service Network (excluding XXX) 50
A.2 Top 10 Countries by Eigenvector Centrality in the Full Co-Service Network (excluding XXX) 50
A.3 Significant Country Associations in Co-Service Network (Bonferroni Corrected $p < 6.89 \times 10^{-6}$) 51
A.4 Top 10 Jurisdictions by Betweenness Centrality in the Full Co-Usage Network (excluding XXX) 52
A.5 Top 10 Jurisdictions by Eigenvector Centrality in the Full Co-Usage Network (excluding XXX) 52
A.6 Significant Jurisdiction Associations in Co-Usage Network (Bonferroni Corrected $p < 6.10 \times 10^{-5}$) 53

Abstract

Intermediaries form the crucial links enabling tax haven use, contributing significantly to global tax avoidance and inequality. While analyses often focus on demand-side factors, this thesis challenges such perspectives by asserting the critical importance of the supply-side network structure and intermediary agency for understanding and regulating offshore finance. Extending the network analysis of Chang et al. (2023) and drawing on Harrington's (2016) micro-sociological evidence, I analyze International Consortium of Investigative Journalists (ICIJ) leak data. Although ICIJ data has limitations for estimating the overall scale of avoidance, it permits robust generalization about intermediary roles within these complex networks. A novel agentic method is employed, enriching ICIJ data by incorporating publicly available online information about intermediaries' professional roles and affiliations.

This thesis presents four key propositions: 1) The overall network exhibits structural vulnerabilities concentrated around central intermediaries. 2) Intermediaries often display cultural or national specificity, catering to distinct clienteles. 3) Different intermediary types occupy distinct network positions and vary in systemic importance, measurable via network centrality. 4) Network structures are dynamic, adapting in response to regulation and financial innovation. This analysis provides critical insights into the architecture and potential regulatory chokepoints within tax haven networks, viewed through the lens of intermediary action.

Introduction & Motivation

1.1 Introduction

The central claim advanced throughout this thesis concerns the critical relevance of examining supply-side dynamics within the offshore financial system. Specifically, it argues that the role of intermediaries – the professional enablers and facilitators of offshore activity – is an incredibly relevant factor. The function and influence of the supply side – encompassing the specialized intermediaries and the specific services offered by various jurisdictions that actively enable and shape offshore activity – remains comparatively under-explored from an empirical standpoint. Building upon recent scholarship that increasingly highlights these supply dynamics (e.g., Laffitte 2024; Alstadsæter et al. 2019), this thesis seeks to extend and generalize insights from qualitative work, such as Harrington's (2016) study of wealth managers, through a quantitative analysis drawing upon the extensive data revealed by the ICIJ leaks.

Primary literature this is building on (contextualising interest in the topic):

- Interest spurred on this by an interest in optimal taxation regimes esp. Saez (2002), and the work of Zucman & Saez (2019) on the optimal taxation of wealth.
- Overall approach from neoclassical public finance and economics. Lectures from Zucman's overviews of tax evasion and avoidance in the modern economic literature has been the primary source. https://gabriel-zucman.eu/publicecon/
- Niche within Political Sociology through Brooke Harrington (2016)'s book and the method's employed in her ethnography of wealth managers. Likewise the tentative work in Chang et al. (2023a and 2023b) on network structure. However, for the latter, they concentrate more on demand-strategies rather than the more interesting supply-side strategies that are the focus of thisthesis.

1.2 Tax Avoidance at the Top of the Income Distribution

While considerable progress has arguably been made in curbing outright tax evasion, tax avoidance remains a substantial challenge, a point emphasized by commentators such as Stiglitz (cited in Alstadsæter et al., 2024). It introduces several clear inefficiencies into the economic system, including the generation of a distinct class and socially unoptimal rents accruing to the intermediaries who facilitate such schemes, the potential for poor allocation of resources as investment decisions are distorted by spurious incentives, and, beyond these economic inefficiencies, a range of normative concerns regarding fairness and the integrity of the tax system that inevitably accompany widespread tax avoidance.

A literature that has grown very prominent in the past two decades or so in A crucial distinction often highlighted is between income and wealth inequality. Income inequality can be somewhat ephemeral in nature; high-earners in one year may retire or experience income fluctuations in the next. Wealth, in contrast, tends to be more permanent, potentially distorting social outcomes over non-transient periods in a more meaningful way. Inordinate wealth accumulation (e.g. Harrington, 2016) has distorted social mobility (as explored in the work of Chetty) and been a key driver of overall inequality trends (e.g. Piketty's main body of work).

With that said, from a (narrow and purely economic) point of view, whether tax avoidance quantifying is actually bad is unclear, so the normative desirability of it at aggregate is still in question. The precise behavioral effects of tax evasion and avoidance on incentives – such as the incentives to work, save, or invest – is not as clear as, for example, studying the effects of tax incentives on MNCs (where it seems generally negative, e.g. Puerto Rico tax credit study from Serrato, 2018; also Garrett & Serrato, 2019). A key complicating factor is the role of expectations; an individual's behavior is likely highly dependent on their expectation of being able to successfully evade or avoid taxes in the future.

1.3 Limitations of Traditional Demand-Side Models

Traditionally, tax evasion and avoidance has been studied from the demand-side. The seminal Allingham-Sandmo (1972) good at explaining tax evasion decisions of the vast majority of the income distribution (Alstadsæter et al. 2019) performs poorly at the top of the distribution (ibid.) the Allingham-Sandmo (1972) model, provides a powerful and often empirically supported framework for understanding tax evasion decisions for the majority of taxpayers. This standard model typically portrays evasion as a individual and rational gamble, where individuals weigh the expected benefits of non-compliance against the probability of detection and the severity of potential penalties (see also Yitzaki & Slemrod). However, under standard assumptions about risk aversion and the structure of penalties and audit probabilities, the model often predicts that wealthier individuals, facing potentially higher stakes and scrutiny, should be less inclined to evade taxes. Yet, empirical evidence, particularly from studies leveraging leaked data (e.g., Alstadsæter et al. 2019), suggests the opposite: offshore tax evasion appears highly concentrated among the ultra-wealthy. The comparative statics do not hold here.

Furthermore, traditional methods for empirically studying tax compliance, such as random audit studies (e.g., Kleven et al. 2011), also face limitations in capturing the full picture of high-end evasion. As highlighted by Alstadsæter et al. (2019), while random audits are invaluable for understanding compliance behavior regarding income streams typically subject to third-party reporting or easily verifiable through standard audits, they often fail to detect the sophisticated, cross-border evasion strategies frequently utilized by the wealthiest segment. Complex offshore structures, shell corporations, and opaque trust arrangements often fall outside the scope of conventional audit procedures, rendering this form of evasion largely invisible to standard demand-side enforcement tools.

This points towards a dynamic of a game of cat and mouse. Demand-side enforcement mechanisms, predicated on detecting and penalizing individual non-compliance, struggle to keep pace with the evolving and increasingly complex strategies developed to obscure wealth and income, often with the assistance of specialized intermediaries. Consequently, relying solely on demand-side models and traditional enforcement metrics provides an incomplete, and potentially misleading, understanding of the phenomenon, especially concerning the significant evasion occurring at the top of the distribution. This underscores the necessity of incorporating supply-side factors and network structures to actually understand these mechanisms enabling tax avoidance at the top of the income distribution.

1.4 The Supply-side: Intermediaries as Gatekeepers

To fully grasp the dynamics of offshore tax evasion and avoidance, it is crucial to clarify what constitutes the "supply-side" (used more-so metaphorically than stringently) in this context. Here, the supply-side refers specifically to the ecosystem of professional intermediaries – such as law firms, banks, trust companies, and specialized advisors – as well as the jurisdictions that provide the legal and regulatory frameworks enabling offshore financial activities. The central argument advanced in this thesis, building on insights from models like Alstadsæter et al. (2019) and qualitative work such as Harrington (2016), is that this supply-side dimension is far more relevant to scrutinize than often acknowledged, potentially offering more effective avenues for understanding and potentially curbing offshore practices compared to a sole focus on demand-side factors.

A primary reason for emphasizing the supply side relates to the concept of elasticity. It is argued here that the elasticity of supply of intermediaries is considerably higher, and therefore potentially more responsive to policy interventions, compared to the elasticity of demand from clients seeking offshore services. Several factors underpin this view:

First, the incentives structuring the behavior of intermediaries are arguably much more sensitive to changes in the regulatory or reputational environment. For these professionals and firms, the provision of offshore services is not merely an option but often a core component of their business model and career trajectory. Their professional existence and profitability are directly dependent on their continued ability to offer these specific services effectively and discreetly. Consequently, factors that threaten this ability – such as increased regulatory scrutiny, heightened enforcement risk, or significant reputational damage – can have a pronounced impact on their willingness and capacity to supply these services. In contrast, the demand for tax minimization or evasion among potential clients, driven by factors like high tax rates or a desire for secrecy, can be seen as a relatively persistent force. While demand might fluctuate, the fundamental desire among some wealthy individuals and corporations to reduce tax burdens is likely to remain, making demand potentially less elastic to targeted interventions than the specialized supply of enabling services.

Second, the micro-sociological account provided by Harrington (2016) and Hoang (2022) offers compelling reasons why intermediaries are so central. Her ethnographic work illuminates the deeply personal, trust-based relationships that often form between wealth managers and their elite clients. These relationships, built over time and predicated on discretion and expertise, are difficult to replace. Clients rely heavily on their chosen intermediaries not just for technical execution but also for navigating the complexities and risks of the offshore world. The non-substitutable nature of these trust-based relationships means that disrupting the intermediary side can significantly impact clients' access to and ability to maintain offshore structures, further highlighting the critical role of the supply-side actors.

Third, the structure of the market itself points towards the strategic importance of intermediaries. There often exists a many-to-one relationship between clients and intermediaries; that is, a relatively small number of specialized intermediary firms or key professionals service a large number of clients seeking offshore solutions. This concentration means that the intermediary sector represents a point of leverage. Regulatory actions or enforcement efforts focused on these key intermediary players could potentially have a cascading effect, impacting a wide network of clients far more efficiently than attempting to identify and pursue each individual client separately. This structural feature makes the intermediary supply-side particularly vulnerable, and thus relevant, from a regulatory perspective.

1.5 Research Gap: Understanding the *Network Structure* to Inform Intermediary Regulation

Considerable research, particularly micro-sociological accounts like Harrington's (2016) ethnography, provides rich insights into the dyadic relationships, motivations, and practices of individual wealth managers and their clients. Ethnography, as a methodology, certainly offers a powerful means of accessing and understanding micro-level dynamics that can illuminate macro-level phenomena or "megatrends,"; of "entering in" an otherwise abstract metanarrative (cf. Neely, 2021; Also Chung 2018(check up; misremeber?)) However, generalizing from these detailed qualitative studies to broader systemic patterns has not really been done.

A nascent thread of literature has begun to explore these structural aspects, often spurred by the availability of large-scale leaked data. Work such as Chang et al. (2023), alongside policy-oriented research emerging from bodies like the EU following disclosures such as the Panama Papers (e.g., research from 2017), represents initial steps in this direction. However, this line of inquiry remains limited thus far, often focusing on specific subsets of countries or actors. The analysis of the network structures inherent in the offshore world is still in a highly exploratory phase. Consequently, the potential held within detailed micro-data sources, such as the ICIJ leaks which map connections between entities, officers, and intermediaries on a vast scale, remains largely underexplored in terms of systematic structural analysis.

The work by Chang et al. (2023) on "Secrecy Strategies" provides a pertinent example. While their primary focus was on analyzing the demand strategies employed by global elites, their findings crucially demonstrate that these strategies are shaped by, and interact with, the supply landscape – the available intermediaries, jurisdictions, and the institutional context of the elites' home countries. Their research, therefore, implicitly highlights the importance of the supply structure by showing how it influences demand patterns, effectively linking the two sides of the market through observable strategic choices.

This points towards the specific research gap addressed herein: the need for a more systematic understanding of the network structure of the supply-side itself. While we have compelling accounts of individual intermediary roles and incentives, a comprehensive picture of how these intermediaries connect to each other, to different types of clients, across various jurisdictions, and through specific service offerings – essentially, the topology of the intermediary network – is lacking. Understanding this structure is potentially crucial for designing more effective regulation targeting these key players.

Therefore, the goal within this thesis is to contribute to bridging this gap, primarily through synthesis and systematization. Drawing upon the existing literature, including the rich ethnographic accounts, the aim here is not necessarily to conduct a novel quantitative network analysis but rather to attempt to codify more generally and quantitatively on some of the more loosely defined observations about intermediaries and their roles. By viewing these observations through the conceptual lens of network structures and positions, the objective is to formulate more general propositions regarding intermediary behavior, influence, and potential vulnerabilities within the broader offshore system.

1.6 RQ: What role do offshore intermediaries play in networks of high-end tax avoidance?

1.7 Roadmap of the Thesis.

Having gone through what motivates the pursuit of this question and situate this thesis, will proceed to the bulk of the paper. First, outline the key concepts and theories I will draw on, then moving on to outline the key propositions this paper will seek to set forth about the role of intermediaries. Then, a brief section will cover the data sources.

Theory

2.1 A Note on Philosophy of Science and Methodological Approach

In line with perspectives advocating for methodological pluralism and the use of qualitative insights for broader theory development (e.g., George & Bennett, 2005), this thesis leverages Harrington's findings for concept formation and hypothesis generation - effectively using her work as a theory-building step. Her work helps define the "intermediary" phenomenon and suggests the importance of factors like trust and expertise, which likely underpin the network structures we observe. This thesis then seeks to assess the generalizability and structural manifestations of these insights across a large dataset, moving from micro-level understanding to macro/meso-level patterns. The objective, therefore, explicitly shifts from *verstehen* to identifying and analyzing recurrent structural patterns within the network as revealed by the ICIJ Offshore Leaks Database *and assuming we can generalise these structures* (more on that later under the methods section).

2.2 Conceptual foundations

This section outlines the necessary conceptual foundations that precede the concrete propositions asserted later in the thesis. These concepts presented here as being analytically requisite for the propositions developed in the subsequent section (2.2).

2.2.1 Global Wealth Chains and the Role of Intermediaries

To understand the significance of the intermediaries central to this thesis – the professional advisors, lawyers, accountants, and wealth managers operating within the offshore financial system – it is helpful to adopt an analytical framework that explicitly centers their role. The overall motivation for focusing on these actors stems from the "Global Wealth Chains" (GWC) approach.

As articulated by Seabrooke & Wigan (2014), this approach offers a distinct perspective compared to analyses focused on global value chains. They argue that: "While actors in value chains share an interest in transparency and coordination, those in wealth chains thrive on rendering movements through the chain opaque. Wealth chains hide, obscure and relocate wealth to the extent that they break loose from the location of value creation and heighten inequality." Adopting this GWC lens necessitates an explicit focus on the intermediaries and professionals. These are the actors who develop and deploy the sophisticated financial and legal innovations required to sustain and manage the complex structures used to hold individual wealth offshore, often obscuring its origins and ownership.

Further elaborating on the socio-legal dynamics underpinning these chains, Seabrooke & Wigan (2022) emphasize the significance of socially constructed legal meaning. They write: "What is significant here is accepted legal assertions,. This happens within interpretative communities, where agreements on legal affordances are secured." The intermediaries operate within these communities, shaping and interpreting the boundaries of legal possibility. Seabrooke & Wigan (2022) also connect this to broader social valuations, noting that "An important element is that within such communities wealth confers honor, where the accrual and transfer of wealth without productive effort is held in high esteem (Veblen, 1899)."

Borrowing from the typology proposed in Seabrooke & Wigan (2022), the networks involving the intermediaries examined in this thesis align closely with their definition of "relational wealth chains." These are characterized as follows: "Relational wealth chains involve the exchange of complex tacit information, requiring high levels of explicit coordination. Strong trust relationships managed by prestige and status interactions make switching costs high." This description of relational wealth chains, emphasizing tacit knowledge, trust, coordination, and high switching costs due to the personal nature of the relationships, is highly with the ethnographic work of Harrington (2016) and how she outlines the structure and dynamics of the networks between wealth managers and their elite clients. This connection is also drawn by Seabrooke & Wigan (2022) themselves, who cite Harrington (2015) alongside related work by Beaverstock & Hall (2016) and de Carvalho & Seabrooke (2016) as evidence supporting the characteristics of relational wealth chains.

Furthermore, a developing body of literature situated within this GWC approach is examining how these professionals actively shape and navigate existing regulatory landscapes (e.g., Christen, 2021; Christensen & Seabrooke). This underscores the analytical purchase of the GWC framework for understanding the pivotal role of intermediaries not just as passive facilitators, but as active agents within the offshore system.

2.2.2 Weaponised Interdependence

The goal here is to outline the theoretical basis for viewing intermediaries not just as facilitators, but as potential points of leverage or vulnerability within the offshore system, thereby informing regulatory strategies.

A lens for such an analysis is provided by the concept of "weaponised interdependence," as developed by Farrell & Newman (2019). Their core argument posits that globalization, far from simply flattening the world or diminishing state power, has often created highly specific network topographies. These global networks—whether in finance, technology, or supply chains—are frequently characterized by asymmetric structures. Power, in this view, does not dissipate but rather concentrates at key hubs or 'chokepoints' within these networks. States or actors who control these chokepoints gain significant leverage over others who depend on access to the network, potentially allowing them to 'weaponize' this interdependence for strategic gain.

This logic of weaponised interdependence has been applied directly to the domain of global tax policy by Christensen (2024). He argues that states have often failed to fully harness the potential regulatory power they could wield by strategically targeting chokepoints within the networks facilitating tax avoidance and evasion. Among the key institutions Christensen (2024) identifies as potential chokepoints relevant to global tax policy are precisely the expert intermediaries – the lawyers, accountants, wealth managers, and corporate service providers – who are central to this thesis. Their specialized knowledge and gatekeeping function position them as critical nodes whose disruption could have widespread effects.

This perspective aligns with and provides a theoretical underpinning for findings across various studies highlighting the importance and potential vulnerability of the intermediary supply-side. Research emphasizing the role of intermediaries (e.g., Harrington 2016; Alstadsæter et al. 2019) implicitly points to their structural significance. For instance, Harrington's (2016) focus on trust-based relationships suggests that disrupting these specific intermediary nodes can create significant friction. Alstadsæter et al.'s (2019) supply-side explanation for high-end evasion similarly underscores the crucial role of these facilitators. More explicitly, recent work analyzing the network structures revealed by leaks, such as Chang et al. (2023), demonstrates the analytical purchase of focusing on these networks. While their specific study examined network structures to understand the effectiveness of sanction regimes against oligarchs, the underlying approach – analyzing network vulnerabilities by focusing on intermediary connections – is directly applicable to the broader question of regulating the offshore system for tax purposes.

All in all, understanding the network structure, particularly the role of intermediaries as potential chokepoints, reinforces the idea that the current state of offshore finance and associated tax evasion is, as Saez & Zucman (2019) argue in a related context, a continued choice shaped by policy and enforcement priorities, rather than an immutable fact of nature.

2.2.3 Network Theory as a Lens for Understanding Illicit networks

To further contextualize the approach taken in this thesis, it is useful to briefly elaborate on how network studies have previously been employed to explore the structure and dynamics of analogous social and economic systems. The application of network analysis provides powerful tools for understanding complex relational patterns, information flows, and vulnerabilities within various types of networks, including those operating in clandestine or illicit domains.

The foundational work in social network analysis, such as Granovetter's (1973) seminal paper on the "strength of weak ties," laid the groundwork for understanding how network structures facilitate crucial processes like information diffusion and resource access. While initially focused on phenomena like job searching, these core insights into how different types of ties (strong vs. weak) and different network positions (e.g., bridges) shape outcomes have proven broadly applicable. Understanding the topology of connections is essential for identifying critical links, potential weaknesses, and influential actors within any network system. This foundational understanding extends to the analysis of illicit networks, where mapping relationships can reveal operational structures and vulnerabilities.

One of the prominent examples demonstrating the application of network analysis to understand illicit operations is the work of Morselli (2009). By examining specific cases, such as the CAVIAR network involved in cross-border drug smuggling, Morselli illustrates how network science concepts (like centrality measures, brokerage roles, and structural holes) can be used to dissect the organizational structure of criminal enterprises. Such analyses move beyond individual actors to understand the relational patterns that enable the illicit activity, potentially identifying key players or structural weaknesses that could be targeted for disruption.

More directly relevant to the subject matter and data source of this thesis, recent studies have begun applying network analysis to the large-scale datasets released by the ICIJ. Chang et al. (2023), for instance, utilized network methodologies on ICIJ data to specifically examine the effectiveness of sanction regimes against oligarchs, analyzing how their embeddedness within offshore networks influenced outcomes. Similarly, related work by the same authors ("Complex Systems of Secrecy," Chang et al. 2023) employed network perspectives to explore patterns related to the types of offshore instruments demanded by elites, linking structural features to strategic choices. These studies exemplify how network analysis can yield substantive insights from the complex relational data contained within the ICIJ leaks, demonstrating its utility for exploring the offshore financial system.

The general principles and analytical techniques drawn upon in such studies are well-established within the broader field of network science, with standard references like Newman's (2010/2018) textbook providing comprehensive overviews of the underlying theory and methodologies. While this thesis may focus more on synthesis and proposition-building informed by network concepts rather than complex quantitative modeling, drawing upon this established body of work provides a robust conceptual and methodological grounding for analyzing the structure and significance of intermediary networks in offshore finance.

2.2.4 A Typology of Intermediaries and Their Role

To proceed with an analysis centered on the supply-side, it is essential to clarify conceptually what exactly is meant by an "intermediary" within the context of offshore finance. These actors play diverse roles in facilitating the creation, maintenance, and utilization of offshore structures. While specific studies, such as Harrington (2016), provide deep insights into the practices of particular intermediary types like wealth managers, a broader classification is useful for systematic analysis.

These are all what Hoang (2022) would call the "small spiders", the "High net worth indivduals" rather than the "Ultra-High net worth individiduals" sitting at the top of the food chain. Anything uncovered, in this respect is extremely limited, because they are able to further obfuscate their position.

For this purpose, this thesis builds upon the typology developed in a 2017 EU report examining the role of advisors and intermediaries as revealed in the Panama Papers. This framework, grounded in empirical observation of a major leak, categorizes intermediaries based on their primary area of expertise and function within the offshore ecosystem. Adopting this typology serves a dual purpose: it provides conceptual clarity for the subsequent discussion and offers a practical schema for efforts to classify the varied intermediary actors identified within the ICIJ dataset, thereby enriching the data for structural analysis.

Based on the EU (2017) framework, we can distinguish the following core types of intermediaries:

- **Tax Experts:** These intermediaries focus primarily on the tax implications of offshore structures. Their core function involves advising clients on tax planning strategies to minimize liabilities (potentially crossing into evasion) and ensuring compliance through the preparation of necessary tax documentation across relevant jurisdictions. This group can include accountants, auditors, and specialized tax advisors, whose advice may vary in aggressiveness.

- **Legal Experts:** This category encompasses professionals providing expertise on the legal design, establishment, and enforcement of offshore structures. Key activities include structuring entities to navigate or exploit laws in multiple jurisdictions, handling incorporation (often via licensed entities), drafting legal documents, arranging nominee services, and providing formal legal opinions or representation. This group includes regulated lawyers, who often have exclusive rights for certain actions like court representation, and potentially notaries involved in document formalization.

- **Administrators:** The primary role of administrators is the ongoing operational maintenance and financial record-keeping of offshore entities. This includes preparing financial accounts, potentially handling tax returns (overlapping with Tax Experts), managing day-to-day administrative tasks, and sometimes auditing accounts (though auditors require independence). Accountants often fall into this category, focusing on financial recording and reporting.

- **Investment Advisors:** Distinct from those setting up the structure, investment advisors focus on managing the assets held within the offshore entity. Their core function is to develop strategies for wealth preservation or growth using the financial instruments (or other assets like property, art, etc.) owned by the offshore structure. Their role is centered on asset management rather than the legal or tax architecture itself.

This typology provides a decent conceptual grounding for analyzing the distinct roles and potential influence of different supply-side actors within the offshore financial network.

2.2.5 Secrecy Strategies: Financial Instruments and Legal Innovations

Goal: Understanding the different financial instruments they use and how they can be innovated on, and used for different purposes. (Mainly Lafitte, 2024; Chang et al. 2023)

Most important type, Bearer instruments:

Harrington (2016) writes of Bearer instruments as follows: *In addition, a few offshore jurisdictions allow the use of "bearer shares," which are a way of issuing corporate stock without specifying a particular owner. Rather, the owner of a bearer share is literally whoever happens to be holding the stock certificate at any moment in time. This provides strong privacy protections, because as long as one does not have the shares in hand, one can say truthfully under oath, "I do not own that firm." And if any officers of the firm are ever questioned about its ownership, they can also truthfully say, "I don't know who owns the company, because bearer shares were issued." In other words, bearer shares make it impossible to know who owns a company, and that makes it impossible to assign legal responsibility for any taxes, fines, or debts the company incurs.*

2.3 Propositions

Data and Methodology

This section details the data sources and methodological approaches employed in this thesis. It begins by describing the primary data source, the ICIJ Offshore Leaks Database (Section 3.1), and the external datasets used for contextualization (Section 3.3). Subsequently, it introduces a novel methodology utilizing agentic AI to enrich intermediary classification (Section 3.4), outlines the general analytical methodologies applied (Section **??**), and finally comments on the use of LLMs in the thesis preparation (Section 3.6).

3.1 The ICIJ Offshore Leaks Database

The primary empirical basis for this thesis is the International Consortium of Investigative Journalists (ICIJ) Offshore Leaks Database. However, before detailing its structure, the inherent complexities and limitations associated with data derived from leaks concerning a domain deliberately designed for opacity will briefly be detailed.

3.1.1 The Challenges of Obtaining and Interpreting Offshore Data

Researching the offshore financial system is inevitably difficult and fraught with challenges due to the pervasive secrecy that is its defining characteristic (Chang et al., 2023c; Christensen et al., 2022). The ICIJ database, while unparalleled in its scale and granularity, is not a comprehensive or randomly sampled representation of the entire offshore world. It is a compilation of data from specific leaks, each with its own origins and potential biases. For instance, a significant portion of the data originates from specific service providers like Mossack Fonseca (Panama Papers) or Appleby (Paradise Papers). Consequently, the observed patterns in clientele, jurisdictions, and service types may, to some extent, reflect the operational focus and market position of these particular firms rather than the offshore industry in its entirety (De Groen, 2017).

Hoang (2022) in her ethnography, notes of one of the ultra-wealthy board directors she interviews (FLAG: Ensure correct description), that despite the entities he was behind were revealed in the Panama Papers, nothing traces back to him. Instead, a group of "fall guys", as she terms them, are the ones that fall victim to the public's search-light. What we're dealing with in the ICIJ data is the comparatively more visible part of the long stem of offshore that's otherwise buried deep where no sunlight can act as Brandeis' disinfectant.

An additional systematic bias arises from the inclusion of data from public corporate registries in certain jurisdictions as part of specific investigations (e.g., Paradise Papers incorporating registry data from Aruba, Bahamas, Barbados, Cook Islands, Lebanon, Malta, Nevis, and Samoa, as per the ICIJ 'sourceID' variable). This data may include local businesses not engaged in typical offshore activities designed.

Nevertheless, despite these inherent limitations, the ICIJ Offshore Leaks Database represents the most extensive publicly available structured dataset on offshore entities and their associated actors. As Kejriwal & Dang (2020, p. 3(FLAG: which page)) note:

> "[...] [P]recisely because the collection maps out a global system, the Panama Papers also present us with a golden opportunity to study the flow of information2 between firms, individuals and intermediaries. From a scientific perspective, the Panama Papers repre- sent a complex system, with entities that range from individuals to companies, many of which serve a specific purpose based on where in the world they are based, to a variety of relationships. Studying the structural properties of this complex system using applied networks science has the potential to reveal interesting trends about how such systems operate across geographies and economies."

This thesis, therefore, acknowledges that it is analyzing the *visible and structured tip of the iceberg*. Consequently, absolute counts or prevalence estimates for the entire offshore world derived from this data will likely be underestimates and must be interpreted with extreme caution. However, findings pertaining to the *types* of activities, the *characteristics* of observed actors (particularly intermediaries, who are often directly named), and the *structural properties* of the revealed networks are more likely to offer robust, albeit partial, insights into the mechanisms of the offshore financial system. The use of the ICIJ database for such analytical purposes is increasingly established in academic research, with studies employing it to gauge propensities for offshore use (Alstadsæter et al., 2019; Londoño-Vélez & Ávila-Mahecha, 2021) or to explore relationships between offshore structures and political contexts (Chang et al., 2023a; 2023b). This thesis will proceed with similar caution, focusing on patterns and relationships rather than definitive global estimates.

3.1.2 Overview of the ICIJ Offshore Leaks Database

Our primary dataset is the **International Consortium of Investigative Journalists (ICIJ) Offshore Leaks Database**. This publicly accessible repository is a comprehensive amalgamation of structured data meticulously extracted from several of ICIJ's landmark global investigations, most notably the Offshore Leaks (2013), Panama Papers (2016), Paradise Papers (2017/18), and Pandora Papers (2021/22). The database is substantial, cataloging information on over 810,000 offshore entities—which encompass a range of structures such as companies, trusts, and foundations—and establishing connections to more than 750,000 individuals and corporate entities. These connections span across a vast geographical landscape of over 200 countries and territories, with the underlying records covering a significant historical period, in some cases extending up to the year 2020.

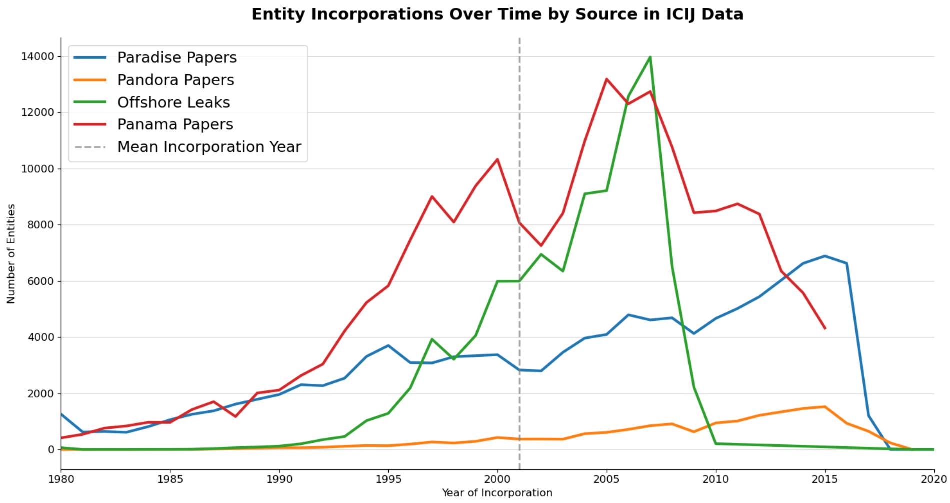

Figure 3.1: Overview of Entity Incorporations Over Time from ICIJ Data

Before getting into the explanation, it is important to note that the overview provided here is relatively cursory and focuses mostly on the attributes and feature engineering specific to this thesis. For those more familiar with network analysis, I'd strongly encourage Kejriwal & Dang's (2020) to get a more in-depth understanding of the data in more graph-theoretical terms.

The fundamental data model leveraged by the ICIJ data is a graph database. This model is used for its ability to represent interconnected information, conceptualizing data as **nodes** (the core informational units) and **edges** (the links defining how these units are connected). For the purposes of our study, the most pertinent node types are:

- **Entities**: These represent the diverse offshore legal structures documented in the leaks, such as Limited companies, S.A. (Société Anonyme), Inc. (Incorporated), trusts, and foundations.

- **Officers**: This category includes individuals or, in some instances, other corporate bodies that fulfill specific roles (e.g., director, shareholder, beneficial owner, trustee, protector, nominee) within an Entity.

- **Intermediaries**: These are the professional facilitators—typically law firms, accounting practices, banks, trust companies, or specialized middlemen—who assist clients in the establishment and ongoing management of offshore entities. They often act as the liaison with offshore service providers like Mossack Fonseca or Appleby.
- **Addresses**: These nodes capture physical location data associated with the other node types, such as the registered office of an Entity or the business address of an Intermediary.

Relationships (edges) within this graph structure explicitly define the nature of the connections, for example, an Officer is an `officer_of` an Entity, or an Intermediary acts as an `intermediary_of` an Entity. The two primary node types of interest for this thesis are **Entities** and, critically, **Intermediaries**. In the ICIJ data model, the role of intermediaries is, with very few exceptions, represented entirely through their connections to Entities. That is, at a high level, a common relational pathway is: Intermediaries are `intermediary_of` Entities, which in turn have Officers (who are `officer_of` these Entities).

3.1.3 Entities

Delving deeper into the **entities**, the information processed from source files such as `nodes-entities.csv` and `relationships.csv` provides a rich set of attributes for each. Key data points include the entity's registered `name`, its `jurisdiction` of incorporation (which is standardized to ISO3 country codes for consistent geographical analysis), and the `country_codes` associated with its operational activities or linked addresses. These `country_codes` are often distinct from its legal `jurisdiction` of incorporation and provide insights into the geographical footprint of the entity's actual business or connections. Further attributes encompass the `incorporation_date`, its operational `status` (e.g., Active, Struck Off, Dissolved), and its specific `entity_type` (e.g., Standard International Company, Trust, Business Company Limited by Shares).

A particularly significant feature derived for each entity is the `bearer_count`. This metric quantifies the number of associated officers explicitly identified as "Bearer" or its linguistic equivalents (e.g., "THE BEARER," "EL PORTADOR"), which are standardized from variations found in the `officers_df`. The presence of bearer instruments, as highlighted by Harrington (2016), is a critical indicator of mechanisms used to obscure true beneficial ownership. In such arrangements, legal ownership follows the physical possession of the share certificate rather than being recorded in a central register, thereby enhancing anonymity (Chang et al., 2023c).

3.1.4 Intermediaries and Feature Engineering

For **intermediaries**, whose foundational data is drawn from `nodes-intermediaries.csv`, the analysis extends beyond basic identifying information. Beyond their `name` and the `countries` associated with their operational addresses, we calculate their `degree`. In this context, the degree represents the total number of distinct entities an intermediary is connected to within the ICIJ network, serving as a proxy for their client base size or activity level.

More extensively, we construct several aggregated metrics that characterize each intermediary based on the collective properties of the entities they service. As our primary research interest lies in understanding the roles and specializations of intermediaries, we aggregate information at the intermediary-level about the entities they are connected to. While graph data models excel at representing complex, interrelated data, extracting features for broader statistical analysis often necessitates such aggregation into key-value attributes (Kejriwal & Dang, 2020).

For every intermediary, we generate the following features:

- `country_counts`: A dictionary detailing the frequency of entities they are connected to, grouped by the `country_codes` associated with those entities. This reflects the geographical spread of the operational links of the entities they service. The derivation of these `country_codes` from address fields in the original leaks means their completeness and precision can vary, a factor considered in interpreting derived metrics.

- `jurisdiction_counts`: A dictionary detailing the frequency of entities they are connected to, grouped by the `jurisdiction` (ISO3 code) in which those entities are incorporated. This captures the intermediary's usage of different offshore legal environments.

- `regime_counts`: A dictionary detailing the frequency of entities they are connected to, grouped by the political regime type (e.g., Liberal Democracy, Closed Autocracy, as per VDem data detailed in Section 3.3) of the entities' associated `country_codes` at the time of entity incorporation. This provides insight into the political contexts linked to an intermediary's client base. Note, VDem do not classify a lot of those countries that are tax havens (e.g. Bahamas, British Virgin Islands etc.), because their methodology is not robust to these countries that are as small (FLAG: Find that appendix in VDem). For the sake of this thesis, when the VDem data cannot be matched, we assign the regime type as "Microstate".

- `legal_tech_counts`: A dictionary detailing the frequency of entities they are connected to, grouped by the predominant types of "legal technologies" (e.g., Banking, Corporate, Dual-Purpose, as per Laffitte (2024), detailed in Section 3.3) prevalent in the entities' jurisdictions of incorporation at the time of their formation. This reflects an intermediary's engagement with specific offshore legal architectures.

Furthermore, we quantify for each intermediary the number of entities they are connected to that have `bearers_connected` (i.e., entities with a `bearer_count` > 0) and calculate the `bearer_share`, representing the proportion of their serviced entities that utilize these anonymity-enhancing instruments.

To measure the diversity of their client entity portfolio across these dimensions, we also compute normalized entropy scores: `country_entropy`, `jurisdiction_entropy`, `regime_entropy`, and `legal_tech_entropy`. More on that in Section 3.5.2. These scores provide a measure of the diversity of the intermediary's client base across the respective dimensions, with higher values indicating a more diverse portfolio.

3.2 Other Data Sources

To enrich the core ICIJ data, several external datasets are integrated, primarily to provide contextual information at the country or jurisdiction level, and to classify intermediaries by type.

- **Laffitte Legal Technologies Data (HTHD)**: This dataset (Laffitte, 2024) is used for connecting historical legal framework changes to entities and their structuring, specifically identifying the "legal technologies" active in a jurisdiction at the time of an entity's incorporation.
- **VDem (Varieties of Democracy) Data**: This provides country-level variables, notably political regime types, for the jurisdictions and countries associated with entities and intermediaries.
- **Intermediary Type Enrichment**: As detailed in Section **??** (updated label), an agentic AI approach is employed to classify a subset of intermediaries based on publicly available information scraped from the internet.

3.3 Other Data Sources

To enrich the core ICIJ data, several external datasets are integrated, primarily to provide contextual information at the country or jurisdiction level, and to classify intermediaries by type.

- **Laffitte Legal Technologies Data (HTHD)**: This dataset (Laffitte, 2024) is used for connecting historical legal framework changes to entities and their structuring, specifically identifying the "legal technologies" active in a jurisdiction at the time of an entity's incorporation.
- **VDem (Varieties of Democracy) Data**: This provides country-level variables, notably political regime types, for the jurisdictions and countries associated with entities and intermediaries.
- **Intermediary Type Enrichment**: As detailed in Section 3.4, an agentic AI approach is employed to classify a subset of intermediaries based on publicly available information scraped from the internet.

At the country and jurisdiction level, we utilize data from the Varieties of Democracy (VDem) project for information on political regime types, and Sébastien Laffitte's (2024) Historical Tax Havens Database (HTHD), developed for his doctoral thesis, which provides detailed information on the evolution of legal technologies in various jurisdictions.

1. The **Varieties of Democracy (VDem) Project data** (specifically `vdem_core.csv`): We utilize the `v2x_regime` variable from VDem's comprehensive dataset to enrich our entity data. This variable classifies countries into categories such as Closed Autocracy, Electoral Autocracy, Electoral Democracy, or Liberal Democracy. By matching an entity's associated `country_codes` (representing operational links) and its `incorporation_year` with the VDem data for the corresponding country and year, we assign a political regime classification to each entity. This entity-level regime information is then aggregated to construct the `regime_counts` at the intermediary level, providing insight into the political environments linked to an intermediary's clientele.
2. **Laffitte's (2024) "The Market for Tax Havens" dataset** (specifically `HTHD.csv`): This dataset offers a historical perspective on the "offshore legal architecture" of various jurisdictions, detailing their adoption of different "legal technologies" such as International Business Company (IBC) laws, trust legislation, or banking secrecy provisions. Laffitte categorizes these into broader types such as "Banking," "Corporate," "Dual-Purpose" (e.g., IBCs serving both personal and corporate needs), and "Personal" (e.g., trust laws). We merge this dataset onto our entity data by matching the entity's `jurisdiction` of incorporation and its `incorporation_year` with the HTHD data. This allows us to identify the specific legal technologies active in an entity's jurisdiction at its time of incorporation. This entity-level characterization is subsequently aggregated to create the `legal_tech_counts` at the intermediary level, reflecting the types of legal environments their serviced entities operate within.

3.4 Agentic AI for Intermediary Classification

Directly at the Intermediaries-level, we also enrich a **subset of intermediaries** (specifically, a random sample of 500 and the top ∼1.5% by degree, chosen to balance representativeness with computational feasibility for the AI agent) with information on their specific "type." This classification is based on a typology adapted from the EU (2017) report on the Panama Papers (De Groen, 2017), which identifies roles such as Tax Expert, Legal Expert, Administrator, and Investment Advisor.

The core idea is to use an AI agent loop to automate the process of gathering information about and classifying the intermediaries listed in the ICIJ data. The basic workflow is illustrated in Figure 3.2.

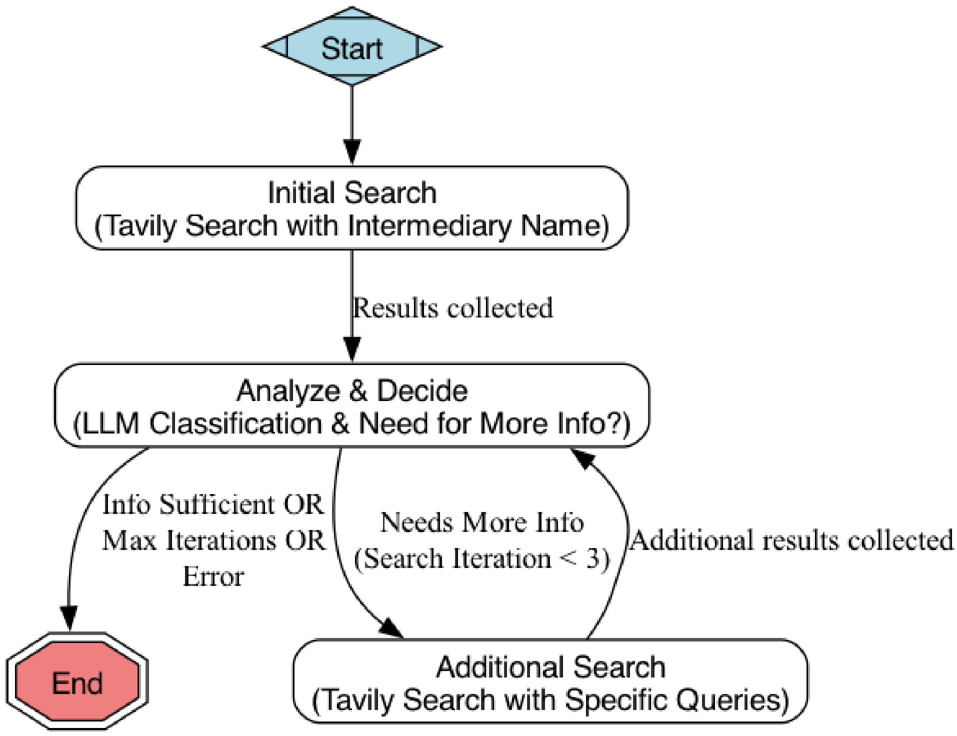

Figure 3.2: Agent Setup for Intermediary Classification

In brief, the process involves an AI agent orchestrating online searches for each intermediary identified in the ICIJ data. It begins with generic searches, reads and interprets the initial results, and then formulates more specific search queries based on the information discovered or identified as lacking. This iterative process involves up to three search queries per intermediary, scouring the top 15 most relevant web results identified through query-result embedding similarity using the Tavily Search API (though the tool is relatively generic and its specific choice is not critical to the methodology). This effectively replaces the time-consuming need for manual searching of the intermediaries.

Based on the information gathered, the AI agent then classifies the intermediary according to the De Groer (2017) typology (Tax Expert, Legal Expert, Administrator, Investment Advisor), adding a few additional relevant fields (e.g., specific job title). To mitigate some of the obvious fallibility of such an enrichment method, the agent also provides a confidence score for its classification judgment, which is filtered on in the analysis.

3.5 Analytical and Statistical Techniques Applied

This section outlines the core analytical techniques applied to the processed data, with a primary emphasis on concepts from network theory for characterizing the structure of the ICIJ data. These network methods are central to the thesis, while other approaches, such as unsupervised learning for pattern discovery and statistical tests for assessing significance, serve as ancillary tools and will be discussed more briefly.

3.5.1 Concepts from Network Analysis

The application of network analysis is fundamental to this thesis, drawing on a tradition of using such methods to understand the nature of complex, often covert or illicit, systems (Morselli, 2009). Specifically, network analysis is employed here to uncover the roles intermediaries play based on their positions within the interconnected offshore financial system revealed by the ICIJ data. As described in Section 3.1, the ICIJ data forms a multi-modal graph (comprising entities, officers, intermediaries, etc.). Directly applying many standard network analysis concepts to such a multipartite graph can be challenging. Therefore, our approach often involves analyzing specific projections or subsets of the global graph to make the analytical tools from network theory applicable. The foundational textbook by Newman (2010) serves as the primary reference for this section.

- **Centrality Scores**: To identify nodes of critical importance within specific network representations, we utilize two fundamental centrality measures. In the context of understanding the key countries for intermediary activity, these measures are applied to a network derived from intermediary incorporation patterns.
  - **Eigenvector Centrality**: This measure assigns scores to nodes based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. It is calculated as the principal eigenvector of the adjacency matrix $\mathbf{A}$ of the network, satisfying $x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j$, where x_i is the centrality score of node i, A_{ij} is 1 if node i is connected to node j and 0 otherwise (or the weight of the edge), and λ is the largest eigenvalue of $\mathbf{A}$ (cf. Perron-Frobenius theorem). Eigenvector centrality is chosen for its ability to identify nodes that are influential not just by having many connections, but by being connected

to other influential nodes, providing a robust reading of which countries are most central in the network of intermediary incorporations.

– **Betweenness Centrality**: This metric quantifies the extent to which a node lies on shortest paths between other pairs of nodes. For a node v, it is defined as $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ is the number of those paths that pass through v. Betweenness centrality is used here to gauge which countries act as crucial "bridges" or conduits within the network, potentially connecting otherwise disparate segments, a role distinct from simply being a high-degree hub.

- **Community Detection: Modularity Maximization**: To uncover clusters or communities of closely related nodes within the country network, we employ modularity maximization. This approach provides an atheoretical method for identifying densely connected groups of countries, which may reflect underlying similarities in how intermediaries utilize them. Such clustering could be influenced by factors like shared regime types (Chang et al., 2023c) or the trust dynamics inherent in relational capitalism. While traditional clustering algorithms could be applied, defining a meaningful distance or dissimilarity metric for nodes in these networks is non-trivial. Modularity maximization, conversely, assesses the quality of a partition by comparing the number of intra-community edges to what would be expected in a random network with similar properties (a null model). The quality of a partition C is measured by the modularity Q:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - P_{ij}\right] \delta(c_i, c_j) \tag{3.1}$$

where m is the total number of edges, A_{ij} is the actual weight of the edge between nodes i and j, P_{ij} is the expected weight of an edge between i and j under the Newman-Girvan null model (a configuration model preserving the degree sequence, where $P_{ij} = \frac{k_i k_j}{2m}$ for unweighted graphs, k_i being the degree of node i), and $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community ($c_i = c_j$) and 0 otherwise. Since finding the optimal partition is an NP-hard (although, to be honest, at the size we reduce our graph sizes, search space isn't an issue...) problem, we utilize the Louvain method (Blondel et al., 2008), an efficient and widely adopted greedy algorithm, as implemented in the `networkx` library.

- **Power-law Distribution**: The distribution of node degrees (number of connections) and other network properties are examined for characteristics of power-law distributions. A power law, $P(k) \sim k^{-\alpha}$, describes a "fat-tailed" distribution where a few nodes (hubs) have a disproportionately high number of connections, while most nodes have few. Such distributions are frequently observed in real-world networks (Clauset et al., 2009; Kejriwal & Dang, 2020) and their presence can indicate significant heterogeneity in node importance.

- **Density of a Graph**: Network density, the ratio of actual edges to the total number of possible edges in the network ($D = \frac{L}{N(N-1)/2}$ for an undirected graph with L edges and N nodes), is used to measure the general level of connectedness. Low density is typical for large, sparse networks and indicates that connections are selective rather than ubiquitous.

3.5.2 Entropy

Drawing on its application in prior studies of offshore finance (e.g., Chang et al., 2023c; Kejriwal & Dang, 2020), Shannon entropy is employed as a measure of diversity or concentration. For a discrete random variable X with n possible outcomes $x_1, ..., x_n$ and probabilities $p(x_i)$, entropy is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i) \tag{3.2}$$

where b is the base of the logarithm (typically $b = 2$, yielding units of bits). Compared to other concentration measures like the Herfindahl-Hirschman Index (HHI), entropy gives more weight to smaller amounts of diversity. This characteristic is particularly useful in this thesis, as intermediaries' activities (e.g., choice of jurisdictions or countries) are often highly concentrated in one or two locations, but variations in minor activities can still be informative. Normalized entropy, calculated by dividing $H(X)$ by the maximum possible entropy ($\log_b n$), is used to provide a standardized measure (0 to 1) for comparing diversity across intermediaries with different breadths of activity. Entropy is used, for example, as a summary statistic at the intermediary-level to quantify the diversity of their client entity portfolio across dimensions like country, jurisdiction, or regime type, enabling subsequent comparisons of these distributions across different intermediary classifications.

3.5.3 Association Analysis

In line with the highly exploratory nature of this thesis, unsupervised learning techniques are employed to discover notable patterns within the data. Association analysis (Hastie et al., 2009) is particularly opportune for identifying non-obvious relationships or co-occurrences in large datasets, such as the ICIJ networks. For example, it can help determine which connections (e.g., between a type of intermediary and the use of a specific jurisdiction or legal technology) are particularly remarkable. This approach relies on a **non-parametric notion** of pattern discovery, aiming to **discover patterns of high**

density or co-occurrence.

Two main tools from association analysis, based on simple set-theoretical notions, are used:

- **Support**: This measures the overall frequency of an itemset (e.g., a specific attribute or combination of attributes) in the dataset. For an itemset X, $Support(X) = P(X) = \frac{\text{count}(X)}{N}$, where N is the total number of transactions (e.g., intermediaries). For an association rule $A \rightarrow B$, $Support(A \rightarrow B) = P(A \cup B)$.

- **Lift**: This measures how much more likely item B is to be present when item A is present, compared to the baseline probability of B. It indicates the strength of an association beyond what would be expected by chance.

$$Lift(A \rightarrow B) = \frac{P(B|A)}{P(B)} = \frac{Support(A \cup B)}{Support(A) \times Support(B)} \tag{3.3}$$

  A lift value greater than 1 suggests a positive association, a value less than 1 suggests a negative association, and a value of 1 suggests independence. **Lift scores** will be used to quantify the strength of associations found, indicating, for example, how much more likely an intermediary of a certain type is to use a specific jurisdiction compared to the overall likelihood.

3.5.4 Multiple Hypothesis Testing

Given that this thesis is highly exploratory and investigates a multitude of potential associations, it is crucial to address the issue of multiple hypothesis testing. When numerous statistical tests are performed, the conventional Type I error rate of 5% ($p < 0.05$) can become inflated, leading to a higher probability of false positives (incorrectly rejecting a true null hypothesis). To counteract this, a highly conservative approach is adopted, opting for the **Bonferroni correction** to control the Family-Wise Error Rate (FWER) at the conventional maximum of 5%. This method adjusts the significance threshold for each individual test to α/m, where α is the desired FWER (e.g., 0.05) and m is the total number of hypotheses tested. Alternatively, individual p-values are multiplied by m, and then compared to α. While known for its conservatism, this choice is made to be particularly cautious about any single false positive claim, given the exploratory nature of the analysis, rather than opting for procedures like the Benjamini-Hochberg method which control the False Discovery Rate (FDR).

3.5.5 Testing Significance of Results

In line with the considerations above, and the nature of the data, specific non-parametric statistical tests are employed to assess the significance of observed differences or associations. These tests are chosen for their robustness to violations of normality assumptions. The following are applied where appropriate, with detailed applications described in the empirical analysis chapter:

- **Mann-Whitney U test**: A non-parametric test used for comparing the distributions of continuous or ordinal variables between two independent groups. It is particularly useful when the data is not normally distributed, as is often the case with metrics like entropy scores or network-derived measures. It assesses whether one distribution is stochastically greater than the other.

- **Fisher's exact test**: Employed for analyzing categorical data, particularly in contingency tables (e.g., 2x2 tables). This test is ideal for assessing associations between categorical variables, such as those resulting from association analysis or when examining the relationship between dummy variables (e.g., whether entities are connected to bearer instruments and intermediary type). It is an exact test, making it suitable for small sample sizes or when expected cell counts are low.

- **Two-sample Kolmogorov-Smirnov test**: Used for comparing the underlying distributions of continuous variables from two independent samples. Unlike tests that compare central tendencies (like the t-test or Mann-Whitney U), the K-S test is sensitive to differences in location, scale, and shape of the distributions, offering a more comprehensive comparison.

3.6 Use of LLMs in the Broader Paper

LLMs have also been used to polish the text of this thesis and used for idea generation.

Used Google Gemini models mainly.

- gemini-2.5-pro-preview-05-06

- gemini-2.5-pro-experimental-03-25

- gemini-2.5-flash-experimental-04-17

Quick edits frequently made using Claude's 3.7 Sonnet model (‘claude-3.7-sonnet-latest‘).

Empirical Analysis

4.1 Overview of the Dataset

This section provides an initial descriptive overview of the aggregated ICIJ dataset, highlighting key structural characteristics that inform the subsequent, more focused and detailed analyses. We begin by examining the geographical concentration of entities, intermediaries, and officers, followed by the degree distribution of intermediaries.

Concentration of Entities, Intermediaries, and Officers

A striking initial observation from the data is the high degree of geographical concentration in offshore activities across all node types. The complex web of global offshore finance, while spanning over 200 countries and territories in the ICIJ data, appears to be significantly anchored in a relatively small number of key locations. Approximately 98% of all entities incorporated within the dataset are registered in just 15 jurisdictions. This intense concentration in a few select jurisdictions aligns with literature suggesting that offshore financial centers (OFCs) often specialize in particular 'legal technologies' or cater to specific clienteles, leading to certain jurisdictions becoming dominant hubs for offshore incorporation (Laffitte, 2024).

This pattern of concentration extends to the actors involved. When examining the top 15 countries or jurisdictions associated with each category, we find that around 70% of entities (by their associated operational countries), 70% of intermediaries (by their listed country of operation), and 70% of officers (by their listed country) are concentrated within these leading locations. Figure 4.1 provides a visual representation of these geographical distributions, illustrating the percentage of entities located in specific jurisdictions, entities active in various countries, intermediaries based in different countries, and officers linked to particular countries. This concentration of intermediaries, for instance, may reflect the clustering of professional services in major global financial centers, as noted by Stausholm and Garcia-Bernardo (2024) - though they concentrate specifically on tax experts - rather than solely in traditional tax havens.

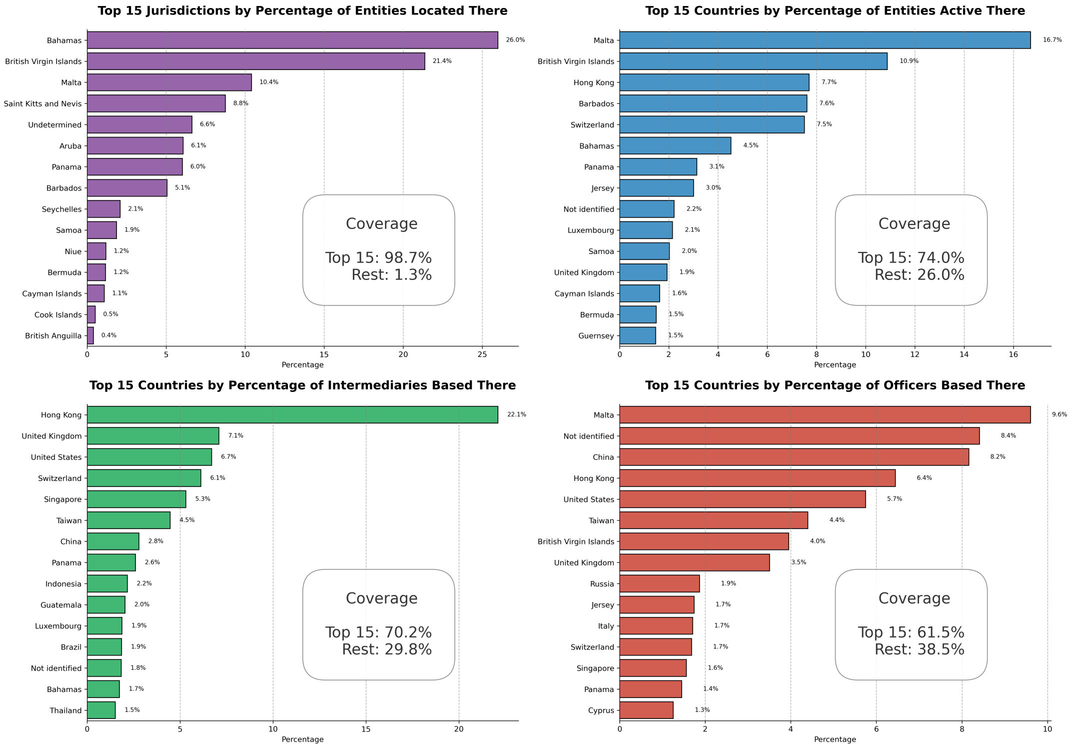

Figure 4.1: Geographical Concentration of Entities, Intermediaries, and Officers. Bar charts illustrate the percentage of total entities incorporated in the top 15 jurisdictions, entities with activity linked to the top 15 countries, intermediaries based in the top 15 countries, and officers associated with the top 15 countries.

Degree Distribution of Intermediaries

A recurring theme, and one that underpins much of the analytical framework of this thesis, is the prevalence of power-law-like distributions in network metrics. This is particularly evident in the degree distribution of intermediaries, as illustrated in Figure 4.2. The degree of an intermediary, in this context, represents the number of distinct entities they are connected to, serving as a proxy for their client load or market reach.

The observed distribution indicates that a small number of intermediaries are connected to a very large number of entities, forming "super-hubs" of activity, while the vast majority of intermediaries have relatively few connections. This characteristic aligns with findings from structural studies of similar datasets, such as Kejriwal and Dang's (2020) analysis of the Panama Papers, which also identified power-law degree distributions. Such a distribution suggests a highly heterogeneous system where certain intermediaries play a disproportionately significant role.

To formally assess this, the fit of the power-law model is compared to that of a log-normal distribution for the intermediary degree data. This comparison yielded a log-likelihood ratio $R = 57.0287$ with a p-value < 0.0001 (detailed in Section 3.5 and implemented as per Clauset et al., 2009). This result strongly suggests that a power-law model provides a statistically significantly better fit than a log-normal model, or at the very least, indicates that the distribution is distinctly heavy-tailed. The implications of this scale-free or heavy-tailed characteristic are profound: it points towards a system where a few key intermediaries may act as critical "enablers" or "chokepoints" (Chang et al., 2023b; Christensen, 2024).

4.2 Geographical Specialisation

This section delves into the geographical patterns exhibited by intermediaries, focusing on the locations of their clients and the jurisdictions they select for entity incorporation. It covers specialization at both the country level of intermediary operation and through network analysis of co-service patterns across entities and jurisdictions.

Intermediary Specialisation at the Country Level

To understand how intermediaries based in specific countries tailor their services, we examine their client footprints and their preferred jurisdictions for incorporation. Heatmaps visually represent these patterns for intermediaries based in the top 15 countries by intermediary count: Hong Kong (HKG), Great Britain (GBR), the United States (USA), Switzerland (CHE), Singapore (SGP), Taiwan (TWN), China (CHN), Panama (PAN), Indonesia (IDN), Guatemala (GTM), Luxembourg (LUX), Brazil (BRA), Bahamas (BHS), Jersey (JEY), and Thailand (THA). Figures 4.3, 4.4, and 4.5 display these heatmaps,

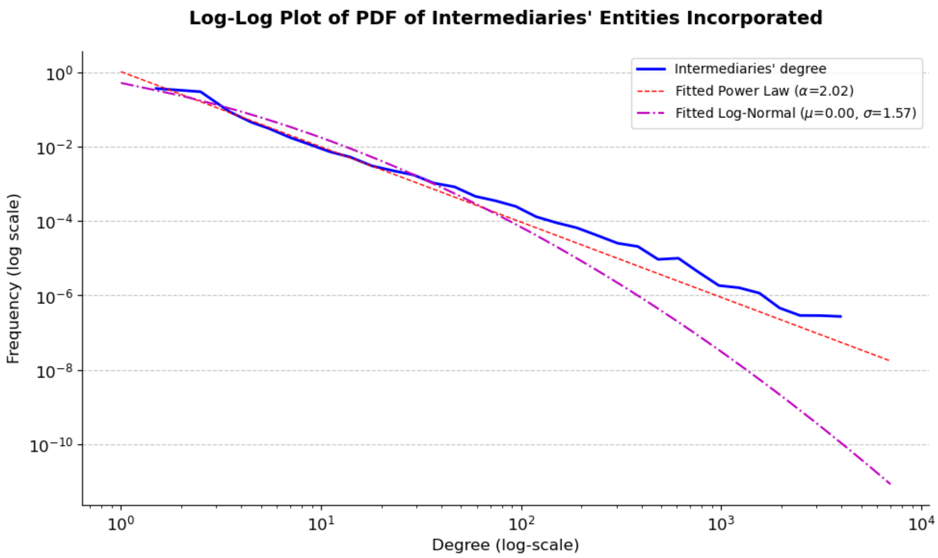

Figure 4.2: Degree Distribution of Intermediaries and Model Fits. The plot shows the probability density function (PDF) of intermediary degrees on a log-log scale. The empirical data (blue line) is compared against a fitted power-law distribution (red dashed line, $\alpha \approx 2.08$) and a fitted log-normal distribution (purple dash-dot line).

detailing the top client countries (by entity activity) and top incorporation jurisdictions for intermediaries in these key operational hubs. These visualizations generally indicate that intermediaries often serve a significant proportion of clients from their own country of operation, while their choice of incorporation jurisdictions tends to be far more outwardly focused and diverse.

The heatmaps reveal distinct national profiles. For instance, intermediaries in major financial centers like Hong Kong and Great Britain show relatively broad international client bases and utilize a range of offshore jurisdictions.

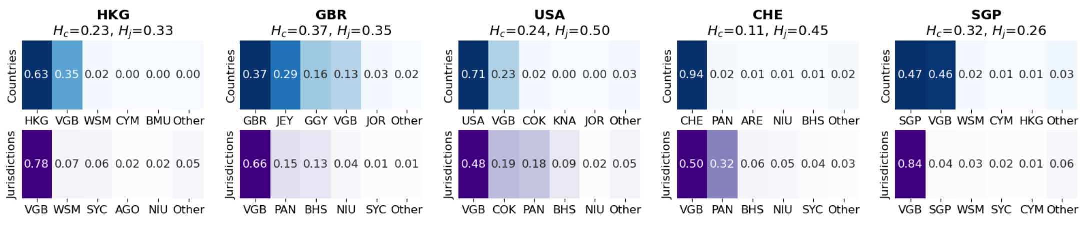

Figure 4.3

A general trend observed is that intermediaries tend to be more diversified in their choice of *incorporation jurisdictions* for their clients' entities compared to the geographical spread of the *countries where their clients' entities are primarily active.* This is quantified by comparing the normalized entropy of jurisdictions used for incorporation (H_j) with the normalized entropy of client entity countries (H_c) at the aggregate level for intermediaries within each country. As shown in Figure 4.6, the distribution of jurisdiction entropy (mean = 0.48) is significantly higher than that of client country entropy

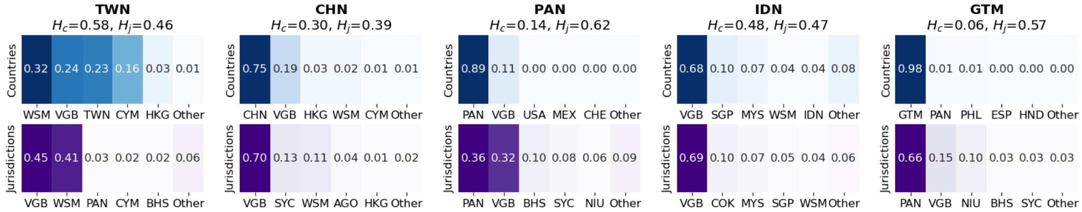

Figure 4.4: Client and Incorporation Jurisdiction Heatmap for Intermediaries in Top 6-10 Countries (TWN, CHN, PAN, IDN, GTM). Panels follow the same format as Figure 4.3.

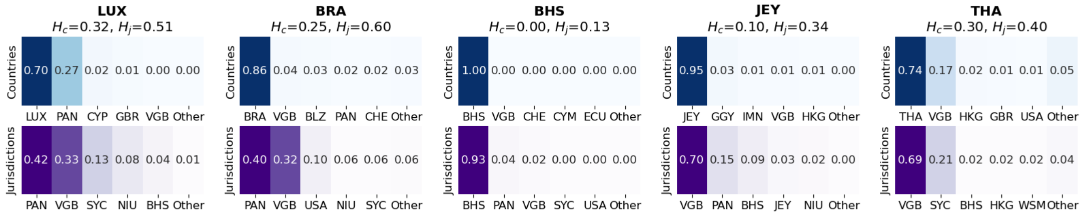

Figure 4.5: Each panel shows the distribution of client entity countries and incorporation jurisdictions for intermediaries based in the specified country. H_c denotes country entropy and H_j denotes jurisdiction entropy for that group of intermediaries.

(mean = 0.23). A two-sample Kolmogorov-Smirnov test confirms that these two distributions are statistically different (KS test: $p \approx 3 \times 10^{-7}$), suggesting that while client bases may be geographically somewhat concentrated, intermediaries draw upon a wider palette of offshore jurisdictions to structure entities. This aligns with the notion that intermediaries strategically select from a global "market for tax havens" (Laffitte, 2024) to meet diverse client needs, whereas client acquisition might be more localized.

An illustrative example of specific geographical specialization is Cyprus (Figure 4.7). Cyprus is well-documented in academic and policy literature for its strong financial links to Russia (e.g., Alstadsæter et al., 2022, note similar patterns for Dubai). While Russia is generally underrepresented in the broader ICIJ dataset, entities incorporated by Cypriot intermediaries show a significant Russian presence, with 12% of such entities linked to Russia. This suggests a strong, specific association (a high "lift" in association analysis terms - though not formally tested here) for the Cyprus-Russia connection, as Russia appears minimally in the client portfolios of intermediaries from most other countries. The entropy values for Cyprus-based intermediaries ($H_c = 0.55, H_j = 0.53$) indicate moderate diversification in both client countries and incorporation jurisdictions.

Network of Countries Served by Intermediaries

While intermediaries operating at the country level exhibit specialization, particularly in their client bases, this section examines the specific clusters of countries served by individual intermediaries. At the individual intermediary level, clientele also tends to be highly concentrated in one or two countries, as shown by the distribution in Figure 4.8.

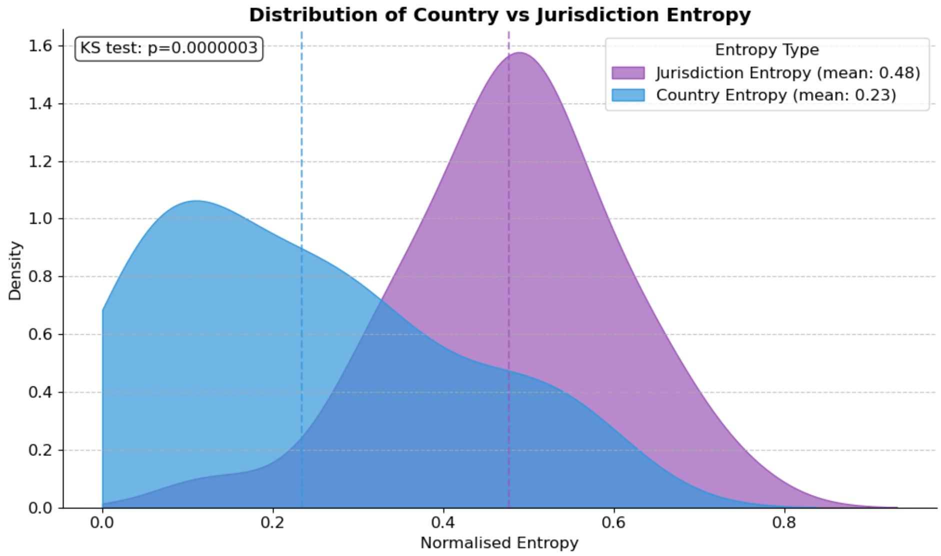

Figure 4.6: Distribution of Normalized Entropy for Client Countries vs. Incorporation Jurisdictions, Aggregated at the Country Level of Intermediaries. The plot shows that intermediaries, when grouped by their country of operation, tend to use a more diverse set of incorporation jurisdictions (H_j, mean = 0.48) than the diversity observed in their clients' countries of activity (H_c, mean = 0.23).

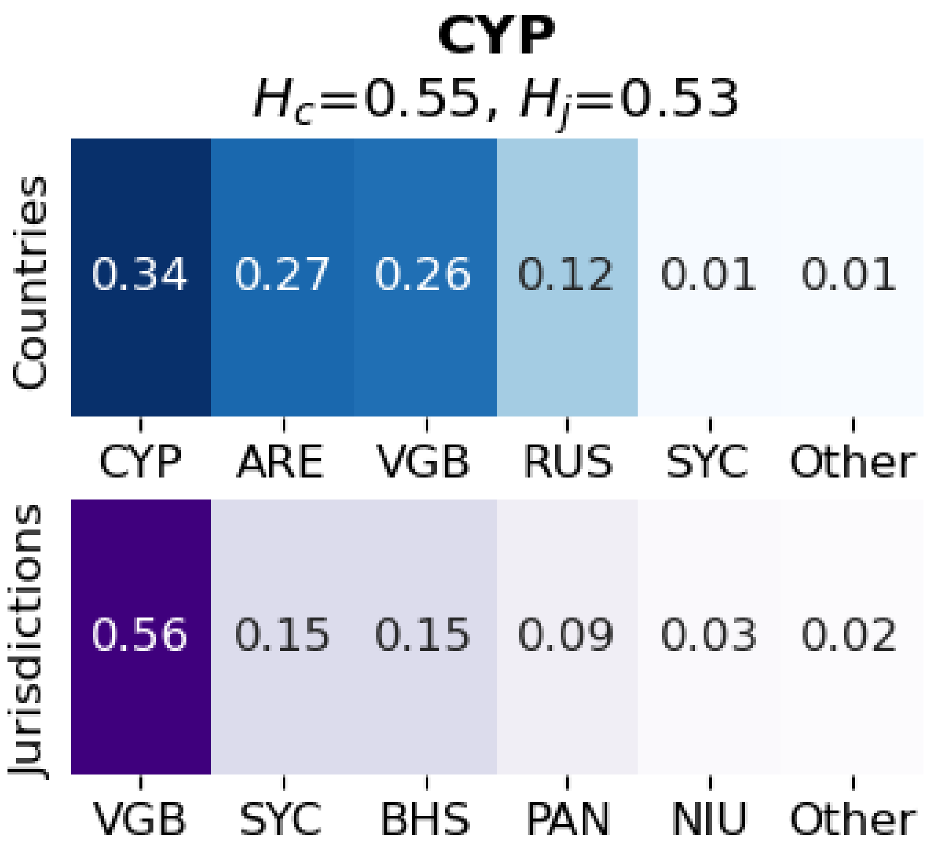

Figure 4.7: Client and Incorporation Jurisdiction Heatmap for Cyprus-based Intermediaries ($H_c = 0.55, H_j = 0.53$). The heatmap shows that 12% of entities serviced by Cypriot intermediaries have links to Russia; a notable concentration.

This indicates that most intermediaries focus their client acquisition efforts narrowly. Interestingly, even as intermediaries grow larger (i.e., serve more entities, reflected by their degree), there is a very low correlation between the number of entities served and the number of distinct countries their clients' activities are linked to, suggesting that scaling often occurs through deeper penetration within existing client geographies rather than broad international expansion of the client base.

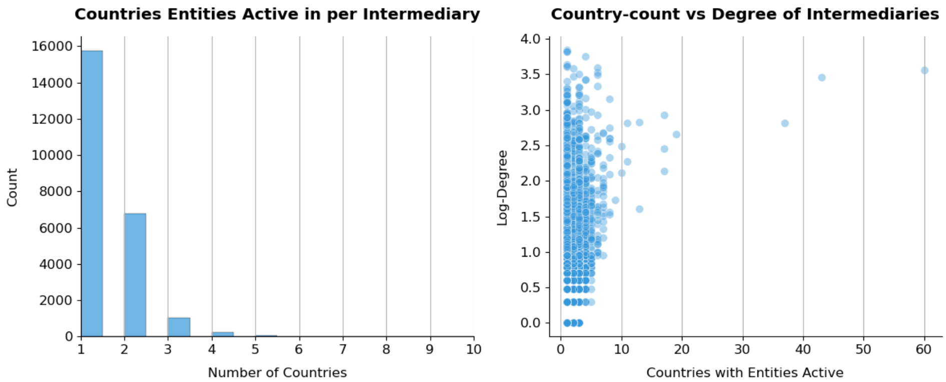

Figure 4.8: Distribution of the Number of Countries Linked to Entities Served per Intermediary. The distribution is heavily skewed, with most intermediaries serving entities linked to only one or a few countries - even high-degree intermediaries.

To explore these co-service relationships further, a network of countries was constructed. In this network, countries are nodes, and an edge exists between two countries if at least one intermediary serves clients (entities) linked to both. The weight of the edge reflects the number of distinct intermediaries serving clients in both countries. The resulting full country network consists of 121 nodes (countries) and 2,716 edges. Key summary statistics for this network are presented in Table 4.1.

Table 4.1: Summary Statistics for the Full Country Co-Service Network

Metric	**Value**
Number of Nodes	121
Number of Edges	2716
Network Density	0.3741
Average Degree	44.89
Average Clustering Coefficient	0.7728

Visualising such dense graphs is incredibly challenging - and to be entirely honest, the rest of the thesis could be filled with differently filtered versions of this graph, illuminating some other aspect of it. Therefore, to identify the most important connections, the network was filtered using principles from association analysis. Edges are displayed only if they 1) meet a minimum support threshold (representing at least 0.008 of all intermediaries' country-pair connections, meaning the pair is co-serviced by at least that fraction of intermediaries who service multiple countries) and 2) a lift score of 1.5 or higher. Lift measures how much more frequently two countries are co-serviced than would be expected if their servicing by intermediaries were independent. This filtering ensures that the visualized connections are not only reasonably frequent but also represent associations significantly stronger than chance, that they are both common links as well as carrying statistical signal. The resulting filtered network, or "backbone," thus highlights the most robust and significant co-service relationships. While the exact number of nodes included is sensitive to the choice of the lift and support thresholds here, this "backbone" as I term it, is relatively stable across a range of thresholds.

The nodes in the network visualization (Figure 4.9) are coloured in two ways: first, by communities identified using the Louvain modularity maximization algorithm (Blondel et al., 2008), which groups densely interconnected countries; and second, by regime type using VDem data, as described in Section 3.3. This dual coloring was intended to explore whether regime type influences intermediary operations and co-service patterns, a factor suggested by literature on offshore secrecy strategies (e.g. Chang et al., 2023b).

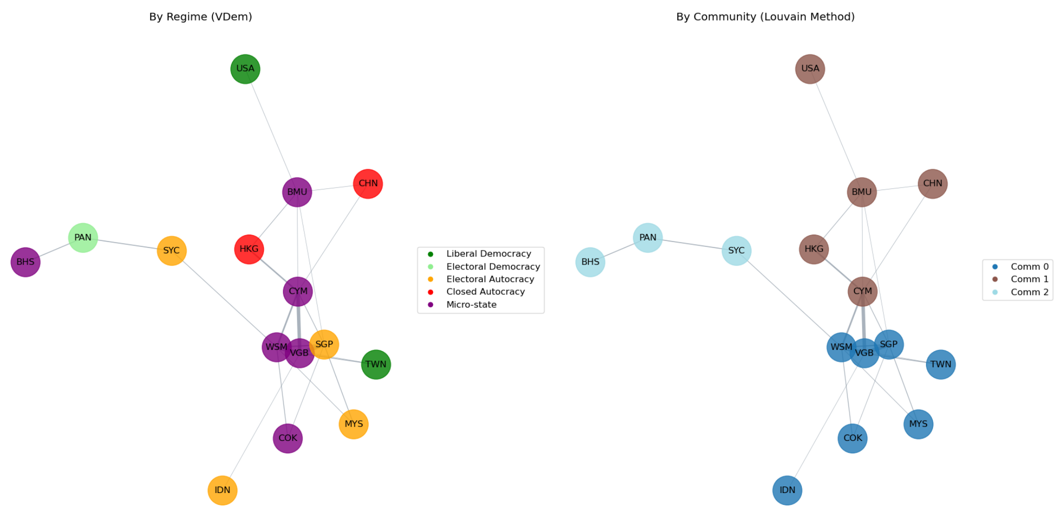

Figure 4.9: Filtered Network of Co-Served Countries, Coloured by Louvain Community (left legend) and Regime Type (right legend). Edges shown have support ≥ 0.008 and lift ≥ 1.5. Node size can be proportional to degree or another centrality measure.

Interpretation of the Filtered Country Network Structure The filtered network (Figure 4.9) reveals a sparse yet highly structured set of relationships, forming a distinct core-periphery structure. A central core of interconnected nodes is evident, particularly involving VGB (British Virgin Islands), CYM (Cayman Islands), and SGP (Singapore), along with their strong links to HKG (Hong Kong) and BMU (Bermuda).

When coloured by regime type, no clear large-scale clustering emerges that aligns strictly with political systems. The central cluster itself is diverse, including Micro-states

(VGB, CYM, BMU), jurisdictions classified as Closed Autocracies (HKG, reflecting its unique status), and Electoral Autocracies (SGP). Liberal Democracies such as the USA and TWN (Taiwan) are present but connect to nodes of various different regime types. This visual evidence supports the notion that regime type, while potentially a factor in individual elite choices (Chang et al., 2023c), is not a primary driver of these strong, systemic co-service relationships at the country-network level. Economic roles, historical ties, and financial infrastructure likely play more dominant roles in shaping this backbone.

The Louvain community detection method, which is data-driven, reveals distinct groupings based on the density of co-service links:

- **Community 0 (Dark Blue):** This is the largest community, featuring prominent offshore centers like VGB and CYM, major Asian economies/financial hubs like SGP, TWN (Taiwan), MYS (Malaysia), IDN (Indonesia), and Pacific jurisdictions like COK (Cook Islands) and likely WSM (Samoa, if present in the filtered graph). This highlights strong ties between several offshore financial centers and key Asian economies.
- **Community 1 (Brown):** This community comprises major economies like the USA and CHN (China), alongside HKG (Hong Kong) and the offshore jurisdiction BMU (Bermuda), indicating a distinct Atlantic-Pacific nexus involving Bermuda.
- **Community 2 (Light Blue):** A smaller, distinct community consisting of PAN (Panama), SYC (Seychelles), and BHS (Bahamas), all of which are significant offshore jurisdictions.

Most nodes in this backbone network are connected within two to three steps, indicating a relatively compact structure despite the filtering.

Centrality in the Country Network Centrality metrics calculated on the full 121-node co-service network (detailed in Appendix Tables A.1 and A.2) identify key players. **VGB (British Virgin Islands)** is dominant, exhibiting the highest betweenness and eigenvector centrality, underscoring its pivotal role in connecting diverse client countries through shared intermediaries. The **USA** ranks second in both measures, reflecting its economic importance and the global reach of its client base serviced by international intermediaries. The USA is linked to BMU (Bermuda) in the filtered graph's Community 1. **HKG (Hong Kong) & CHN (China)** also feature prominently in centrality scores and are central to Community 1. Numerous **Micro-states** (BMU, BHS, CYM) show high centrality, consistent with their specialized roles in offshore finance. **SGP (Singapore)** is another key, highly central node, bridging various parts of the network. In general, high centrality in the full network translates to a significant structural role in this filtered backbone, indicating that the most connected countries in the overall system also form the core of the strongest co-service relationships.

Significant Country Associations Lift scores from the association analysis (top associations detailed in Appendix Table A.3, filtered for co-occurrences ≥ 20) reveal particularly strong and statistically significant pairings, many of which are visualized in Figure 4.9 (those with lift ≥ 1.5). Key findings include:

- Strong **Micro-state synergies** are evident. For instance, the WSM-CYM (Samoa-Cayman Islands) pairing shows a high lift of 6.78, and VGB-CYM (British Virgin Islands-Cayman Islands) has a lift of 1.91. These indicate that intermediaries servicing clients in one of these micro-states are substantially more likely to also service clients in the other, suggesting complementary service offerings or established pathways for specific client types. The CYM-BMU (Cayman Islands-Bermuda) link is exceptionally strong with a lift of 13.5.
- A critical **China-Bermuda nexus** emerges with CHN-BMU showing a very high lift of 15.3. This suggests Bermuda acts as a particularly favored intermediary hub for clients linked to China. This is complemented by the USA-BMU link (lift 4.92), highlighting Bermuda's role in Community 1 of the filtered network, connecting major economic powers.
- Robust **Asian connections** are underscored by pairs like SGP-MYS (Singapore-Malaysia, lift 5.27). Singapore (SGP) also shows strong co-service patterns with various Micro-states such as WSM (Samoa, lift 3.04) and CYM (Cayman Islands, lift 3.89), reinforcing its role as a key hub in Community 0.
- A distinct **PAN-SYC-BHS nexus** (Community 2) is confirmed with pairings like PAN-SYC (Panama-Seychelles) having a lift of 3.89.
- Crucially, high lift values are common **across different regime types**. For example, China (Closed Autocracy) has a very high lift with Bermuda (Micro-state), and the USA (Liberal Democracy) also has a significant lift with Bermuda. This reinforces the earlier observation that factors beyond regime similarity, such as specialized financial services, established legal and commercial pathways, or historical ties, are potent drivers of these strong co-service relationships.

Network of Jurisdictions Used by Intermediaries

Shifting focus from client locations to incorporation locations, this section analyzes the network of jurisdictions that intermediaries use in combination. The full jurisdiction co-usage network, where an edge exists if an intermediary incorporates entities in both jurisdictions (weighted by the number of such intermediaries), comprises 41 nodes and 347 edges. Summary statistics are provided in Table 4.2. The distribution of the number of distinct jurisdictions used per intermediary is shown in Figure 4.10, indicating that most intermediaries utilize a small portfolio of jurisdictions, though some use many.

Table 4.2: Summary Statistics for the Full Jurisdiction Co-Usage Network

Metric	**Value**
Number of Nodes	41
Number of Edges	347
Network Density	0.4232
Average Degree	16.93
Average Clustering Coefficient	0.8155

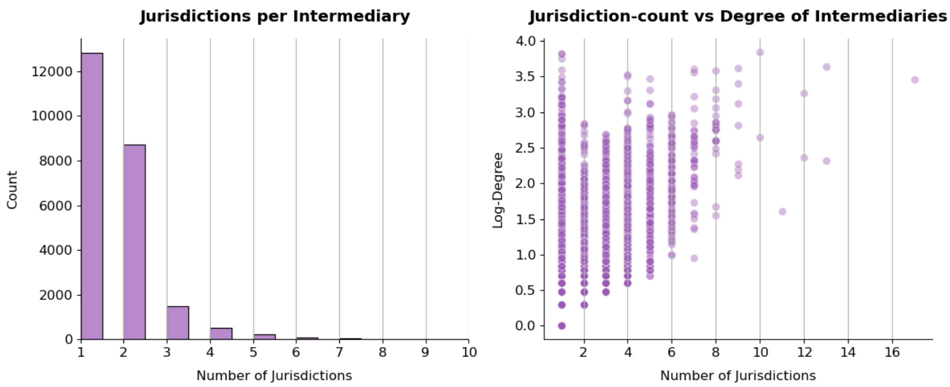

Figure 4.10: Distribution of the Number of Distinct Jurisdictions Used per Intermediary. Most intermediaries use only one or two jurisdictions for incorporation, but a tail of intermediaries uses a broader portfolio.

Figure 4.11 presents a filtered "backbone" of these co-usage patterns, applying the same support (≥ 0.008) and lift (≥ 1.5) thresholds as for the country co-service network. Nodes are coloured by their predominant legal technology profile (derived from Laffitte, 2024, as detailed in Section 3.3) and by Louvain communities. The image displays the most prominent nodes in this filtered network, including CRI (Costa Rica), SGP (Singapore), CYP (Cyprus), GBR (Great Britain), BLZ (Belize), AGO (Angola), HKG (Hong Kong), CYM (Cayman Islands), COK (Cook Islands), MYS (Malaysia), BHS (Bahamas), SYC (Seychelles), PAN (Panama), NIU (Niue), WSM (Samoa), and USA.

Interpretation of the Filtered Jurisdiction Network Structure The filtered jurisdiction network (Figure 4.11) reveals a central, densely connected core. Key jurisdictions in this core include BHS (Bahamas), SYC (Seychelles), AGO (Angola), WSM (Samoa), NIU (Niue), PAN (Panama), USA, and HKG (Hong Kong).

When coloured by their predominant legal technology profile (Laffitte, 2024), the central cluster is overwhelmingly dominated by jurisdictions offering **"Dual-Purpose"** legal technologies (e.g., International Business Companies - IBCs). This strongly supports

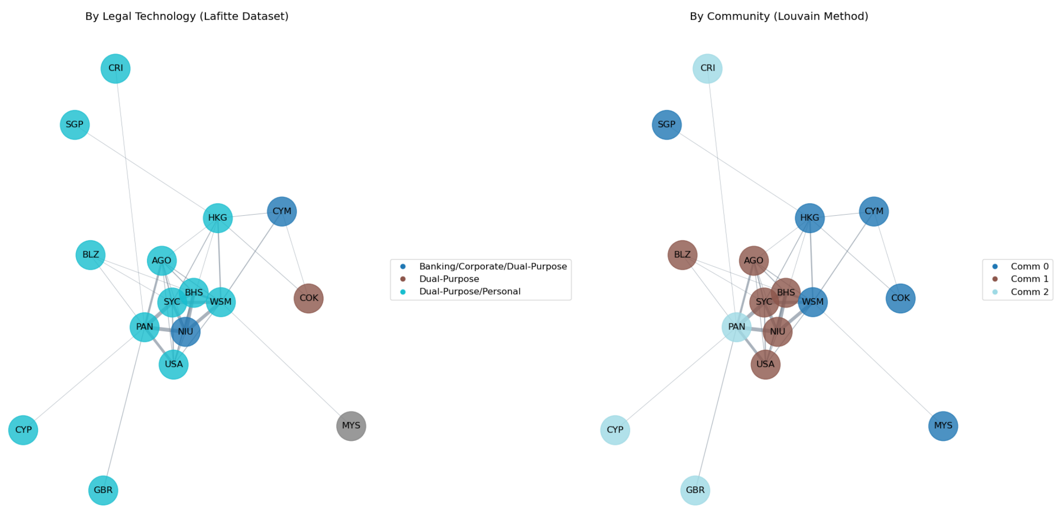

Figure 4.11: Filtered Network of Co-Used Jurisdictions, Coloured by Predominant Legal Technology (left legend) and Louvain Community (right legend). Edges shown have support ≥ 0.008 and lift ≥ 1.5.

the observation that central jurisdictions in this co-usage network are those providing flexible, widely applicable corporate vehicles suitable for both corporate and personal wealth structuring.

Louvain community detection identifies the following groupings based on co-usage patterns:

- **Community 1 (Brown):** This is the largest and most central community, encompassing jurisdictions like USA, PAN, NIU, BHS, SYC, AGO, and WSM. These are largely characterized by "Dual-Purpose" legal technologies.
- **Community 0 (Dark Blue):** This community includes HKG, CYM (Cayman Islands), and COK (Cook Islands), combining jurisdictions known for "Banking/Corporate/Dual-Purpose" (like HKG) with those strong in "Dual-Purpose/Personal" (like CYM, COK).
- **Community 2 (Light Blue):** This community is more peripheral in the filtered network and includes SGP (Singapore), CRI (Costa Rica), CYP (Cyprus), GBR (Great Britain), BLZ (Belize), and MYS (Malaysia), representing a mix of financial centers and specialized offshore jurisdictions.

Centrality in the Jurisdiction Network Centrality metrics for the full 41-jurisdiction co-usage network (detailed in Appendix Tables A.4 and A.5) are revealing. **VGB (British Virgin Islands)** ranks first in both betweenness and eigenvector centrality in the full network, confirming its paramount importance as an incorporation jurisdiction. However, it is strikingly absent from the filtered graph in Figure 4.11. This implies that while VGB is co-used with many other jurisdictions by numerous intermediaries, these individual pairings might not meet the specific high support and lift thresholds chosen for this backbone view (requiring at least 20 co-occurrences and lift ≥ 1.5). This suggests VGB's role might be more as a general-purpose, widely connected jurisdiction whose strong pairings are numerous but perhaps more diffuse, rather than concentrated in extremely high-lift niche combinations that also meet the co-occurrence threshold. **BHS (Bahamas)** and **PAN (Panama)** rank second and third, respectively, in overall centrality and are visibly central within the filtered graph, particularly in Community 1. **HKG (Hong Kong)** and **CYM (Cayman Islands)** are also highly central in the full network and form a core part of Community 0 in the filtered view. Most other top-ranked jurisdictions by centrality align with their prominence in the backbone, with VGB being the main exception due to the filtering criteria.

Significant Jurisdiction Associations Association analysis, focusing on lift scores from statistically significant pairs with at least 20 co-occurrences (detailed in Appendix Table A.6), highlights robust co-usage patterns:

- The dominant **Community 1 (largely "Dual-Purpose" hubs)** shows very high mutual lift values. For instance, BHS-NIU (Bahamas-Niue) has a lift of 4.6 (support 0.016), and NIU-WSM (Niue-Samoa) has a lift of 5.3 (support 0.012). NIU (Niue) appears as a critical connector within this cluster of Pacific and Caribbean jurisdictions, also showing strong lift with SYC (Seychelles, lift 4.1, support 0.010).
- **Community 0 (financial centers and specialized OFCs)**: The HKG-CYM (Hong Kong-Cayman Islands) pairing shows a strong lift of 5.8 (support 0.0018), indicating a significant tendency for intermediaries using one to also use the other. WSM-CYM (Samoa-Cayman Islands) also shows a notable lift of 4.6 (support 0.0031).
- Strong co-usage is observed between jurisdictions offering **similar legal technology profiles**. For example, many of the high-lift pairs within Community 1 involve jurisdictions predominantly offering "Dual-Purpose" or "Dual-Purpose/Personal" technologies.

4.3 Functional Specialisation of Intermediaries

This section transitions from the geographical patterns of intermediary activity to an exploration of their functional roles. Drawing upon the typology developed by the EU (2017) (De Groen, 2017), which distinguishes between roles such as Tax Expert, Legal Expert, Administrator, and Investment Advisor, we investigate whether these classifications correspond to distinct operational characteristics. A central theme is the differentiation between intermediaries primarily offering *personalised advice* versus those focused on *aid in incorporation* and ongoing entity management.

The analysis in this section primarily utilizes a classified random sample of intermediaries. This approach is adopted because, as will be shown, intermediaries with the highest degrees (i.e., those connected to the largest number of entities) are not representative of the broader intermediary population in terms of functional type (see Figure 4.14). The detailed filtering process for this enriched random sample, ensuring high-confidence classifications, is illustrated in Figure 4.12 (and further detailed in Appendix **??** if you have such an appendix section).

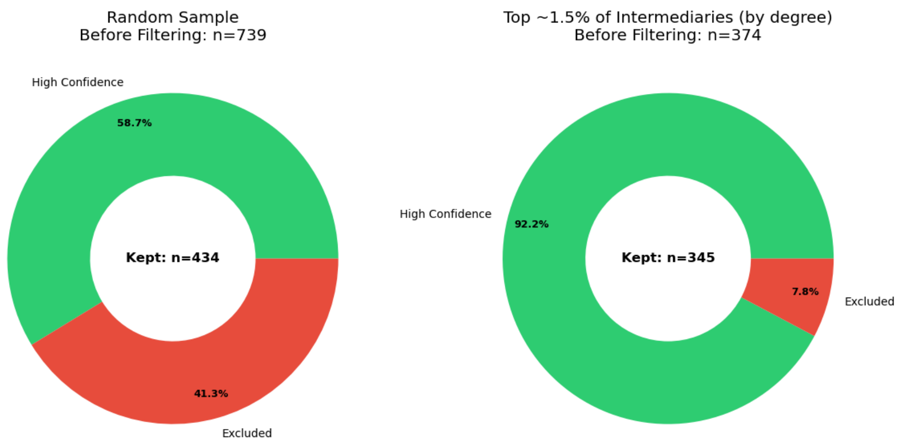

Figure 4.12: Filtering Process of the Enriched Random Sample for Functional Classification. This flowchart illustrates the steps taken to arrive at the final set of intermediaries with high-confidence functional classifications used in this section's analysis.

Different Levels of Connectivity: Personalised Advice vs. Aid in Incorporation

A key differentiator among intermediary types is expected to be their scale of operation, proxied by their degree (the number of entities they are connected to). Intermediaries providing bespoke, personalised advice (e.g., Tax Experts, Investment Advisors) might typically serve fewer clients than those offering more standardized services like entity incorporation and administration (e.g., Administrators, and some Legal Experts specializing in volume).

Figure 4.13 presents the Cumulative Distribution Function (CDF) of degrees for each intermediary classification within the random sample. Visual inspection suggests differences in these distributions, particularly between groups like Tax Experts and Administrators.

To formally test these observations, a Kruskal-Wallis H test was performed, which

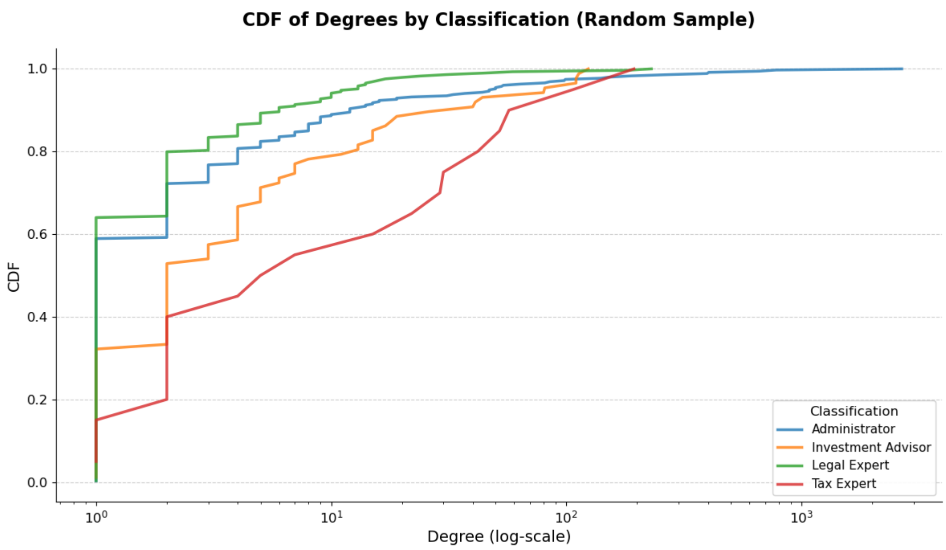

Figure 4.13: Cumulative Distribution Function (CDF) of Degrees by Intermediary Classification (Random Sample). The plot shows distinct degree profiles, with Tax Experts and Investment Advisors generally having lower degrees than Administrators and Legal Experts.

indicated a statistically significant difference in degree distributions among the four classifications (H(3) = 51.243, $p < 0.0001$). Subsequent pairwise two-sample Kolmogorov-Smirnov (KS) tests, with Bonferroni correction for the six unique pairs (corrected $\alpha = 0.05/6 \approx 0.0083$), were conducted to pinpoint specific differences:

- **Administrator vs. Investment Advisor:** Significant difference (KS = 0.267, $p_{corr} \approx 0.0024$).
- **Administrator vs. Legal Expert:** No significant difference (KS = 0.077, $p_{corr} = 1.0000$).
- **Administrator vs. Tax Expert:** Significant difference (KS = 0.439, $p_{corr} \approx 0.0282$).
- **Investment Advisor vs. Legal Expert:** Significant difference (KS = 0.318, $p_{corr} < 0.0001$).
- **Investment Advisor vs. Tax Expert:** No significant difference (KS = 0.285, $p_{corr} = 1.0000$).
- **Legal Expert vs. Tax Expert:** Significant difference (KS = 0.490, $p_{corr} \approx 0.0042$).

These results support the notion of two broader functional groups based on connectivity. The "personalised advice" types (Tax Experts and Investment Advisors) do not significantly differ from each other in degree distribution. Similarly, the "aid in incorporation/management" types (Legal Experts and Administrators) do not significantly differ from each other. However, comparisons across these broader groups (e.g., Tax Expert vs. Legal Expert, Investment Advisor vs. Administrator) reveal significant differences, with the personalised advice types generally having lower degrees.

This distinction is further highlighted when comparing the composition of the random sample with a sample of the top approximately 1.5% of intermediaries by degree (Figure 4.14). The top-degree sample is overwhelmingly composed of Administrators, who constitute a much smaller fraction of the random sample. Conversely, the random sample shows a higher prevalence of Legal Experts, Investment Advisors, and Tax Experts, who are less dominant among the highest-degree intermediaries. This underscores that intermediaries specializing in high-volume entity administration are distinct from those providing more individualized services.

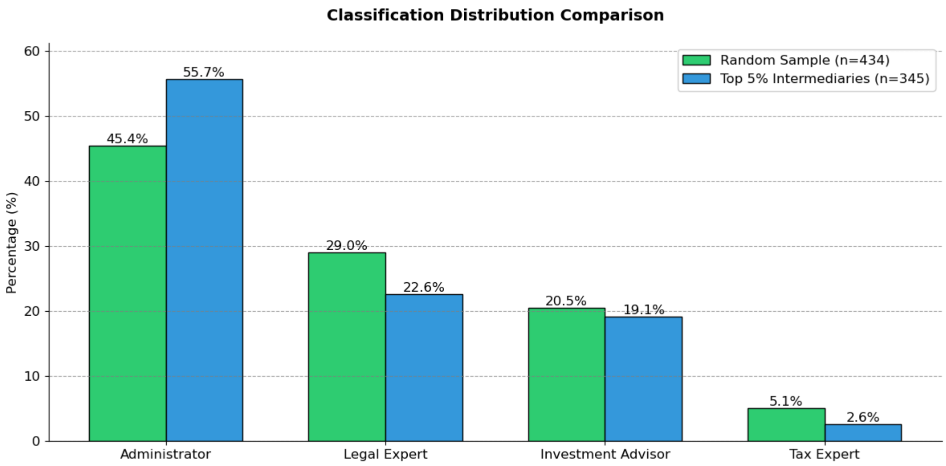

Figure 4.14: Distribution of Intermediary Classifications: Random Sample vs. Top ≈1.5% by Degree. The random sample shows a more diverse mix of functional types, while the top-degree sample is heavily dominated by Administrators.

Different Activities: Instruments and Service Offerings

Beyond connectivity, we examine if intermediary types differ in their operational activities, using five key metrics:

1. **Jurisdiction Entropy**: Diversity in the jurisdictions where they incorporate entities.

2. **Client Country Entropy**: Diversity in the countries their clients' entities are linked to.

3. **Regime Entropy**: Diversity in the political regimes of the countries where client entities are linked.

4. **Legal Technology Entropy**: Diversity in the types of legal technologies (Laffitte, 2024) prevalent in the jurisdictions they use for incorporation.

5. **Bearer Instrument Usage**: A binary indicator of whether the intermediary has serviced entities using bearer instruments.

Figure 4.15 displays the average values of these metrics for each intermediary classification in the random sample. Pairwise comparisons using Mann-Whitney U tests for entropy measures and Fisher's exact test for bearer instrument usage were conducted, with a Bonferroni correction applied for the 30 comparisons (5 metrics $\times$ 6 pairs, corrected $\alpha = 0.05/30 \approx 0.00167$).

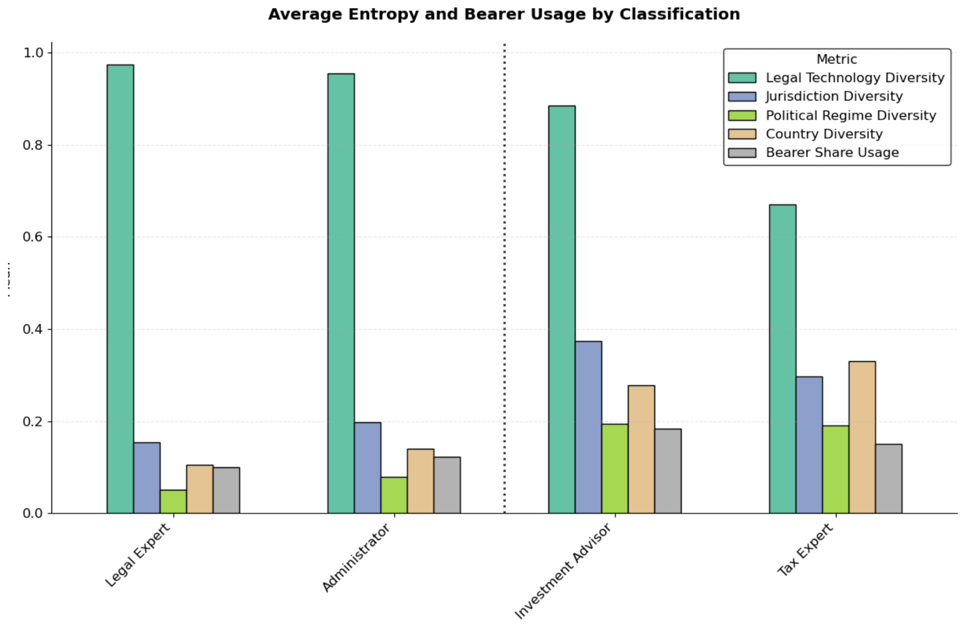

Figure 4.15: Average Entropy Measures and Bearer Instrument Usage by Intermediary Classification (Random Sample). Error bars could represent standard errors if available.

The statistical tests revealed several significant differences (detailed results in Appendix **??** if you have one, otherwise summarize key findings):

Legal Technology Entropy: Legal Experts exhibit significantly higher diversity in the legal technologies of the jurisdictions they use compared to Investment Advisors ($U = 15330.5, p_{corr} < 0.0001$) and Tax Experts ($U = 4560.5, p_{corr} < 0.0001$). Administrators also show significantly higher diversity than Investment Advisors ($U = 17801.5, p_{corr} \approx 0.0047$) and Tax Experts ($U = 5401.5, p_{corr} < 0.0001$). This suggests that Legal Experts and Administrators engage with a broader array of jurisdictional legal frameworks. No significant difference was found between Legal Experts and Administrators, nor between Investment Advisors and Tax Experts.

Jurisdiction Entropy: Legal Experts show significantly higher diversity in incorporation jurisdictions compared to Investment Advisors ($U = 9271.0, p_{corr} < 0.0001$).

Similarly, Administrators demonstrate greater jurisdiction diversity than Investment Advisors ($U = 12187.5, p_{corr} \approx 0.0012$). Other pairwise comparisons for jurisdiction entropy did not yield statistically significant differences after Bonferroni correction. This indicates that Legal Experts and Administrators tend to utilize a wider range of jurisdictions than Investment Advisors.

Regime Entropy: Legal Experts exhibit significantly higher diversity in the political regimes of their client countries compared to Investment Advisors ($U = 10185.0, p_{corr} < 0.0001$) and Tax Experts ($U = 2100.5, p_{corr} \approx 0.0001$). Administrators also show significantly higher regime diversity than Investment Advisors ($U = 13170.5, p_{corr} \approx 0.0044$). This pattern suggests that Legal Experts and Administrators may cater to clienteles originating from a more diverse set of political environments.

Client Country Entropy: Legal Experts demonstrate significantly higher diversity in their client countries compared to Investment Advisors ($U = 10164.5, p_{corr} \approx 0.0001$) and Tax Experts ($U = 1863.5, p_{corr} < 0.0001$). Administrators also have significantly more diverse client countries than Tax Experts ($U = 2462.0, p_{corr} \approx 0.0057$). This implies that Legal Experts, and to some extent Administrators, engage with clients from a broader range of countries.

Bearer Instrument Usage: After Bonferroni correction, **no statistically significant differences** were found in the propensity to use bearer instruments among any of the intermediary classifications. This suggests that, within this classified random sample, the use of these high-anonymity instruments is not strongly associated with a particular functional type of intermediary.

In summary, Legal Experts and Administrators tend to exhibit greater diversity across various geographical and legal-technical dimensions of their service offerings compared to Tax Experts and Investment Advisors. This aligns with the notion that the former group, often involved in broader incorporation and management services, may engage with a wider array of jurisdictional options and client origins. The "personalised advice" groups (Tax Experts and Investment Advisors) appear more focused in these respects. The lack of differentiation in bearer instrument usage is a notable finding, suggesting other factors may drive their deployment.

Discussion

5.1 Three Propositions

Three core propositions developed

Not in themselves particularly novel - none of them come as a particular surprise - but viewed through a set of new empirical data.

5.1.1 Proposition 1: Specialisation of Intermediaries

Functional Specialisation: Even though there's bound to be measurement error with the approach taken here, yet still significance for at least two distinct sets of roles. Taking outset from the typology of De Groen (2017).

First set of roles:

- Administrators
- Legal Experts

Second set:

- Tax Experts
- Investment Advisors
- Both the significance of the degree distribution only between those two sets,
- likewise entropy measures and bearer instrument share clearly visually distinct and significant across those two groups, but not very much within

Correlate with operational scale (degree of connectivity) and the diversity of their service portfolios (jurisdictional and legal-tech entropy). Personalised wealth management versus the mass provision of legal services.

Geographical Specialisation: Distinct development of preferential client corridors linking specific client origination countries to favored offshore jurisdictions, sometimes driven by historical ties, linguistic affinity, or specialized demand. As we'll see in proposition 3, still universal hubs that connect all these corridors.

- Country heatmaps,
- Distribution of degree and country connections

5.1.2 Proposition 2: Duality of Intermediary Focus - Local Anchors, Global Reach

Intermediaries often exhibit a primary client concentration within their own operational countries, likely leveraging local networks and trust (Granovetter, 1973; Stausholm, 2024; Harrington, 2016; Hoang, 2022). However, their core value proposition and a key driver of their specialization lies in their capacity to connect these local clients to a diversified global offshore architecture.

- Heatmaps showing intermediaries in HKG, GBR, USA etc., serving a notable portion of clients from their home country.
- Lower client country entropy compared to jurisdiction entropy, suggesting more concentration in client origin.

5.1.3 Proposition 3: Structural Centrality of Microstates in Intermediation Network

A core set of Offshore Financial Centers (OFCs), that in heavy part are those infamous microstates. Highly central in it.

Interesting to note, that this may also be due to their offering of versatile "legal technologies" like "Dual-Purpose" vehicles as for example is proposed in Laffitte (2024). Empirical confirmation that they form the structural backbone of the global offshore network, and are countries that intermediaries from all countries whose residents they do business and jurisdiction they incorporate. Critical hubs and bridges in chains of intermediation, facilitating complex offshore strategies regardless of client or intermediary home country.

- High centrality (betweenness, eigenvector) of jurisdictions like VGB, BHS, PAN, CYM, HKG in your co-service and co-usage networks.
- Dominance of "Dual-Purpose" legal technologies in the central core of the jurisdiction co-usage network.
- High lift values between key OFCs in association analysis.

5.2 Implications for Regulation

The potential to escape the multi-level games; most intermediaries serving officers in their own countries. Reclaiming lost tax revenue, can be done by targeting local citizens (though still the problem of changing citizenship).

Functional specialisation very much seeming to correlate with the degree: Layered due diligence or liability regimes could be interesting.

Affirming the centrality of microstates in the offshore network, and the potential for targeted regulation. Switzerland and Luxembourg as less active tax havens as case study.

And not least, affirming the importance of intermediaries.

5.3 Limitations

Where to begin. PLACEHOLDER

5.4 Future Research

- A lot within the current dataset that could be done - like extending a lot of the analyses. More importantly, the whole temporal dimension is currently left out; are the patterns persistent? As seen, network here has entities over a very large time frame, countless of them that are not active anymore.

Conclusion

Appendix

A.1 Country Network Centrality and Associations

Table A.1 lists the top 10 countries by betweenness centrality and Table A.2 by eigenvector centrality in the full co-service network. Table A.3 details significant country associations.

Table A.1: Top 10 Countries by Betweenness Centrality in the Full Co-Service Network (excluding XXX)

Node	Betweenness	Eigenvalue	Appearances	Regime
VGB	0.18	0.14	6285	Micro-state
USA	0.053	0.13	1042	Liberal Democracy
CHE	0.039	0.13	1545	Liberal Democracy
GBR	0.028	0.13	1258	Liberal Democracy
MUS	0.024	0.13	139	Liberal Democracy
BHS	0.021	0.13	489	Micro-state
BMU	0.020	0.13	103	Micro-state
PAN	0.020	0.096	1203	Electoral Democracy
SGP	0.019	0.13	578	Electoral Autocracy
URY	0.017	0.031	318	Liberal Democracy

Table A.2: Top 10 Countries by Eigenvector Centrality in the Full Co-Service Network (excluding XXX)

Node	Eigenvalue	Betweenness	Appearances	Regime
VGB	0.14	0.18	6285	Micro-state
USA	0.13	0.053	1042	Liberal Democracy
GBR	0.13	0.028	1258	Liberal Democracy
HKG	0.13	0.016	2865	Closed Autocracy
JEY	0.13	0.013	390	Micro-state
CHN	0.13	0.0085	320	Closed Autocracy
CAN	0.13	0.0088	195	Liberal Democracy
BHS	0.13	0.021	489	Micro-state
SGP	0.13	0.019	578	Electoral Autocracy
CYM	0.13	0.012	363	Micro-state

Table A.3: Significant Country Associations in Co-Service Network (Bonferroni Corrected $p < 6.89 \times 10^{-6}$)

	u	v	u_regime	v_regime	support	lift	p_value
76	VGB	WSM	Micro-state	Micro-state	0.017	1.87	7.24e-46
498	WSM	CYM	Micro-state	Micro-state	0.0035	6.78	1.34e-45
105	VGB	CYM	Micro-state	Micro-state	0.0076	1.91	9.58e-23
775	CHN	BMU	Closed Autocracy	Micro-state	0.00088	15.3	3.36e-19
2405	CYM	BMU	Micro-state	Micro-state	0.00088	13.5	4.46e-18
2032	SGP	MYS	Electoral Autocracy	Electoral Autocracy	0.0016	5.27	5.66e-17
102	VGB	SGP	Micro-state	Electoral Autocracy	0.010	1.60	9.75e-17
351	PAN	SYC	Electoral Democracy	Electoral Autocracy	0.0020	3.89	8.08e-16
501	WSM	COK	Micro-state	Micro-state	0.0015	4.84	1.57e-15
496	WSM	SGP	Micro-state	Electoral Autocracy	0.0025	3.04	1.89e-14
2044	SGP	BMU	Electoral Autocracy	Micro-state	0.00083	8.06	5.07e-13
488	WSM	SYC	Micro-state	Electoral Autocracy	0.0014	4.14	2.36e-12
2033	SGP	CYM	Electoral Autocracy	Micro-state	0.0014	3.89	1.75e-11
2035	SGP	COK	Electoral Autocracy	Micro-state	0.0010	4.63	2.23e-10
1190	USA	BMU	Liberal Democracy	Micro-state	0.00092	4.92	4.62e-10
497	WSM	TWN	Micro-state	Liberal Democracy	0.00083	5.48	5.27e-10
768	CHN	CYM	Closed Autocracy	Micro-state	0.00096	4.75	8.31e-10
2034	SGP	MUS	Electoral Autocracy	Liberal Democracy	0.00071	5.07	4.46e-08
419	JEY	BMU	Micro-state	Micro-state	0.00050	7.16	1.17e-07
642	HKG	CYM	Closed Autocracy	Micro-state	0.0032	1.75	6.61e-07
650	HKG	BMU	Closed Autocracy	Micro-state	0.0013	2.52	6.83e-07
502	WSM	MYS	Micro-state	Electoral Autocracy	0.0012	2.75	1.54e-06
2039	SGP	TWN	Electoral Autocracy	Liberal Democracy	0.00054	5.04	1.85e-06
2307	AUS	IRL	Liberal Democracy	Liberal Democracy	0.00021	22.0	3.25e-06

A.2 Jurisdiction Network Centrality and Associations

Table A.4 shows the top 10 jurisdictions by betweenness centrality and Table A.5 by eigenvector centrality in the full co-usage network. Significant jurisdiction associations are detailed in Table A.6.

Table A.4: Top 10 Jurisdictions by Betweenness Centrality in the Full Co-Usage Network (excluding XXX)

Node	Betweenness	Eigenvalue	Appearances	Jurisdiction Legal Technology
VGB	0.20	0.26	13533	Dual-Purpose/Personal
BHS	0.084	0.26	2099	Banking/Corporate/Dual-Purpose/Other Technologies/Personal
PAN	0.060	0.25	6533	Banking/Corporate/Dual-Purpose
HKG	0.058	0.24	625	Banking/Corporate/Other Technologies
CYM	0.048	0.21	290	Banking/Corporate/Dual-Purpose
WSM	0.027	0.20	1352	Dual-Purpose/Personal
USA	0.019	0.23	387	None
COK	0.018	0.12	954	Banking/Corporate/Dual-Purpose/Personal
CYP	0.017	0.22	45	Banking/Corporate/Dual-Purpose
SGP	0.013	0.19	355	Banking/Other Technologies

Table A.5: Top 10 Jurisdictions by Eigenvector Centrality in the Full Co-Usage Network (excluding XXX)

Node	Eigenvalue	Betweenness	Appearances	Jurisdiction Legal Technology
VGB	0.26	0.20	13533	Dual-Purpose/Personal
BHS	0.26	0.084	2099	Banking/Corporate/Dual-Purpose/Other Technologies/Personal
PAN	0.25	0.060	6533	Banking/Corporate/Dual-Purpose
HKG	0.24	0.058	625	Banking/Corporate/Other Technologies
USA	0.23	0.019	387	None
CYP	0.22	0.017	45	Banking/Corporate/Dual-Purpose
CYM	0.21	0.048	290	Banking/Corporate/Dual-Purpose
WSM	0.20	0.027	1352	Dual-Purpose/Personal
JEY	0.20	0.011	28	Dual-Purpose/Other Technologies
SGP	0.19	0.013	355	Banking/Other Technologies

Table A.6: Significant Jurisdiction Associations in Co-Usage Network (Bonferroni Corrected $p < 6.10 \times 10^{-5}$)

	u	v	u_legal_technology	v_legal_technology	support	lift	p_value
72	BHS	NIU	Bnk/Corp/Dual/Oth Tech/Pers	Dual-Purpose	0.016	4.6	1.80e-165
108	NIU	WSM	Dual-Purpose	Dual-Purpose/Personal	0.012	5.3	9.68e-134
106	NIU	SYC	Dual-Purpose	Dual-Purpose/Personal	0.010	4.1	1.89e-85
122	SYC	WSM	Dual-Purpose/Personal	Dual-Purpose/Personal	0.011	3.3	4.50e-73
3	PAN	SYC	Bnk/Corp/Dual-Purpose	Dual-Purpose/Personal	0.027	1.6	2.50e-49
73	BHS	SYC	Bnk/Corp/Dual/Oth Tech/Pers	Dual-Purpose/Personal	0.013	2.3	2.62e-48
123	SYC	AGO	Dual-Purpose/Personal	None	0.0043	4.6	2.20e-40
4	PAN	USA	Bnk/Corp/Dual-Purpose	None	0.0095	2.2	1.33e-39
121	SYC	USA	Dual-Purpose/Personal	None	0.0041	4.2	9.61e-36
143	USA	AGO	None	None	0.0021	8.6	1.49e-32
2	PAN	NIU	Bnk/Corp/Dual-Purpose	Dual-Purpose	0.018	1.6	1.88e-32
174	WSM	CYM	Dual-Purpose/Personal	Bnk/Corp/Dual-Purpose	0.0031	4.6	1.31e-29
1	PAN	BHS	Bnk/Corp/Dual-Purpose	Bnk/Corp/Dual/Oth Tech/Pers	0.033	1.4	3.87e-29
74	BHS	USA	Bnk/Corp/Dual/Oth Tech/Pers	None	0.0042	3.0	1.33e-23
162	WSM	HKG	Dual-Purpose/Personal	Bnk/Corp/Oth Tech	0.0043	2.9	4.59e-23
204	HKG	CYM	Bnk/Corp/Oth Tech	Bnk/Corp/Dual-Purpose	0.0018	5.8	3.16e-21
107	NIU	USA	Dual-Purpose	None	0.0025	3.9	1.05e-19
6	PAN	AGO	Bnk/Corp/Dual-Purpose	None	0.0074	1.8	3.80e-18
163	WSM	AGO	Dual-Purpose/Personal	None	0.0027	3.2	1.30e-16
140	USA	WSM	None	Dual-Purpose/Personal	0.0026	2.8	8.95e-14
7	PAN	GBR	Bnk/Corp/Dual-Purpose	None	0.0023	2.4	1.50e-13
76	BHS	AGO	Bnk/Corp/Dual/Oth Tech/Pers	None	0.0032	2.4	3.06e-13
128	SYC	BLZ	Dual-Purpose/Personal	Bnk/Corp/Dual-Purpose/Pers	0.00092	6.0	3.06e-12
75	BHS	WSM	Bnk/Corp/Dual/Oth Tech/Pers	Dual-Purpose/Personal	0.0080	1.6	3.94e-12
10	PAN	CRI	Bnk/Corp/Dual-Purpose	None	0.0011	3.1	7.44e-11
109	NIU	AGO	Dual-Purpose	None	0.0017	2.8	2.41e-09
81	BHS	BLZ	Bnk/Corp/Dual/Oth Tech/Pers	Bnk/Corp/Dual-Purpose/Pers	0.00083	3.7	1.18e-07
34	VGB	NIU	Dual-Purpose/Personal	Dual-Purpose	0.025	1.1	4.69e-07
9	PAN	CYP	Bnk/Corp/Dual-Purpose	Bnk/Corp/Dual-Purpose	0.0012	2.3	9.42e-07
207	HKG	SGP	Bnk/Corp/Oth Tech	Bnk/Oth Tech	0.0011	2.8	2.53e-06
125	SYC	HKG	Dual-Purpose/Personal	Bnk/Corp/Oth Tech	0.0028	1.8	3.98e-06

A.3 Classification of Intermediaries

To instruct the AI agent on how to perform the classification and the specific structure of the information to return, the following prompt template is utilized. This prompt defines the categories, provides keywords for guidance, and specifies the desired output fields. The agent's output for each intermediary is a structured data record, typically resembling a JSON object or a Python dictionary, which includes the fields detailed in the prompt.

Classification Prompt

The core prompt provided to the AI agent for classification is as follows (where `{intermediary_name}` and `{log_summary_for_classification}` are dynamically inserted):

```
Classify the intermediary: {intermediary_name}

Based *only* on the information gathered in the following search log.
{log_summary_for_classification}

Classify this intermediary into ONE of these categories based on their
likely primary role in offshore activities:
- Tax Expert: Focuses on tax planning, compliance, advisory. Keywords:
  tax advisory, international tax, tax compliance, tax returns,
  transfer pricing, VAT, tax structuring.
- Legal Expert: Focuses on legal structuring, compliance, incorporation,
  representation. Keywords: legal services, corporate law, entity formation,
  incorporation, contracts, litigation, legal opinions, regulatory
  compliance, M&A legal, lawyer, attorney, solicitor.
- Administrator: Focuses on accounting, auditing, financial reporting,
  company administration. Keywords: accounting, bookkeeping, audit,
  financial statements, reporting, company secretarial, payroll,
  administration services, domiciliation, accountant, auditor.
- Investment Advisor: Focuses on managing financial assets and investments.
  Keywords: investment management, wealth management, asset management,
  portfolio management, financial planning, investment strategy,
  securities, funds, financial advisor.

Provide a structured classification including:
- classification (Enum: Tax Expert, Legal Expert, Administrator, Investment Advisor)
- role_muddled (bool: true if the role seems mixed or unclear)
- role_muddled_reasoning (str: explanation if role_muddled is true)
```

```
- is_individual (bool: based on the name and findings, is this likely a person?)
- job_title (str: inferred job title if possible, e.g., "Lawyer", "Accountant",
  "Director", or "Unknown")
- confidence (Enum: Low, High - Use Low if evidence is sparse, contradictory,
  or confidence in the source/relevance is low)
- justification (str: detailed reasoning for the classification, referencing the
  search log)
- key_evidence (list[str]: specific snippets or findings from the search
  results supporting the classification)

Analyze the content of the search results carefully. Prioritize information
directly describing the intermediary's services or professional role.
```

Examples of Dynamic Search and Structured Output

The agent's search process is dynamic. It begins with a general query (the intermediary's name) and, based on the retrieved information's relevance and completeness, may formulate up to two additional, more specific queries. For instance, if initial results for a company are vague, subsequent queries might include terms like "services offered" or "business activity." The classification is then made based on the entirety of the gathered search logs.

Sometimes it's just not possible to find anything useful, hence the confidence section. Any cases where 'confidence' is low, they are excluded from the analysis sections.

The output for each intermediary is a structured record. While the 'key_evidence' field in the prompt requests specific snippets, for comprehensiveness in these examples, it contains the full, somewhat verbose, search log detailing each iteration of the dynamic search process.

The following examples illustrate this process and the resulting structured output, taken directly from the enrichment process. The first provides an example of where it's impossible to find clear information linked to the interemediary resulting in low confidence, while the second is a clear case of a legal expert.

Example 1: OPULENT OPAL HOLDINGS LIMITED

Dynamic Search Summary:

- **Iteration 0 Query:** 'OPULENT OPAL HOLDINGS LIMITED'
- **Findings:** Results were ambiguous, pointing to "OPULENT OPAL CONSTRUCTION LIMITED," a subsidiary link to "Opal Holdings Pte Ltd" (itself a subsidiary

of CapitaLand, suggesting a holding structure), and unrelated products (paint, home decor).

- **Iteration 1 Query:** ‘OPULENT OPAL HOLDINGS LIMITED primary business activities‘
- **Findings:** Retrieved general information on Bank Secrecy Act compliance for businesses, potentially triggered by "holdings," and another mention of the CapitaLand subsidiary link.
- **Iteration 2 Query:** ‘OPULENT OPAL HOLDINGS LIMITED services offered‘
- **Findings:** Largely irrelevant results (gemstones, furniture), again highlighting the difficulty in pinpointing specific services for this entity.

Structured Output (Simplified Representation):

- **intermediary_name:** OPULENT OPAL HOLDINGS LIMITED
- **classification:** Administrator
- **confidence:** Low
- **role_muddled:** True
- **role_muddled_reasoning:** "The role is muddled as the search results are ambiguous and could indicate construction, holding company activities, or other possibilities. There is no clear indication of their primary business activities."
- **is_individual:** False
- **job_title:** Unknown
- **justification:** "The search results provide limited information about OPULENT OPAL HOLDINGS LIMITED. One result mentions 'OPULENT OPAL CONSTRUCTION LIMITED' which suggests involvement in construction, potentially as a specialty trade contractor. Another result indicates that 'Opal Holdings Pte Ltd' is a subsidiary of CapitaLand, implying a holding company structure. The lack of direct evidence about their services or role leads to a low confidence classification as an administrator, potentially involved in company administration or holding activities."
- **key_evidence (Full Search Log):** (Content of the JSON-like string representing the three search iterations and their results, as provided in the original example, would be detailed here. For brevity, it is summarized above and omitted in this representation.)

Example 2: Cheo Yeoh & Associates

Dynamic Search Summary:

- **Iteration 0 Query:** ‘Cheo Yeoh & Associates‘
- **Findings:** Immediately yielded relevant results identifying the entity as "CHEO YEOH & ASSOCIATES LLC," a Singapore-based company providing legal services, with details on its registration and directors.
- **Iteration 1 Query:** ‘Cheo Yeoh & Associates services‘
- **Findings:** Reinforced the legal nature, with results like "Cheo Yeoh & Associates provides legal services" and official listings specifying "LEGAL ACTIVITIES."
- **Iteration 2 Query:** ‘Cheo Yeoh & Associates client testimonials‘ (or similar refinement if needed)
- **Findings:** Could uncover legal case mentions, lawyer profiles, or further business directory listings, solidifying its role. (The provided log indicates successful identification primarily from the first two iterations).

Structured Output (Simplified Representation):

- **intermediary_name:** Cheo Yeoh & Associates
- **classification:** Legal Expert
- **confidence:** High
- **role_muddled:** False
- **role_muddled_reasoning:** (Empty or N/A)
- **is_individual:** False
- **job_title:** Lawyer (Implicitly, as a law firm)
- **justification:** "Cheo Yeoh & Associates LLC provides legal services in Singapore. The company’s principal activity is LEGAL ACTIVITIES (EXCLUDING ONLINE MARKETPLACES). Johnny Cheo is a Director at the firm. The firm has been operating for 23 years. The search results consistently point to legal services."
- **key_evidence (Full Search Log):** (Content of the JSON-like string representing the search iterations and their results, as provided in the original example, would be detailed here. For brevity, it is summarized above and omitted in this representation.)