

GNN1-2025 - Assignment 1

In Assignment 1, we received a dataset from the ICIJ Offshore Leaks, which contains information on more than two million documents—covering companies, directors (officers), and addresses. I have decided to focus on the most connected address, that is, the address with the highest number of relationships—and analyze its role and significance within the dataset.

- **Finding the top 10 nodes with most relations**

To begin, I examine the top 10 nodes with the most relationships to see if any address exhibits an unusually high number of connections. The Cypher query to find nodes with the most relationships:

```
MATCH (n)-[r]-()
WITH n, COUNT(r) AS connections
RETURN n, connections
ORDER BY connections DESC
LIMIT 10
```

Here we see the results for the top 10 nodes with the most relationships. The node with the highest number of relationships is an address: "Portcullis TrustNet Chambers P.O. Box 3444 Road Town," located in the British Virgin Islands. It has 37,338 connections, which is an unusually high number of connections for an address suggests that it might be involved in atypical or even suspicious behavior. This anomaly will be analyzed further.

Node Description	connections
(:address {sourceID: "Offshore Leaks", address: "Portcullis TrustNet Chambers P.O. Box 3444 Road Town, Tortola British Virgin Islands (w.e.f 9 December 2005)", country_codes: "VGB", node_id: "236724"})	37,338
(:intermediary {sourceID: "Offshore Leaks", name: "Portcullis TrustNet (BVI) Limited", country_codes: "THA;VGB;IDN;SGP", node_id: "54662"})	36,374
(:intermediary {sourceID: "Paradise Papers - Nevis corporate registry", name: "Morning Star Holdings Limited", country_codes: "KNA", node_id: "230000018"})	35,359
(:intermediary {sourceID: "Bahamas Leaks", name: "MOSSACK FONSECA & CO. (BAHAMAS) LIMITED", country_codes: "BHS", node_id: "23000136"})	14,902
(:address {sourceID: "Pandora Papers - Alemán, Cordero, Galindo & Lee (Alcogal)", address: "3rd Floor, Yamraj Building, Market Square, P.O. Box 3175 Road Town, Tortola British Virgin Islands", country_codes: "VGB", node_id: "240000001"})	11,962
(:intermediary {sourceID: "Bahamas Leaks", name: "UBS TRUSTEES (BAHAMAS) LTD.", country_codes: "BHS", node_id: "23000147"})	9,719
(:address {sourceID: "Paradise Papers - Appleby", address: "Clifton House", country_codes: "CYM", node_id: "81027146"})	9,268
(:address {sourceID: "Paradise Papers - Appleby", address: "Canon's Court", country_codes: "BMU", node_id: "81027090"})	8,558
(:intermediary {sourceID: "Bahamas Leaks", name: "CREDIT SUISSE TRUST LIMITED", country_codes: "BHS", node_id: "23000330"})	8,302
(:intermediary {sourceID: "Bahamas Leaks", name: "TRIDENT CORPORATE SERVICES (BAH) LTD", country_codes: "BHS", node_id: "23000156"})	8,287

- **Identify Key Companies/Intermediaries for chosen address**

From the previous analysis, where I identified an address exhibiting unusual behavior, I will now proceed to pinpoint key companies connected to this address. To do this, I will use a Cypher query that retrieves the companies or intermediaries with the highest total number of relationships, specifically filtering for those that have this address registered as their "registered_address."

```
MATCH (a:address {node_id: "236724"})<-
[:registered_address]-(c)
OPTIONAL MATCH (c)--()
WITH c, count(*) AS TotalConnections
RETURN c.name AS CompanyName, labels(c) AS
Labels, TotalConnections
ORDER BY TotalConnections DESC
LIMIT 5;
```

When examining the relationships associated with the addresses, it becomes clear that the node with the highest connectivity is an intermediary called Portcullis TrustNet (BVI) Limited, which has 36,374 connections. Notably, this intermediary is also registered at the address we previously flagged as suspicious, suggesting that it may play a significant role in the anomalous activity observed.

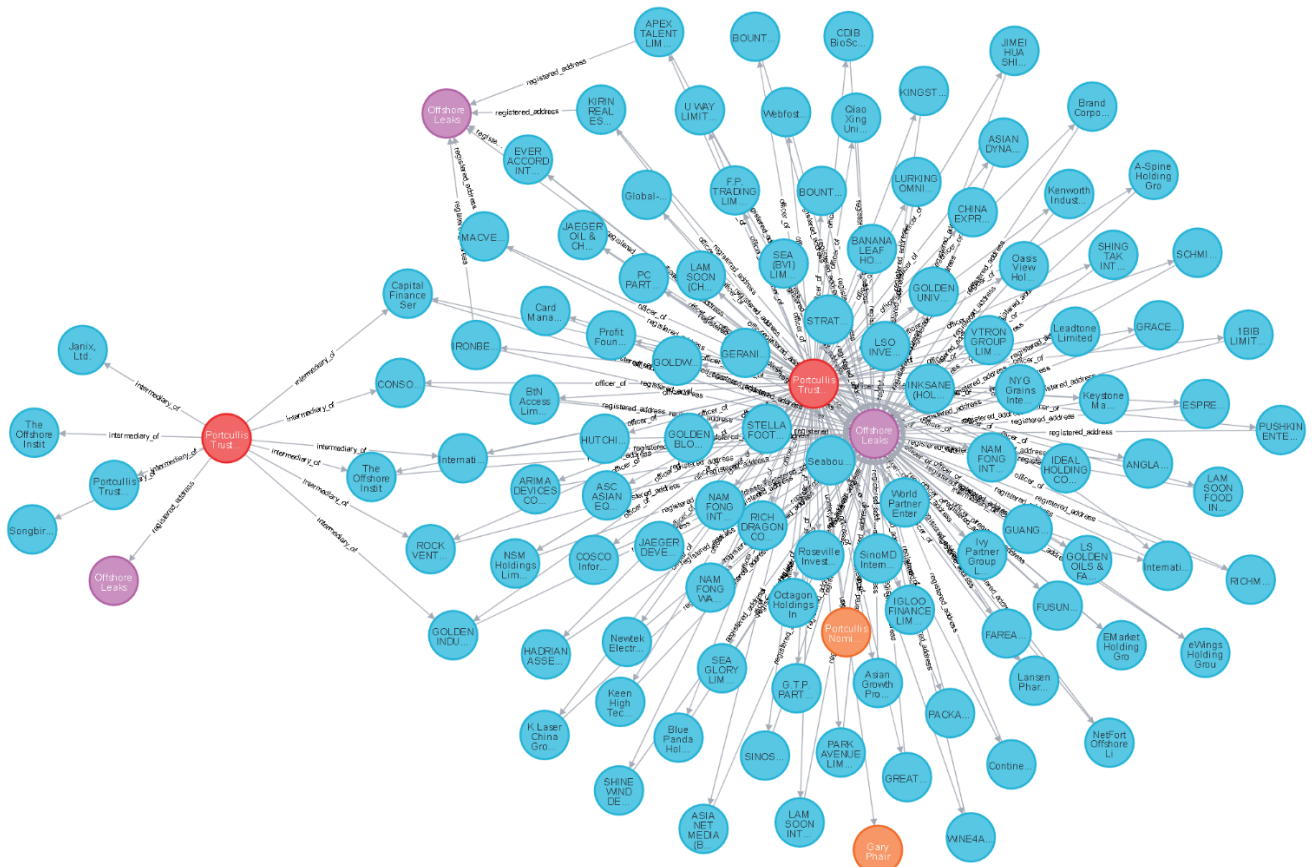
Company Name	Labels	Total Connections
"Portcullis TrustNet (BVI) Limited"	intermediary	36,374
"Sharecorp Limited"	officer	2,695
"Execorp Limited"	officer	2,050
"Acticorp Limited"	officer	1,423
"Sherper Limited"	officer	698

- **Analyzing the intermediary**

The intermediary stands out not only because it is linked to a suspicious address, but also because it has the highest number of overall connections. This makes it a prime candidate for further investigation. I will use the following Cypher query to retrieve the top 100 connected nodes for deeper analysis:

```
MATCH (i:intermediary {name: "Portcullis TrustNet (BVI) Limited"})
MATCH (i)-[r]-(n)
WITH i, n, size([ (x)-[]-(n) | 1 ]) AS totalConnections
ORDER BY totalConnections DESC
WITH i, collect({node: n, connections: totalConnections}) AS neighbors
UNWIND neighbors[0..100] AS topNeighbor
RETURN i, topNeighbor.node AS Neighbor, topNeighbor.connections AS TotalConnections;
```

This query will help us focus on the most significant relationships, narrowing down the network to the top 100 nodes (officers and entities) connected to this intermediary, which can then be visualized and analyzed further.



Below is a partial visualization of the top 100 nodes connected to Portcullis TrustNet (BVI) Limited (in red). This intermediary functions as a major hub, linking numerous officers (in blue) and entities (in orange). Additionally, there is a smaller pink node to the left that represents a separate cluster of relationships. Overall, Portcullis TrustNet (BVI) Limited stands out for its high connectivity and is therefore a prime candidate for further investigation. At the same time, other nodes in the diagram may also play central roles in the broader network and potentially be part of any illicit or suspicious activities.

- **Further analysis**

To refine the search within the top 100 nodes identified earlier, we could apply more advanced analytical tools to prioritize key connections. One approach would be to implement an algorithm like PageRank, which assigns a score to each node based on its relationships. By weighing each connection, we can filter out less relevant or non-suspicious relationships, ensuring that only the most influential nodes are highlighted. Combining this score with the existing relationship counts would provide a clearer indication of a node's true influence within the network, helping us determine where to focus our investigation first.