

---

# Offshore Leaks

*An Exploration in  
Graphs & Networks*

By Azin Faghihi

---

# Problem statement:

How does the network of Offshore Leaks look like?

- *Making a platform to study similar data in future.*

*Data: [Offshore Leaks Database](#)* by

The International Consortium of Investigative Journalists

---

# Motivation

---



An ICIJ Investigation



# PANDORA PAPERS

The largest investigation in journalism history  
exposes a shadow financial system that benefits  
the world's most rich and powerful.

# About

## Learn more about the data in the ICIJ Offshore Leaks database

---

This database contains information on more than 785,000 offshore entities that are part of the Paradise Papers, the Panama Papers, the Offshore Leaks and the Bahamas Leaks investigations. The data links to people and companies in more than 200 countries and territories.

The real value of the database is that it strips away the secrecy that cloaks companies and trusts incorporated in tax havens and exposes the people behind them. This includes, when available, the names of the real owners of those opaque structures. In all, the interactive application reveals more than 720,000 names of people and companies behind secret offshore structures. They come from leaked records and not a standardized corporate registry, so there may be duplicates. In some cases, companies are listed as shareholders for another company or a trust, an arrangement that often helps obscure the flesh-and-blood people behind offshore entities.

The International Consortium of Investigative Journalists obtained the data through four massive leaks:

Source: <https://offshoreleaks.icij.org/pages/about>

---

# General look at the **data**

---

## offshore\_leaks.edges.csv



## offshore\_leaks...des.entity.csv



## offshore\_leaks...des.officer.csv



## offshore\_leaks...s.address.csv



offshore\_leaks...ermediary.csv



## offshore\_leaks.edges.csv



CSV



CSV



## offshore\_leaks...s.address.csv



:offshore\_leaks...ermediary.csv



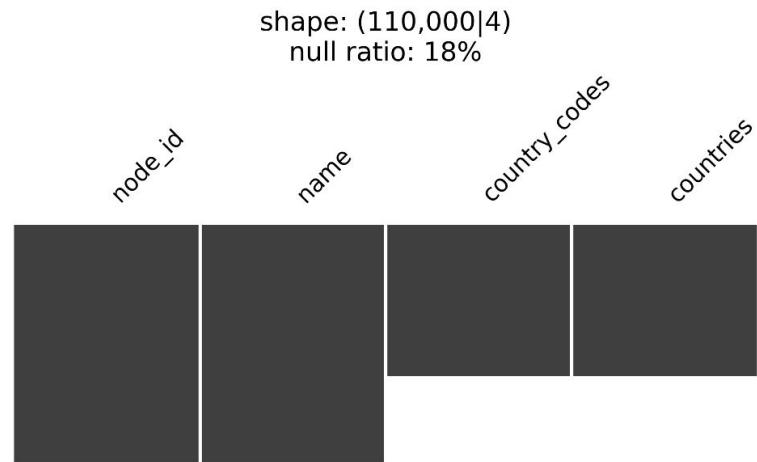
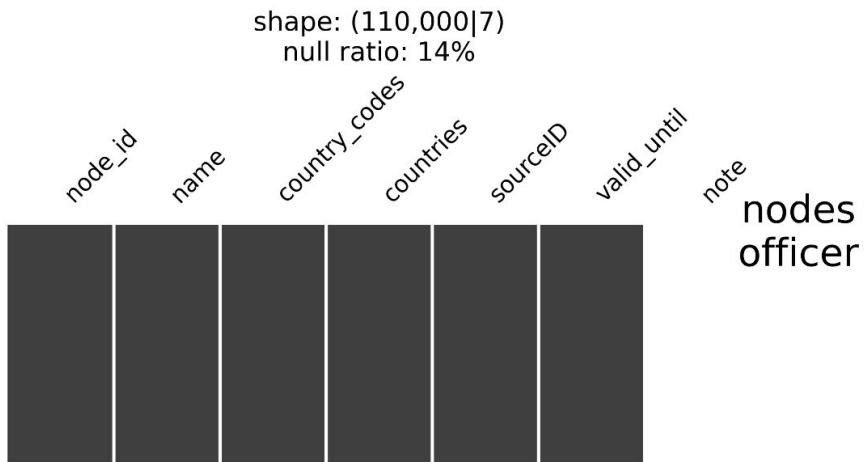
---

Tables:

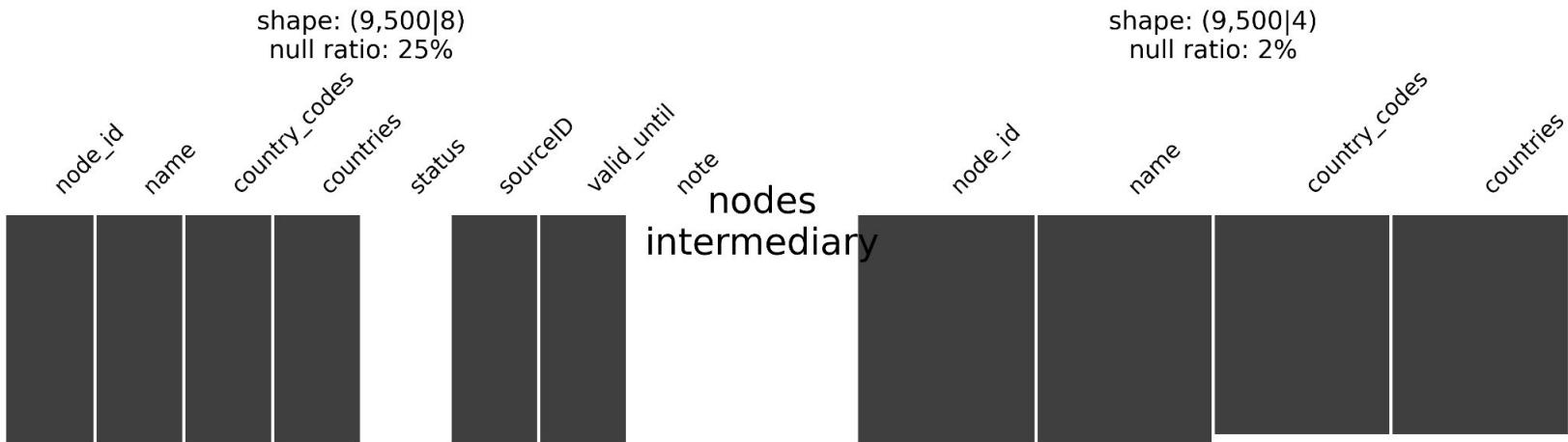
*Nodes & Edges*

---

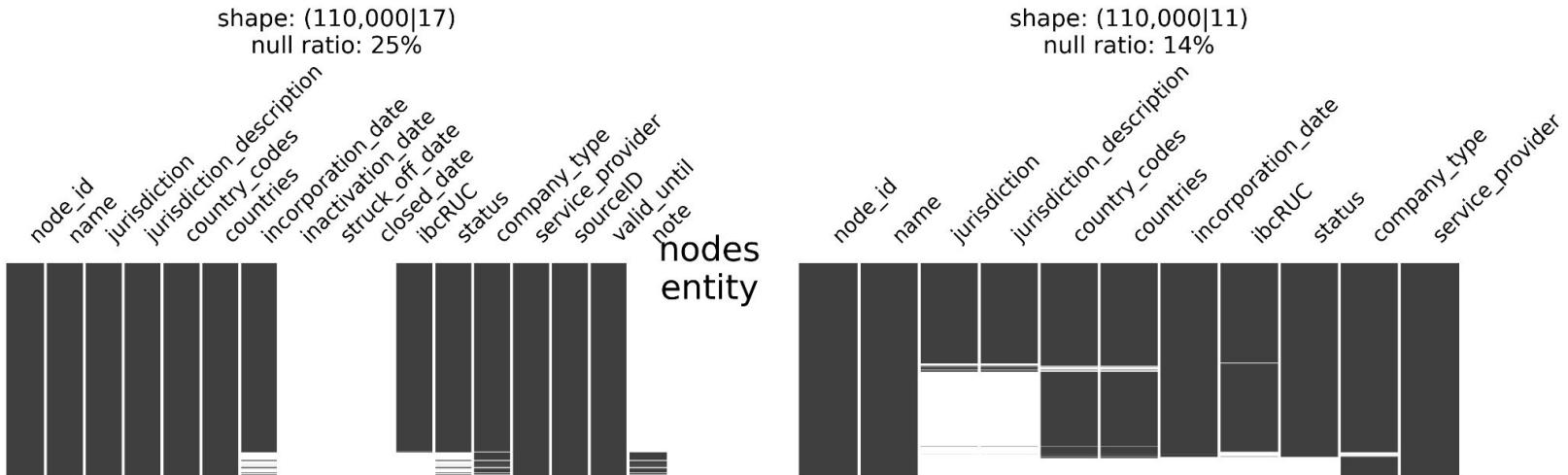
# Nodes



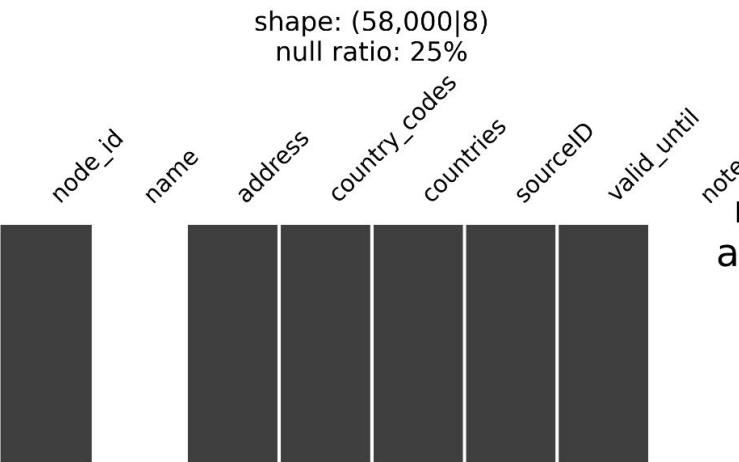
# Nodes



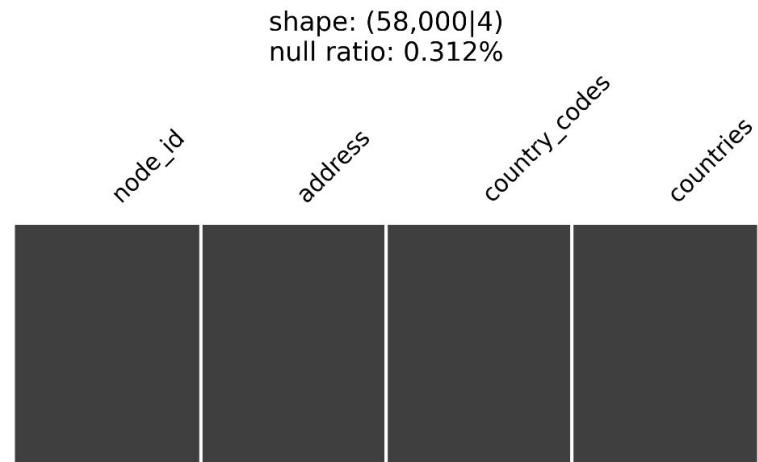
# Nodes



# Edges



nodes  
address



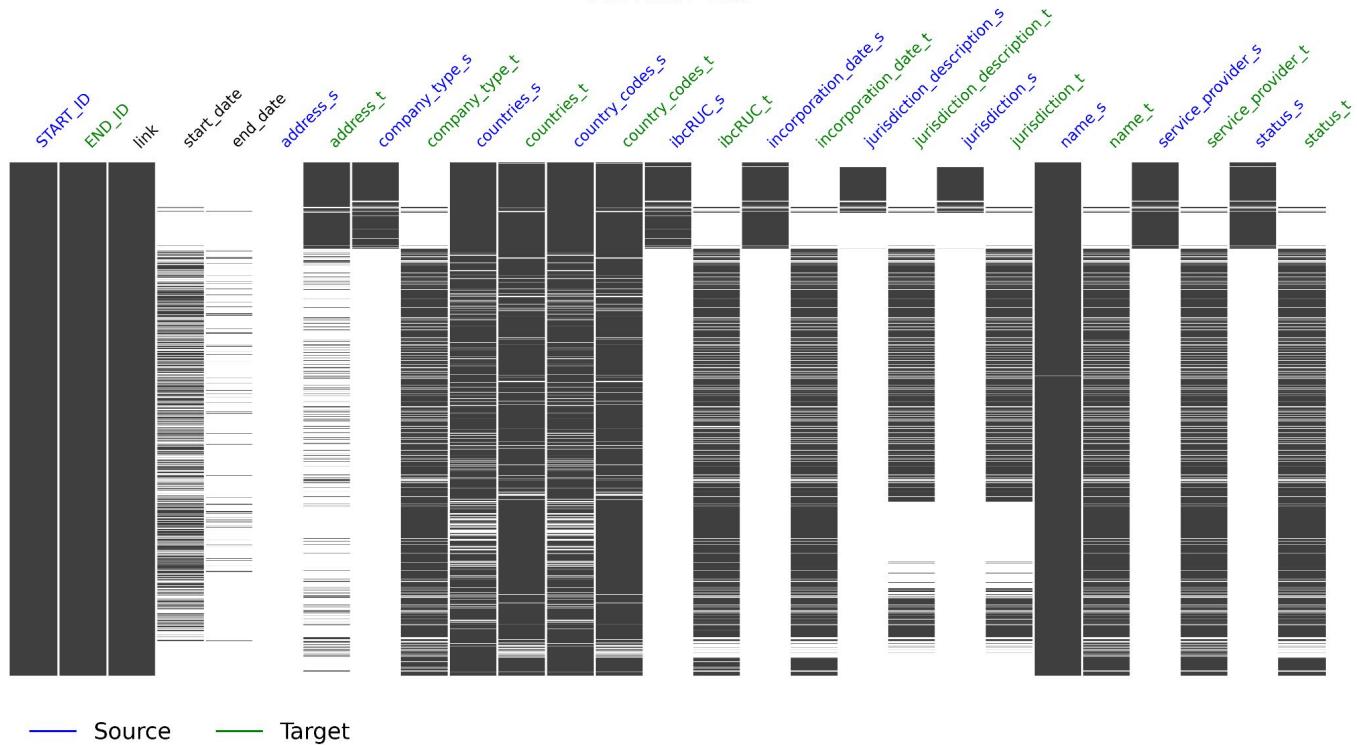
---

# Combine Tables

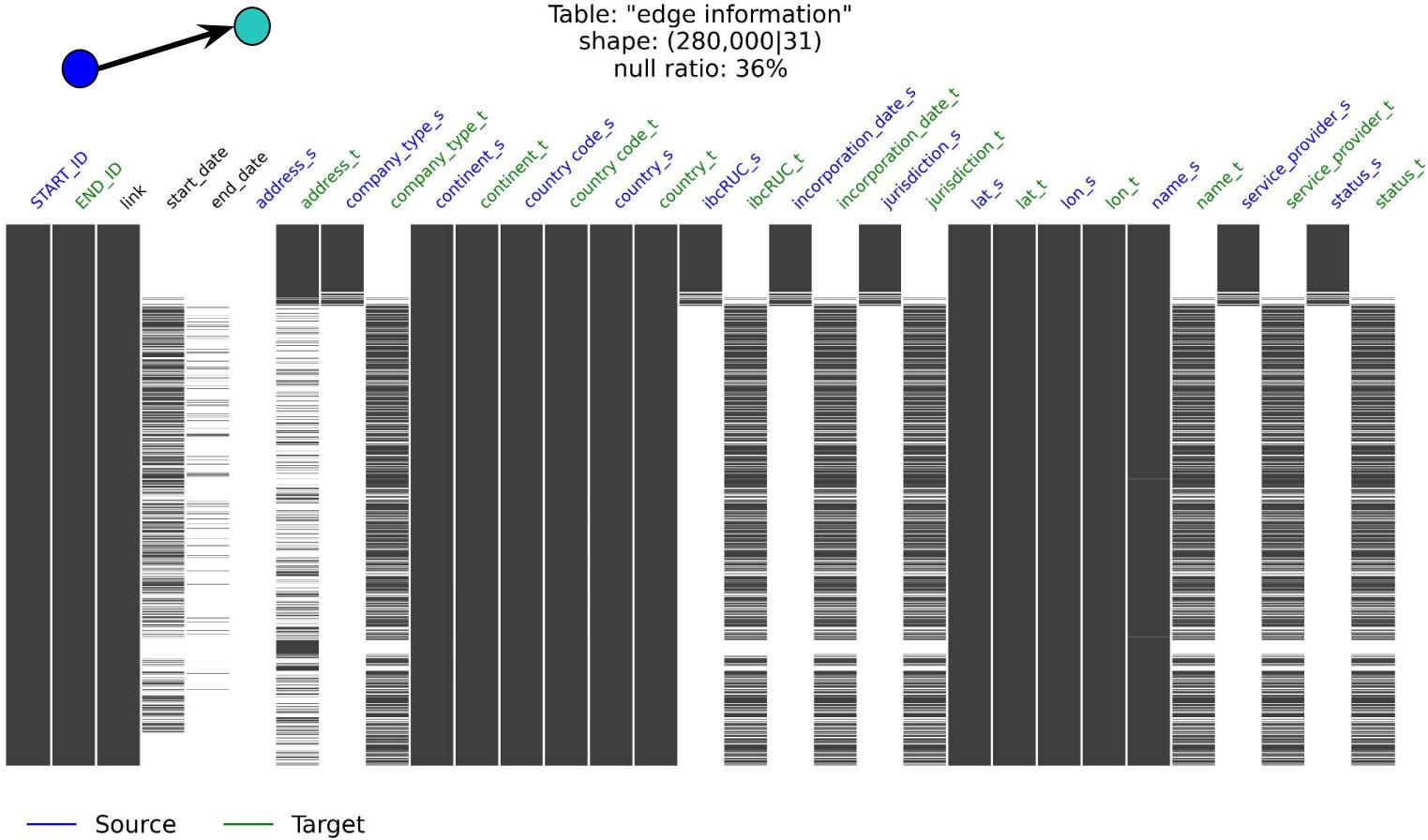
---

# Nodes & Edges

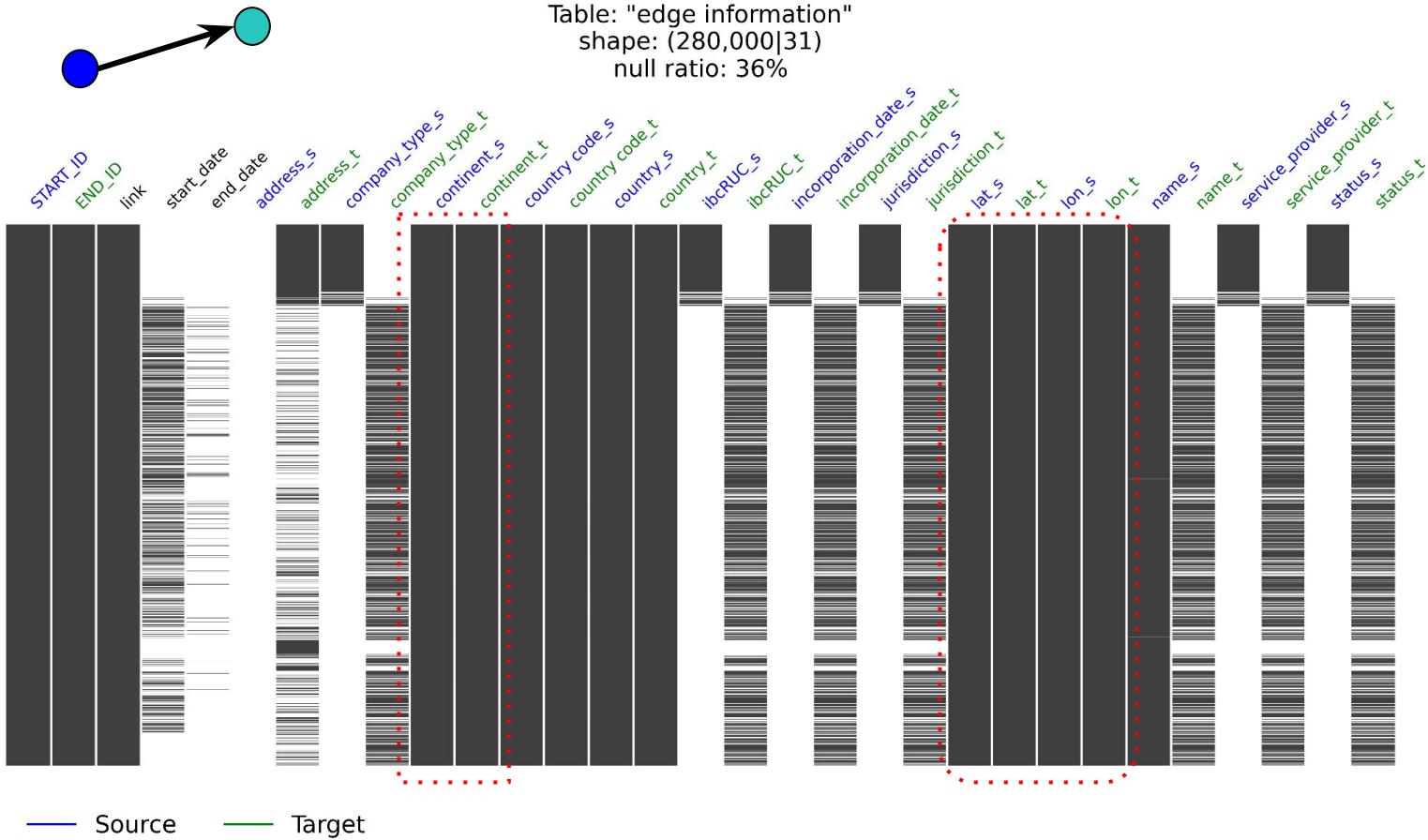
Edges & Nodes  
shape: (560,000|27)  
null ratio: 46%



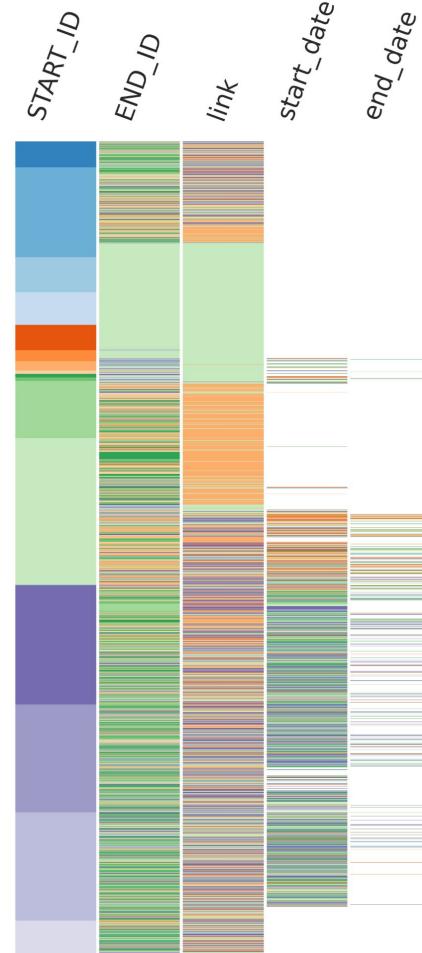
# Add Geographic Attributes



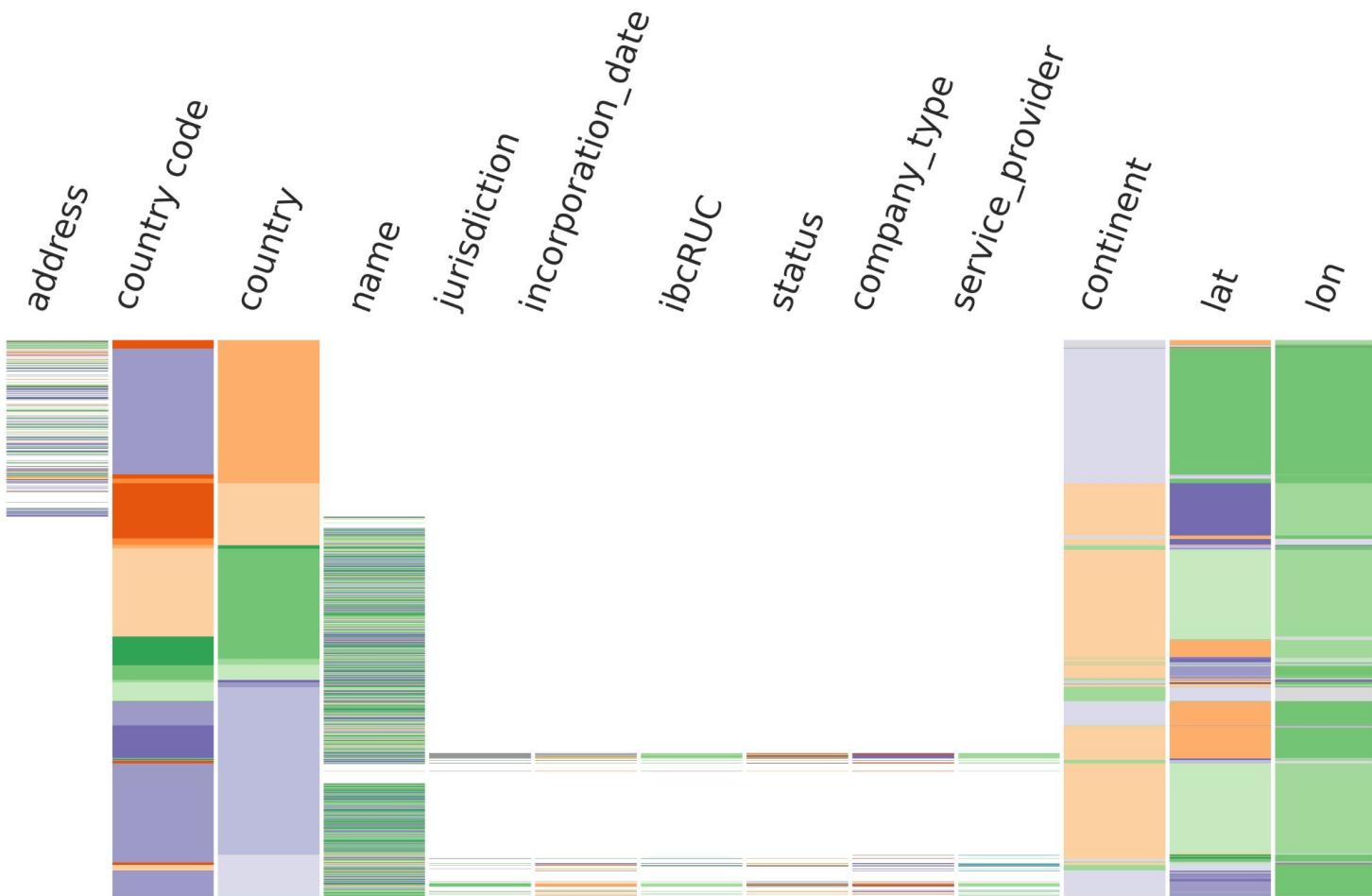
# Add Geographic Attributes



## Edges Table Columns



# Nodes Columns



## Edges Columns



---

# Graph Components:

## *Nodes & Edges*

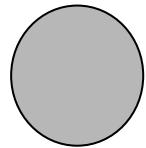
---

# **Node Properties**

status

name

service provider

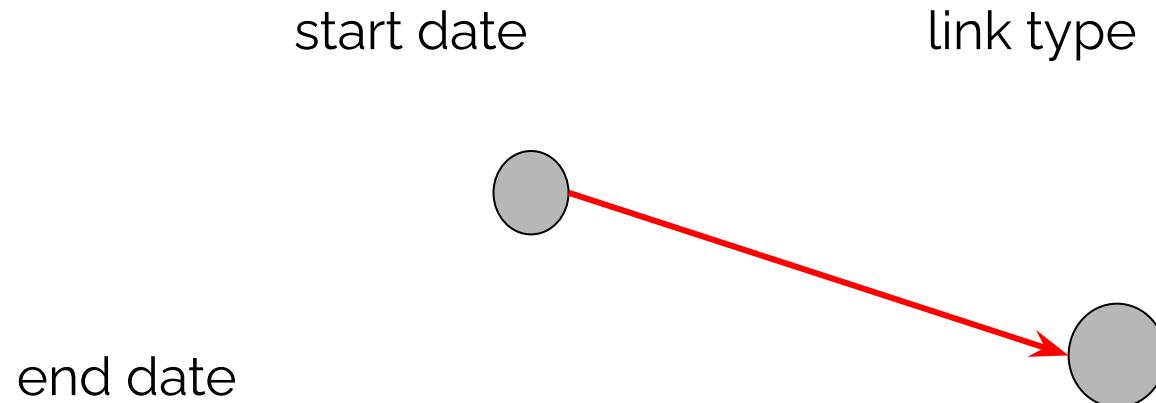


address

company type

status

# **Edge Properties**

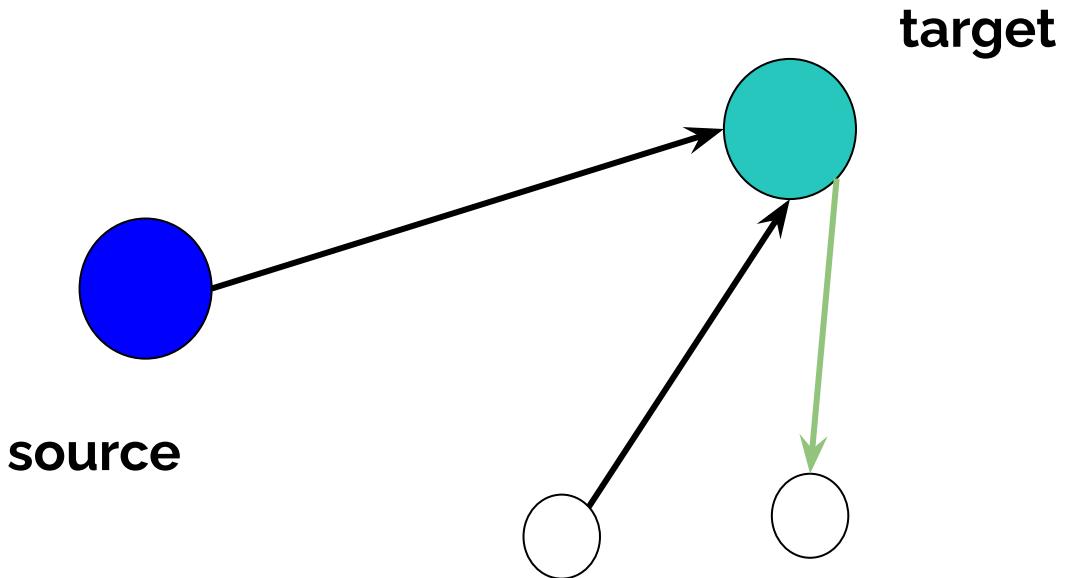


---

# Graph properties: *Degrees*

---

# Degrees in a Directed Graph

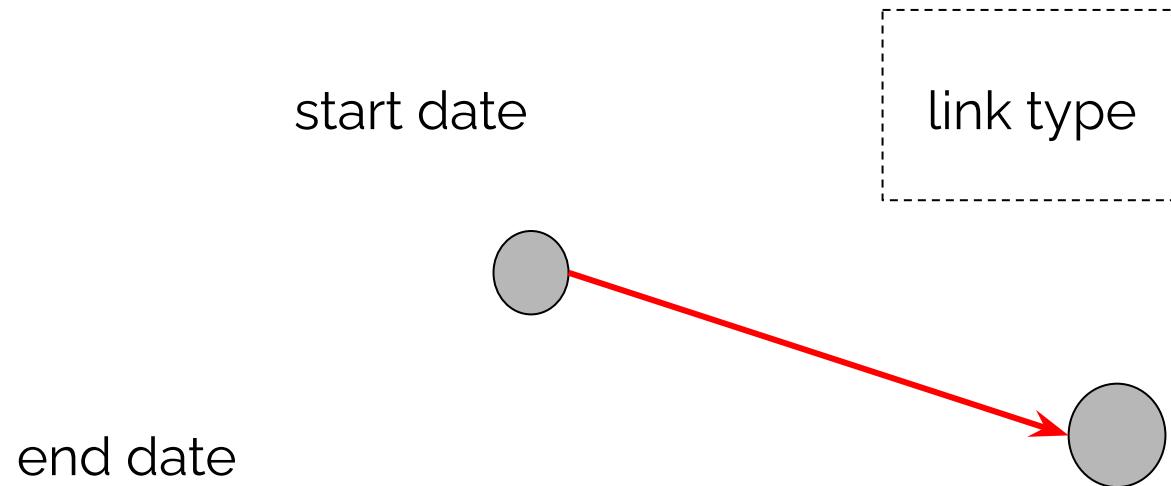


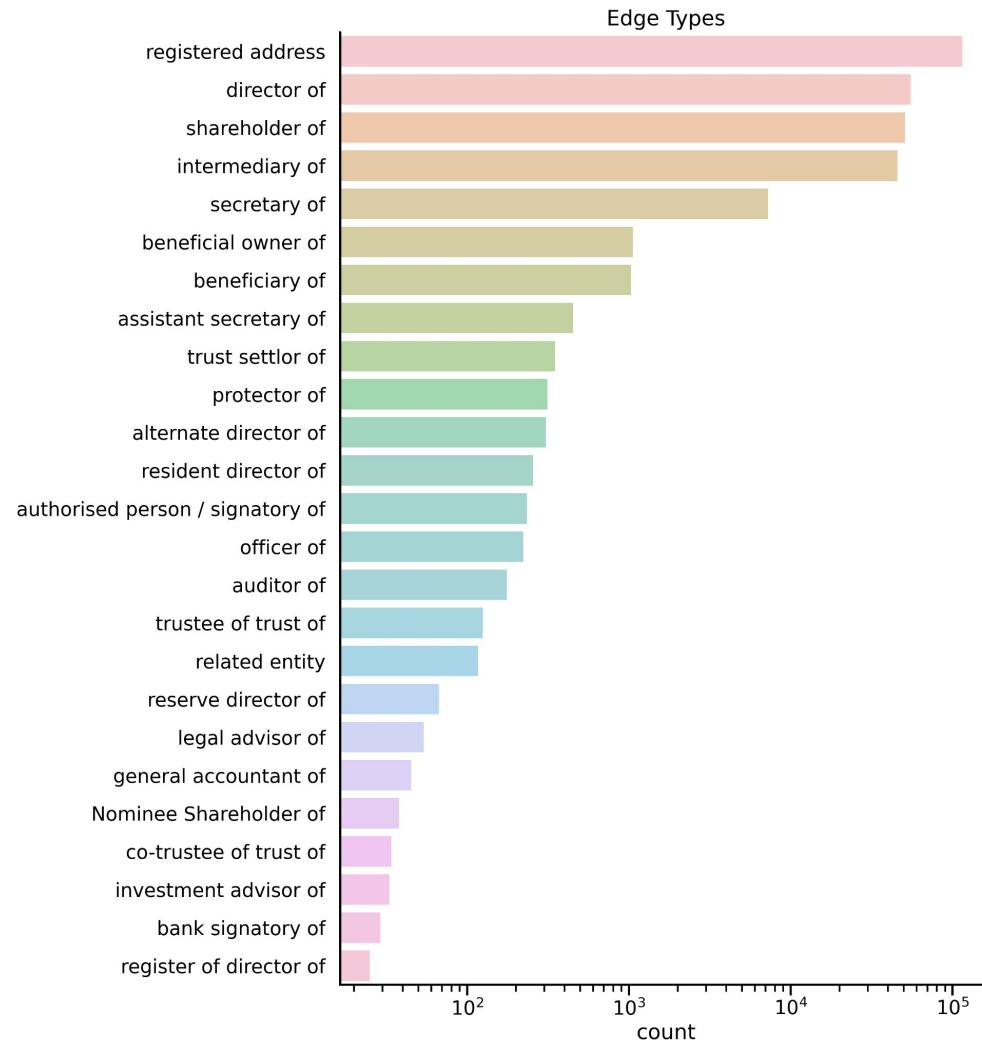
---

# Edge Properties: “*link*” type

---

# Edge Properties



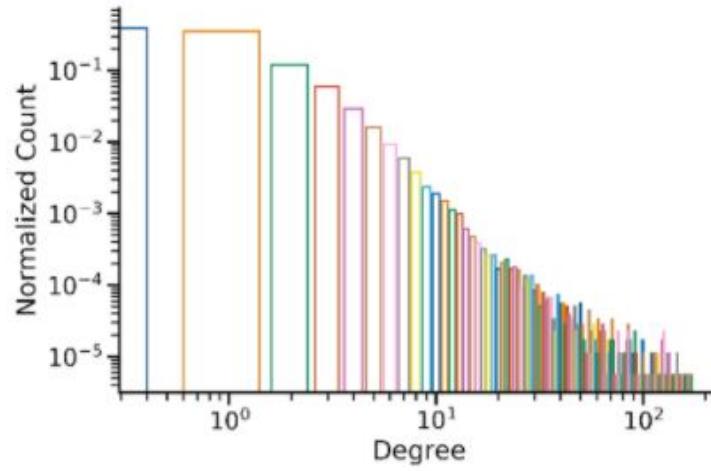


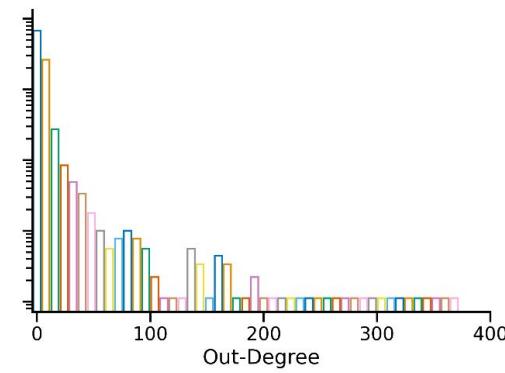
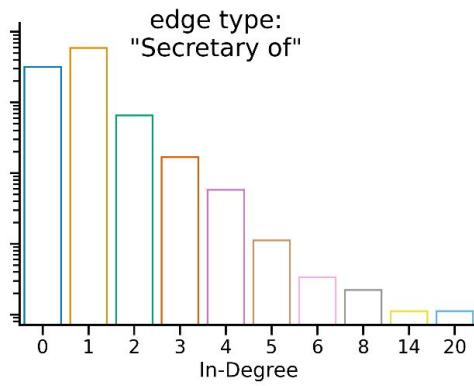
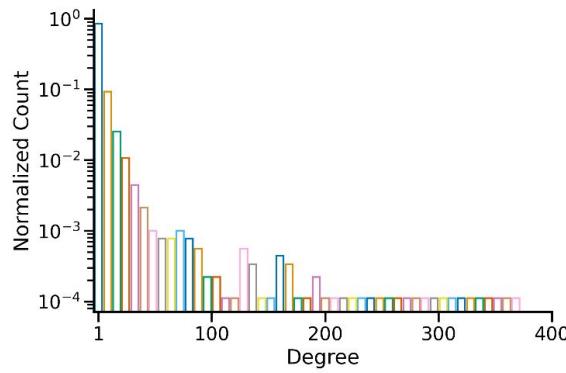
---

# **Subgraphs**

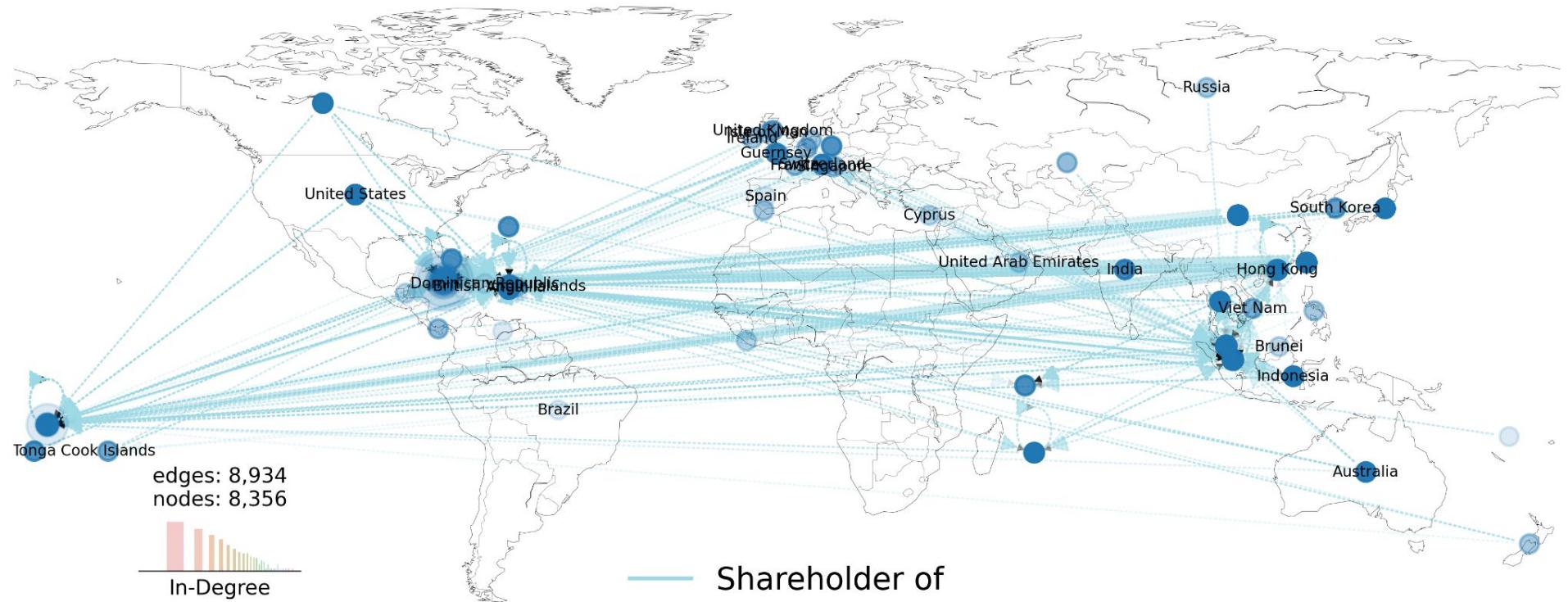
---

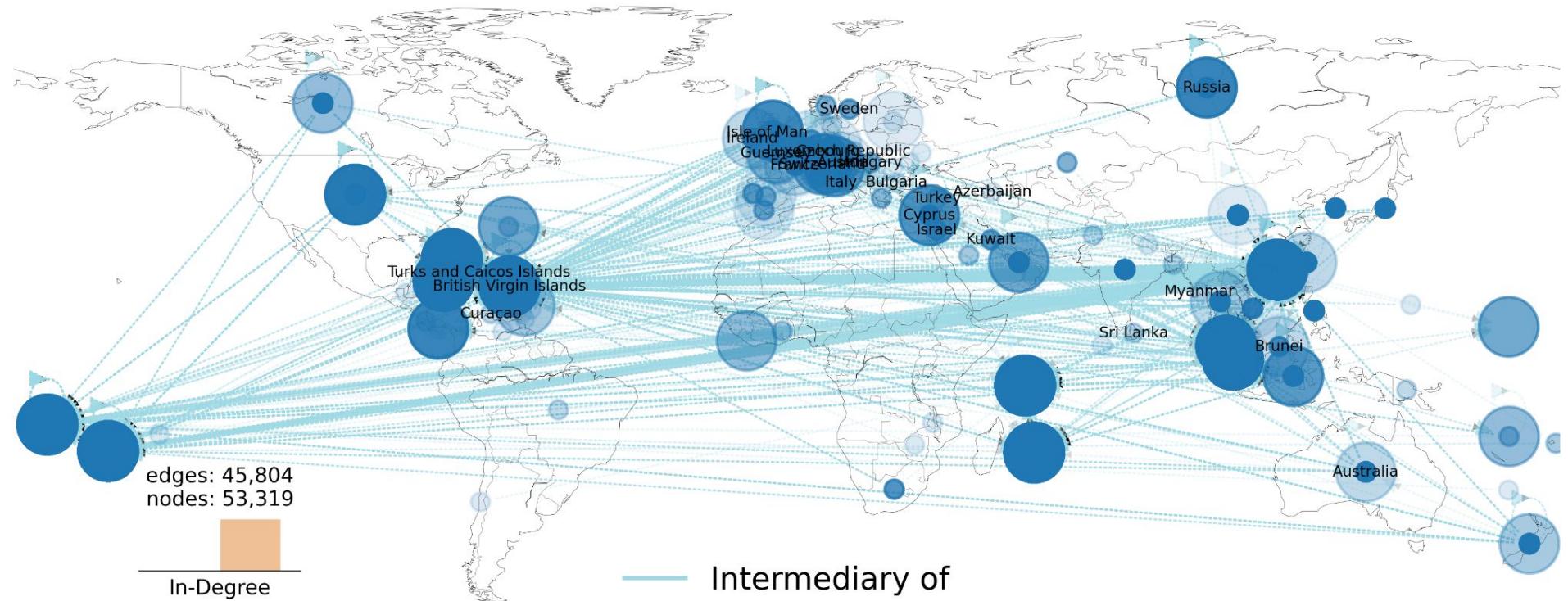
# Degrees

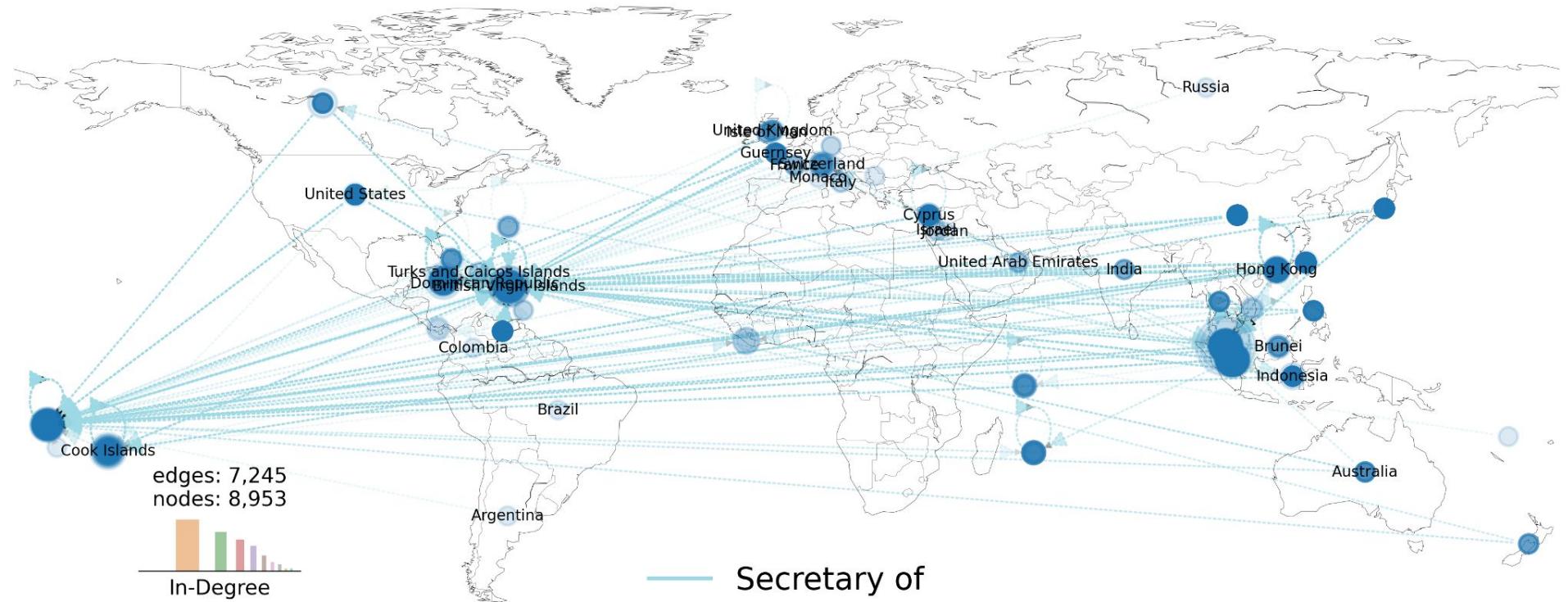


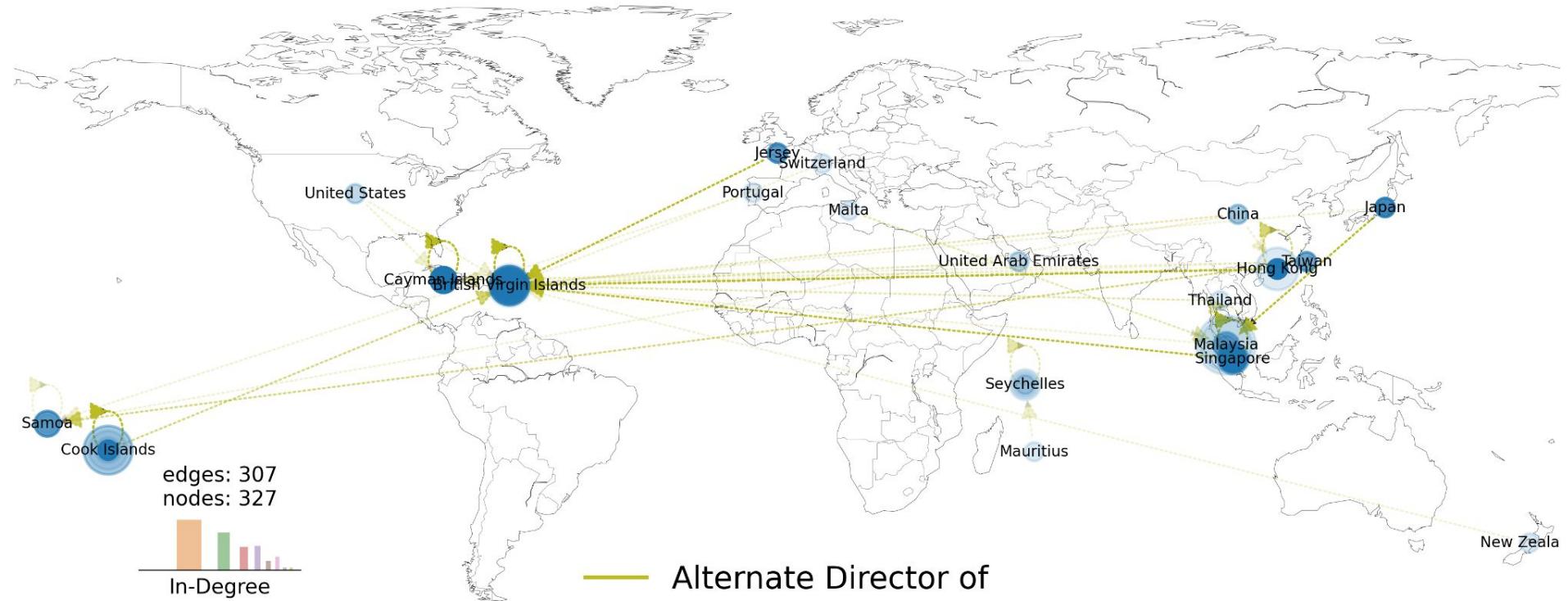


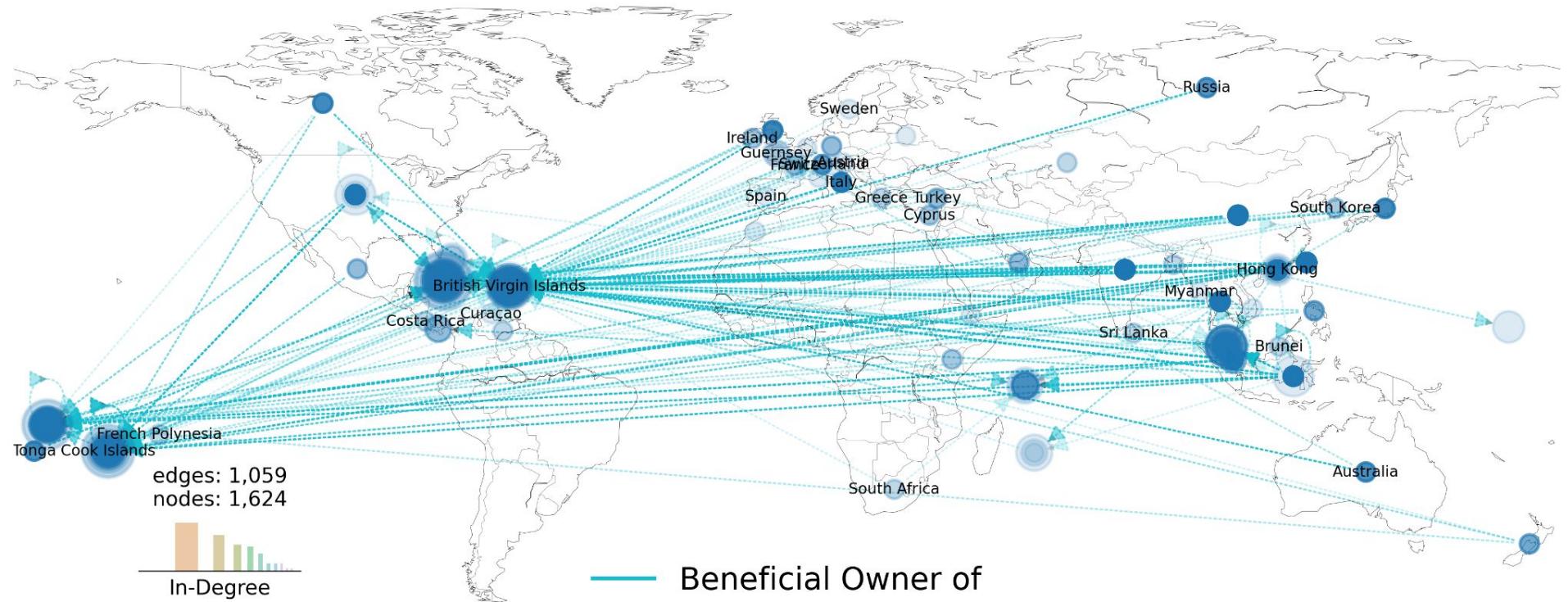




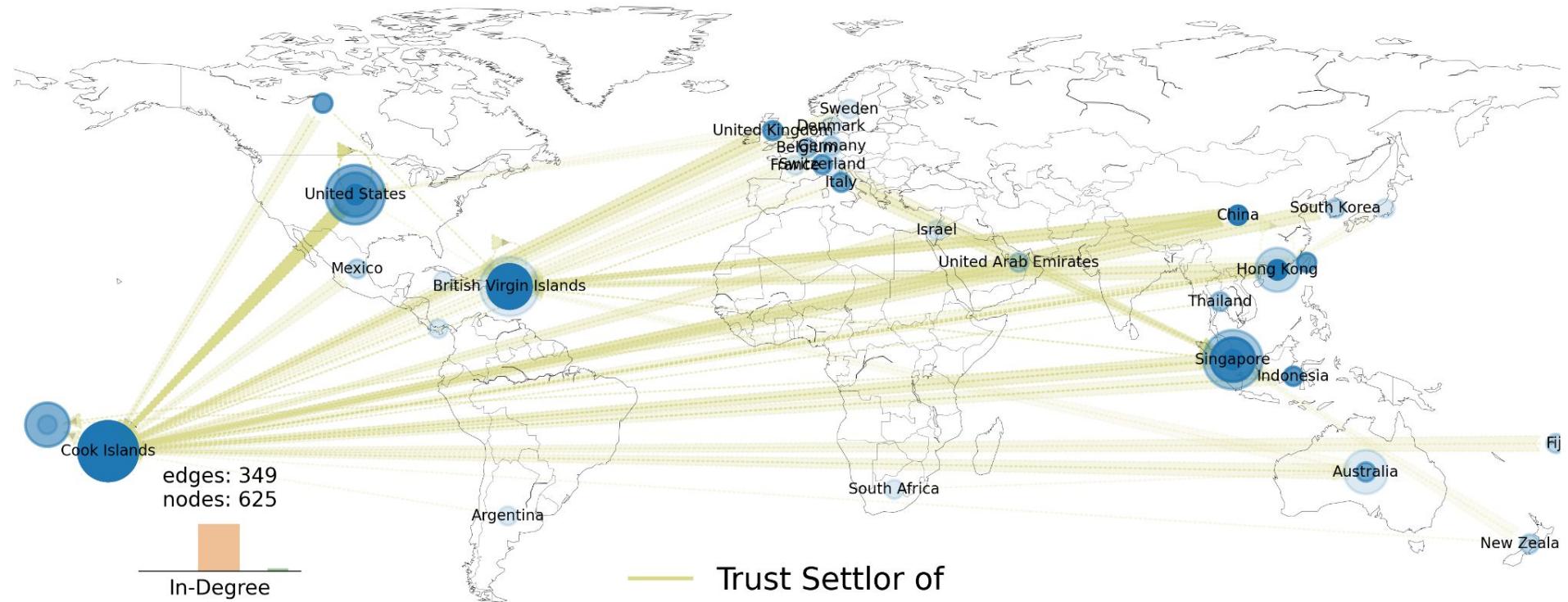


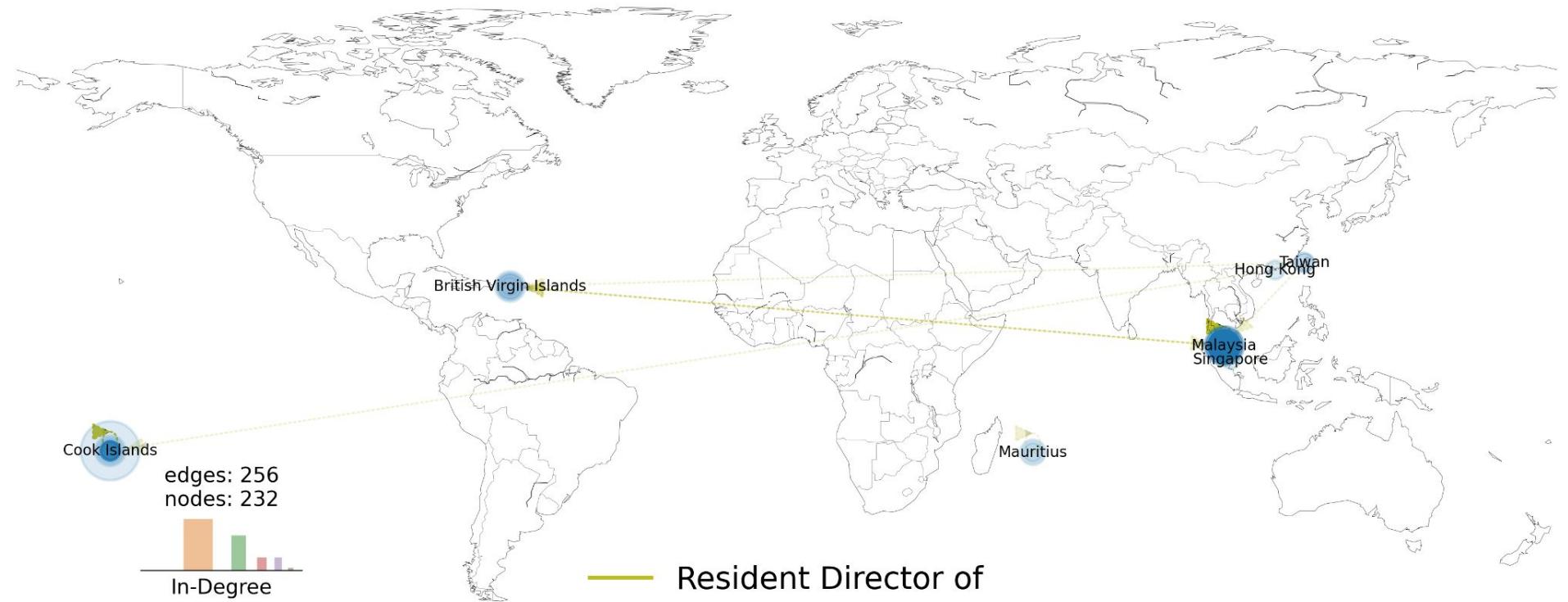


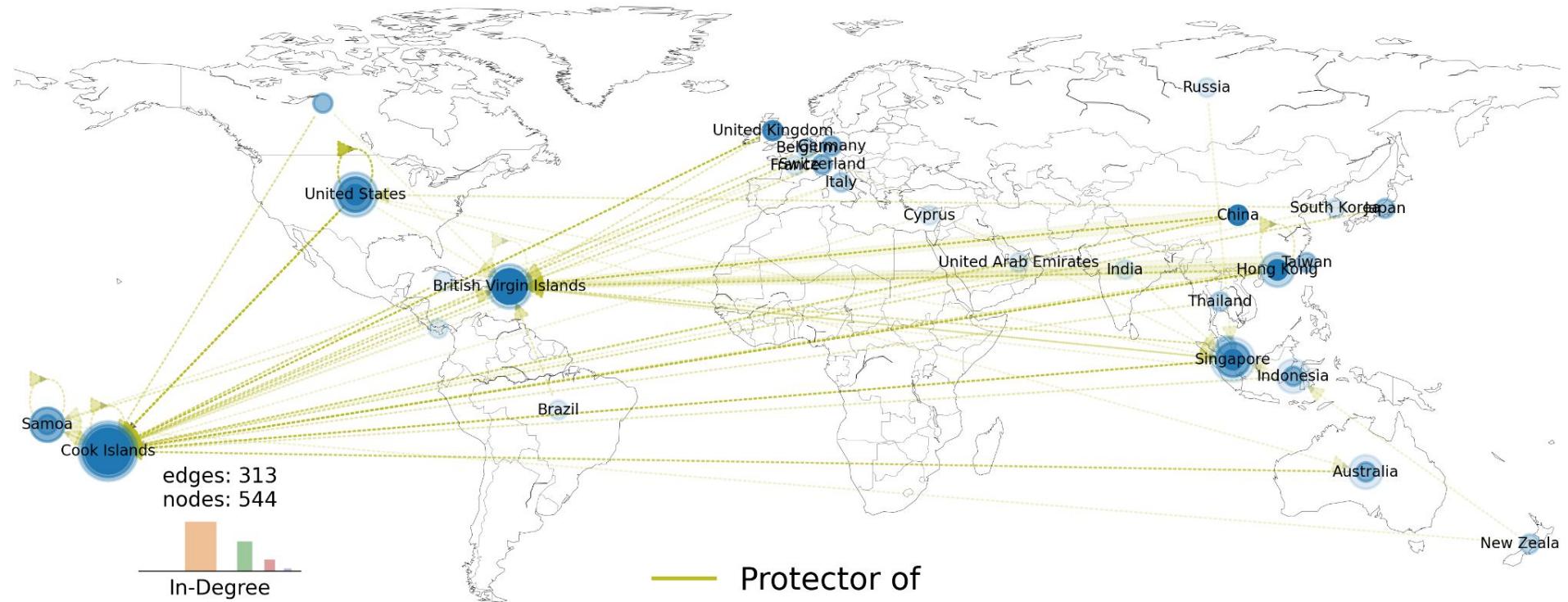




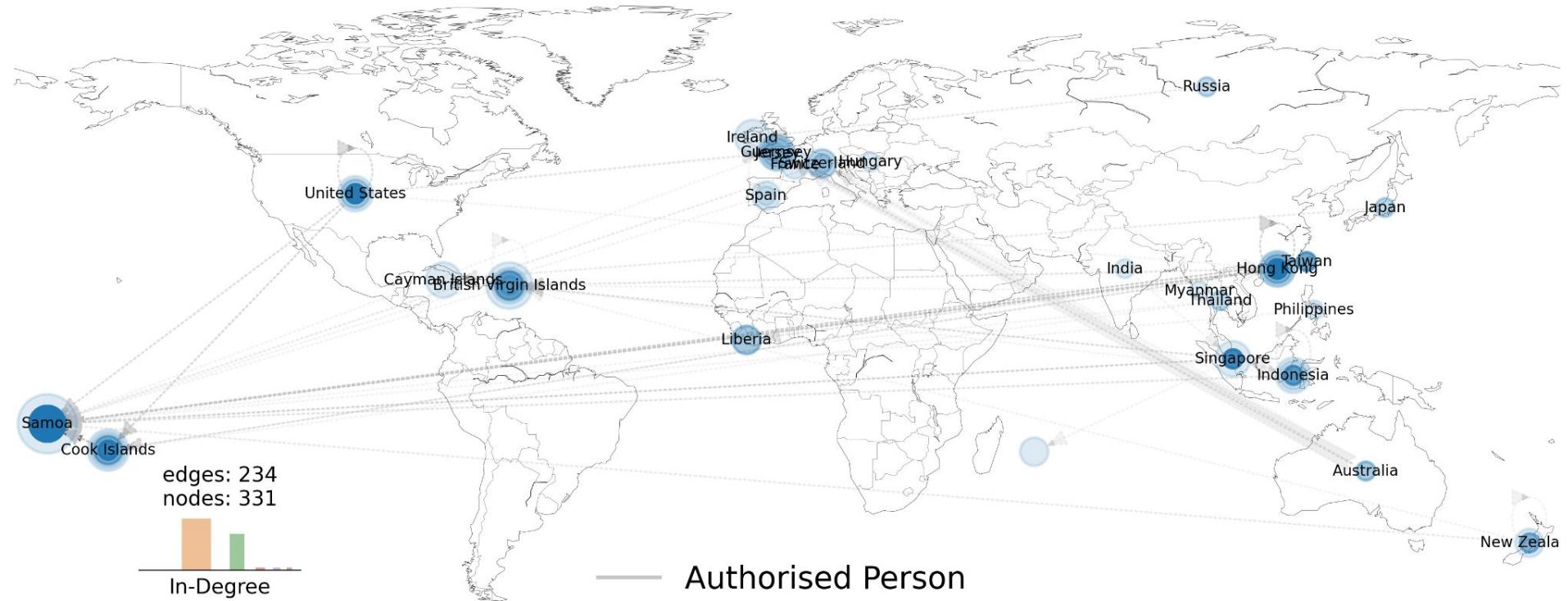


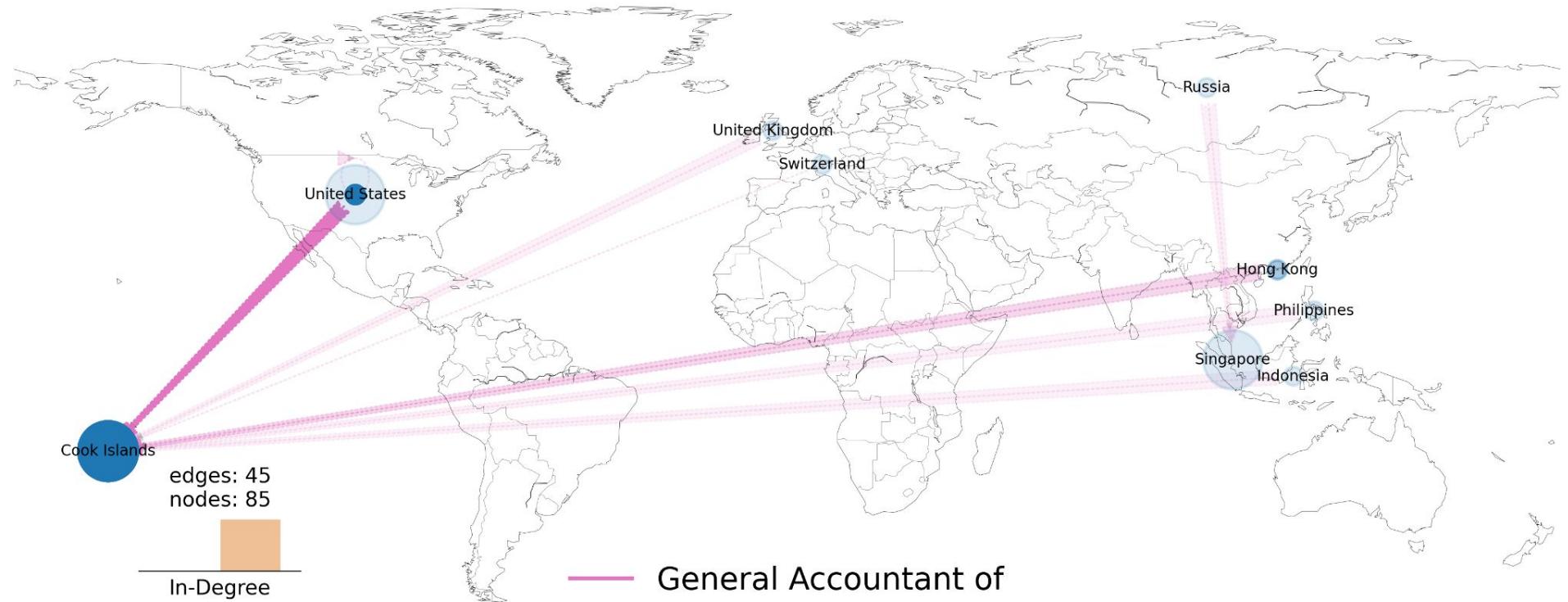


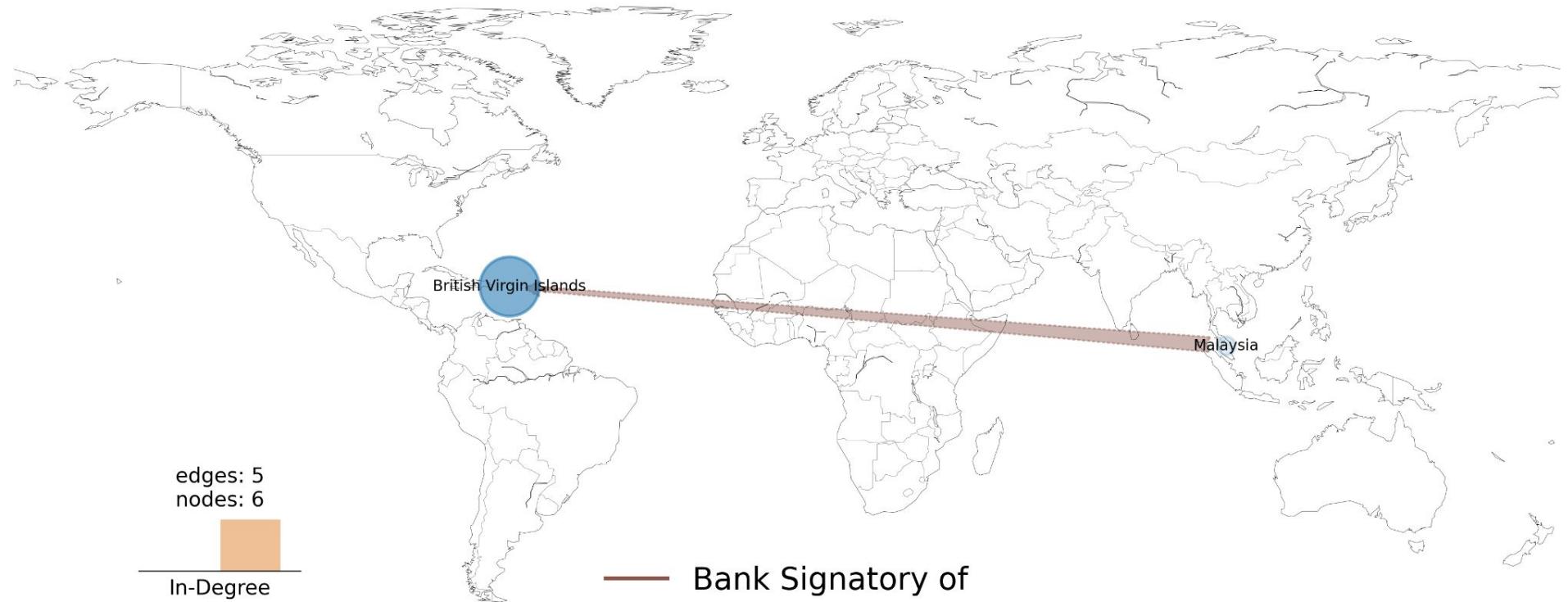


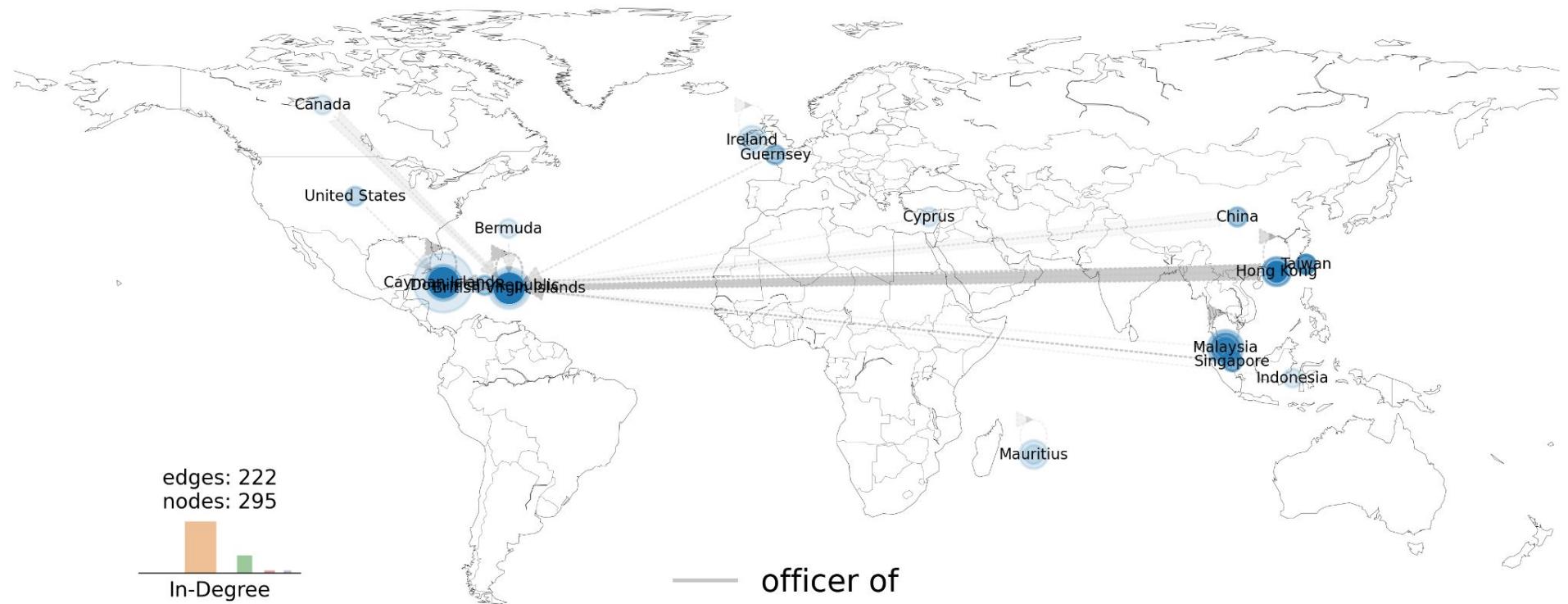






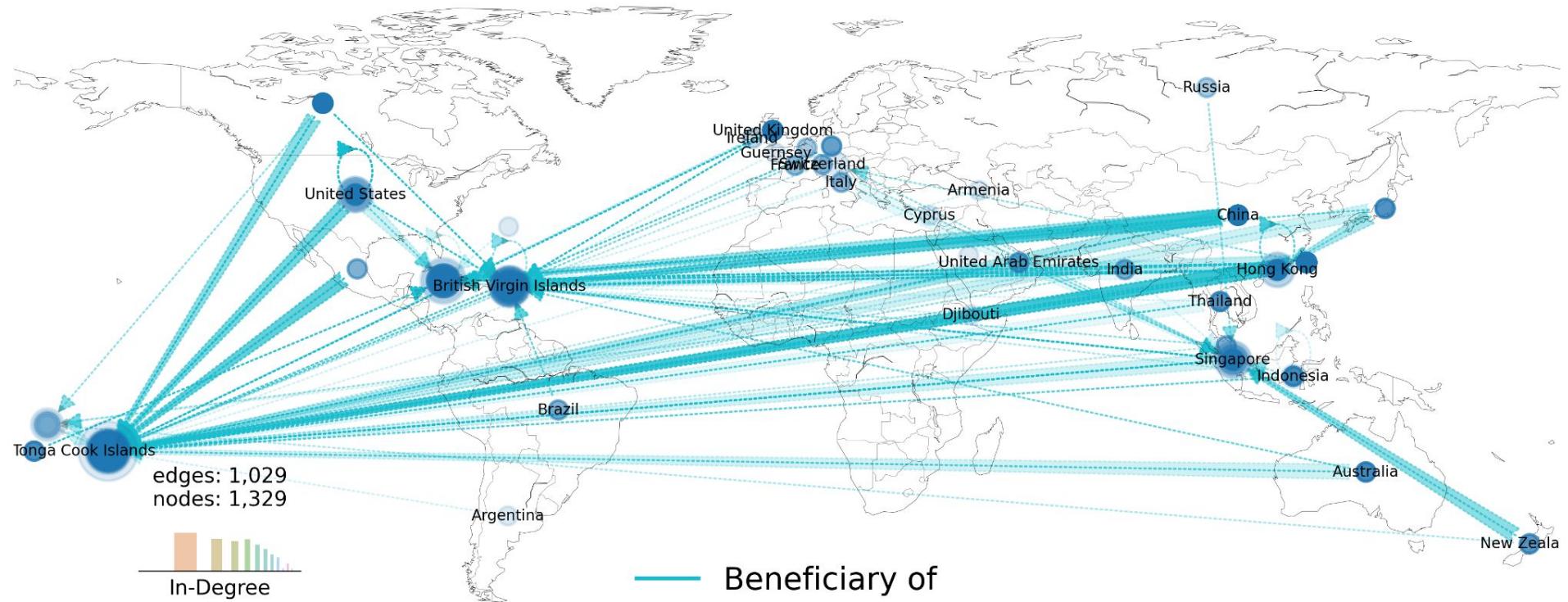


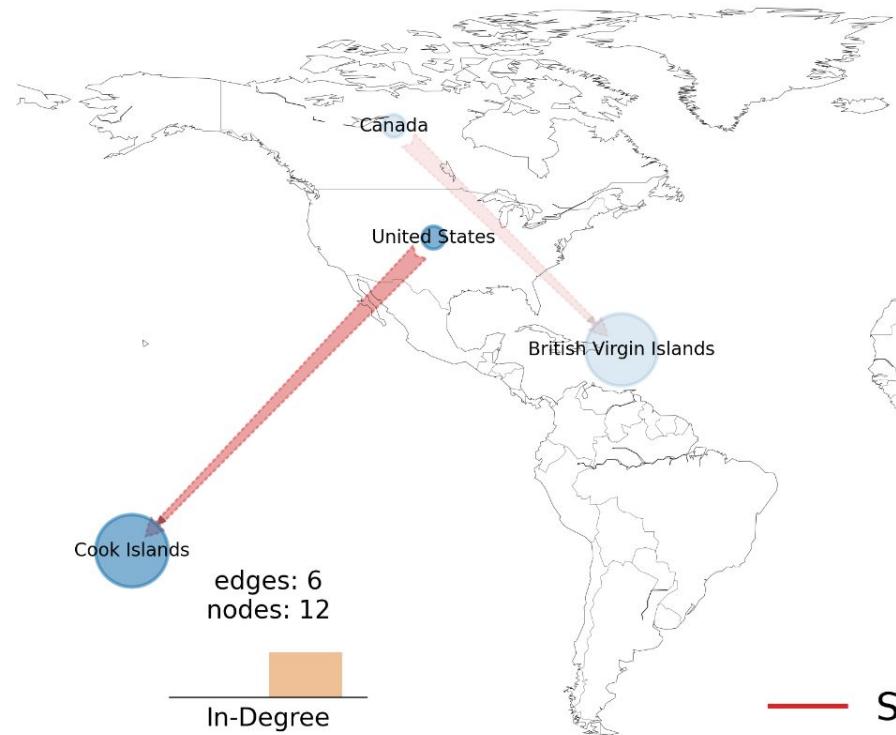




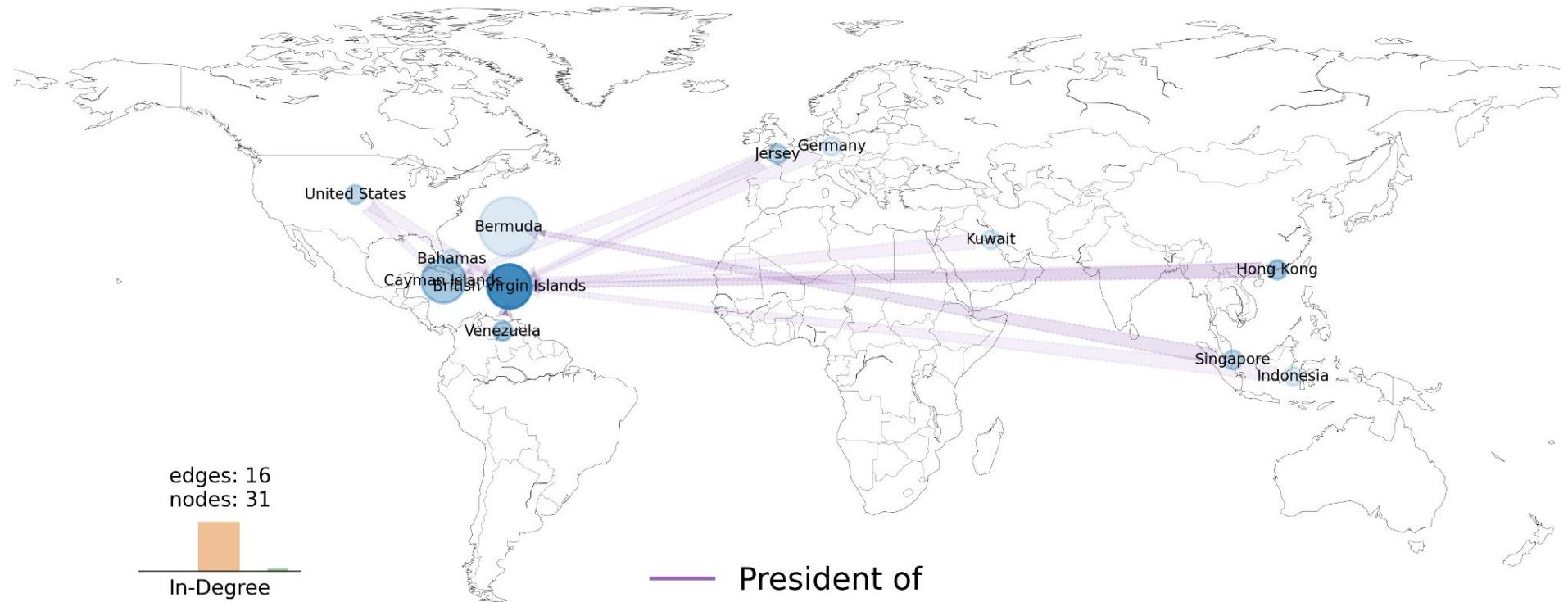


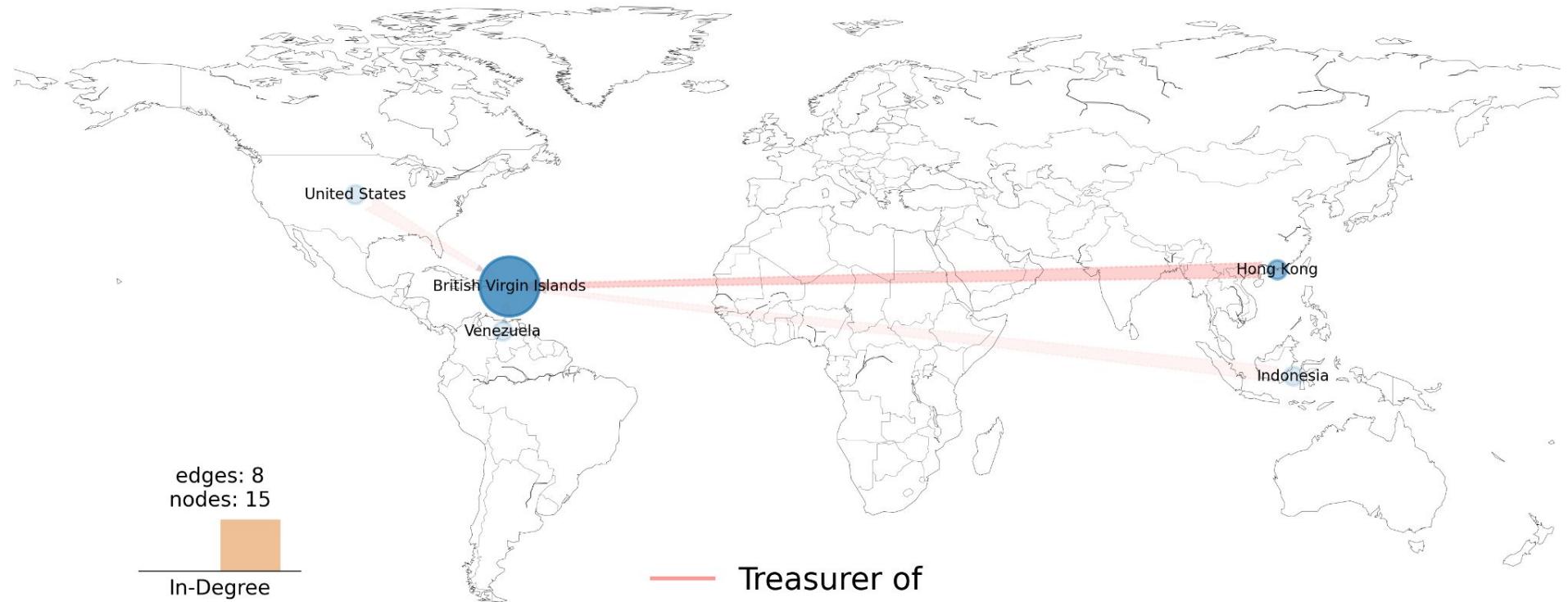




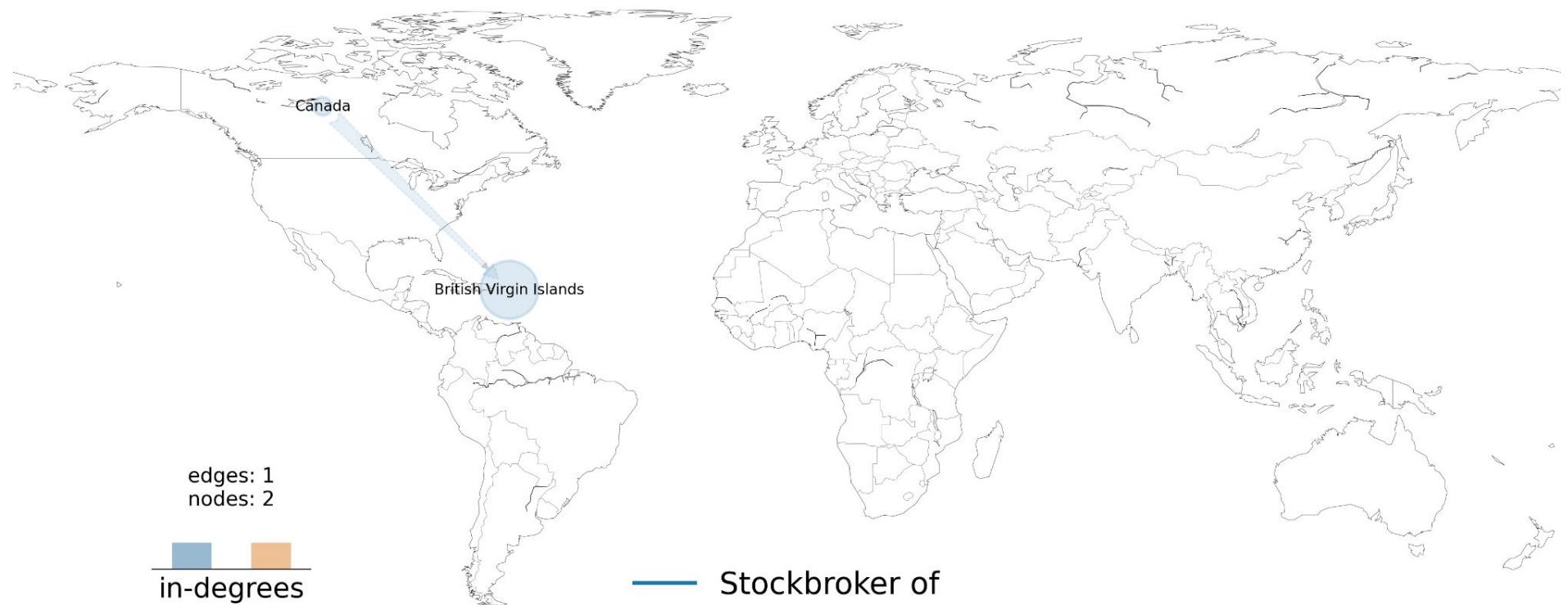










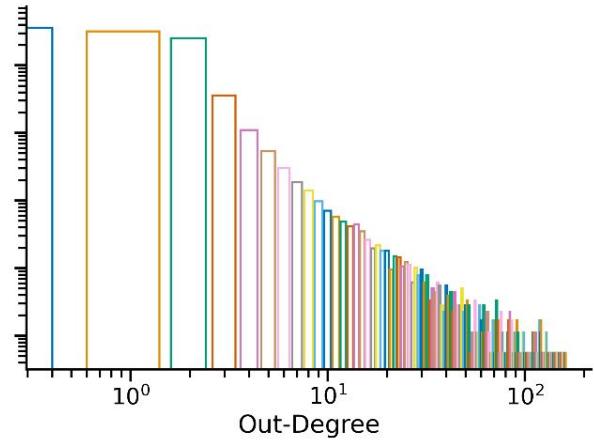
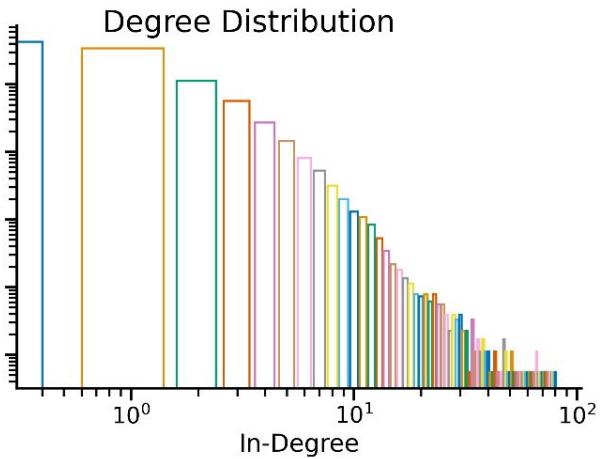
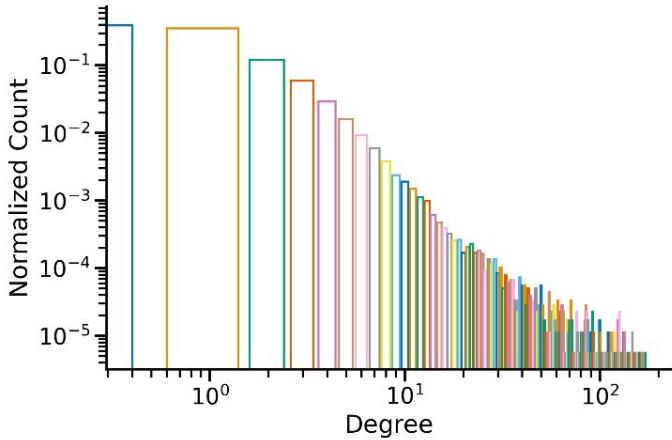


---

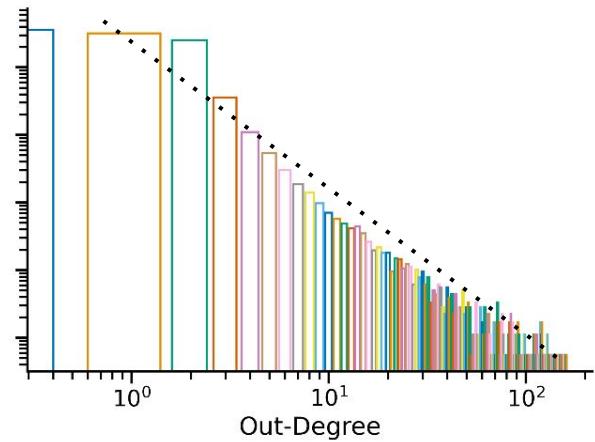
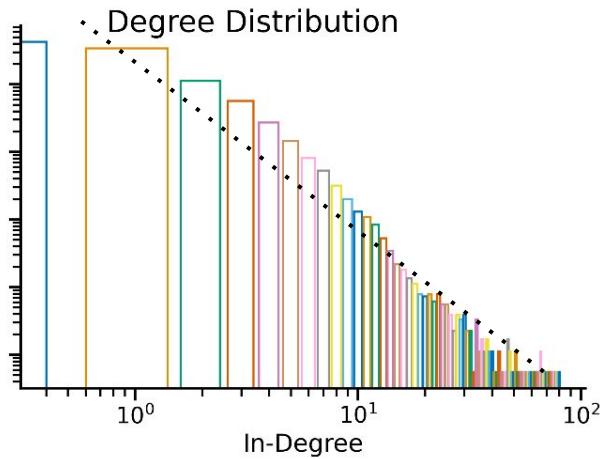
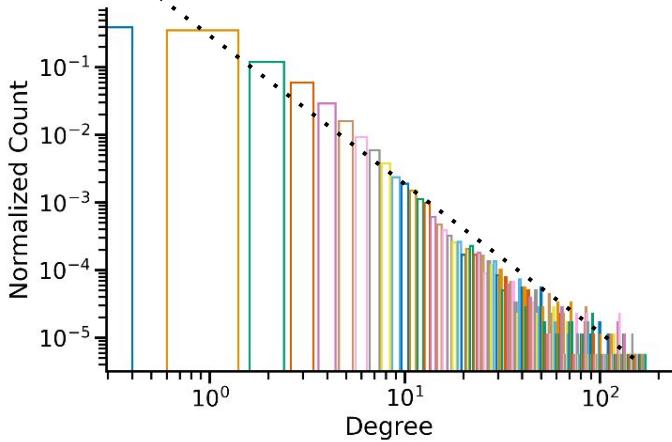
# Degree Distribution

---

# Degrees

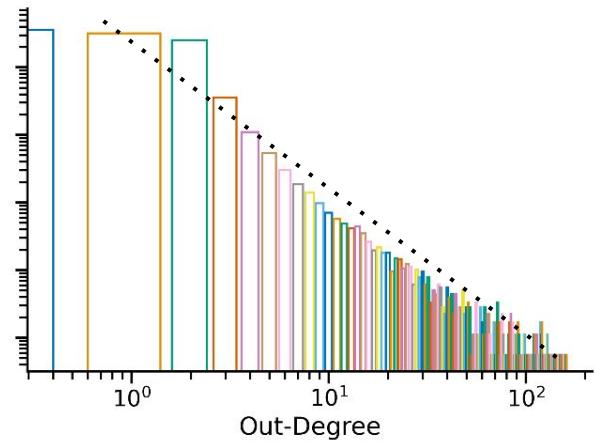
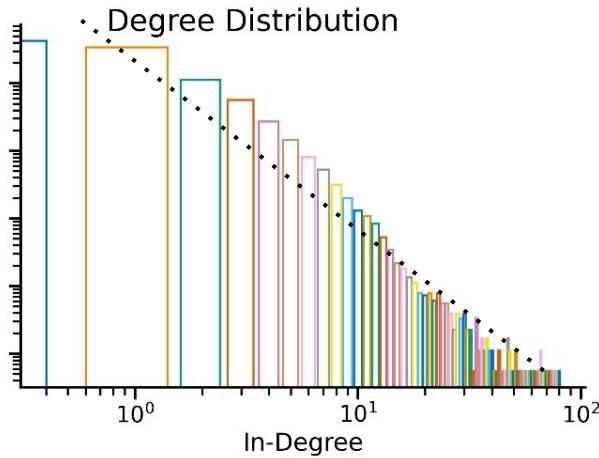
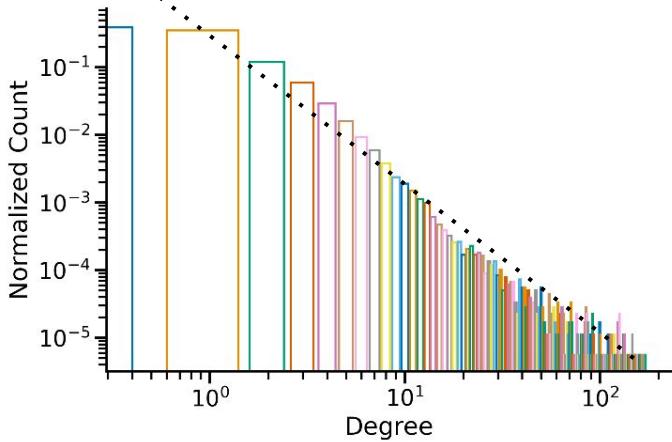


# Degrees



Power Law

# Degrees



Preferential attachment model

---

# Interactive App:

Offshore Leaks Exploration [App](#)

---

---

## **Future Work:**

- Use preferential attachment model assumption to predict future links between nodes
  - Find pattern in network development in time
-

