

# Getting Started with Project Hooked

User guide



Project Hooked will allow the users to explore data related to the fishing industry through 3 different lenses:



Network Risks



Financial Risks



Country Risks

Users can then build a summary heatmap based on customised factors to highlight key groups of interests.

# Overview – Landing Page

The screenshot shows the landing page of a Shiny application titled "A Fishy Business". At the top is a blue navigation menu with the word "Menu" on the left and five links: "Introduction", "Network Risk", "Revenue Risk", "Country Risk", and "Customised Heatmap". The "Introduction" link is highlighted with a yellow border. Below the menu, the title "A Fishy Business" is displayed. A welcome message follows: "Welcome to our Shiny App where you can explore the following modules:". Below this is a teal-bordered box containing three paragraphs of text, each starting with a bolded module name: "Financial Risks", "Country Risks", and "Network Risks". A teal line with a circle at the end points from the "Introduction" link in the menu to the top of the teal-bordered box. Another teal line with a circle at the end points from the bottom of the teal-bordered box to a callout text at the bottom of the slide.

Easy and clear navigation to toggle between different modules on the Shiny App.

Menu Introduction Network Risk Revenue Risk Country Risk Customised Heatmap

## A Fishy Business

Welcome to our Shiny App where you can explore the following modules:

**Financial Risks:** With this module, users can investigate the financial aspects of the data, such as Total Revenue by subgraph, Total Revenue per company, and Total Revenue per Beneficial Owner. The objective is to identify suspicious clusters that may indicate illegal activities.

**Country Risks:** The primary goal of this module is to enable users to explore the data from a country perspective. It addresses questions like "Which subgraph spans across the most countries?" and "Are there any distinctive characteristics of these subgraphs?"

**Network Risks:** This module offers users the capability to visualize the interconnections within each subgraph, facilitating a deeper analysis of the relationships among companies, beneficial owners, and company contacts.

We hope your data exploration journey here leads you to meaningful insights. Together, let us get hooking!

Brief explanation on what each module entails.

# Network Risk – Basic Features



Option to search by either Subgraph ID or Node ID.  
For the given data, the Node ID is the name of the entity

Select a subgraph or entity of interest.  
There are 648 subgraphs (ID 1 to 648) in our Seafood Business Community

Filter to show only companies' Node IDs for search selection

Select the hops to display for the graph entities. Larger hop number will result in more connected entities and relationships being displayed

Select the layout algorithm to display the network graph

Menu

Introduction

Network Risk

Revenue Risk

Country Risk

Customisable Heatmap

Search By

☐ Subgraph ID

☒ Node ID

Select a Seafood Company or its Related Person

Congo Rapids Ltd. Corporation

☒ Filter droplist above to show only companies

Select No. of Hops for the graph

1

5

Select Layout Algorithm

Davidson and Harel(DH)

Show Similar Graphs

☐

Select by id

Select by group

Beneficial Owner

Seafood Company

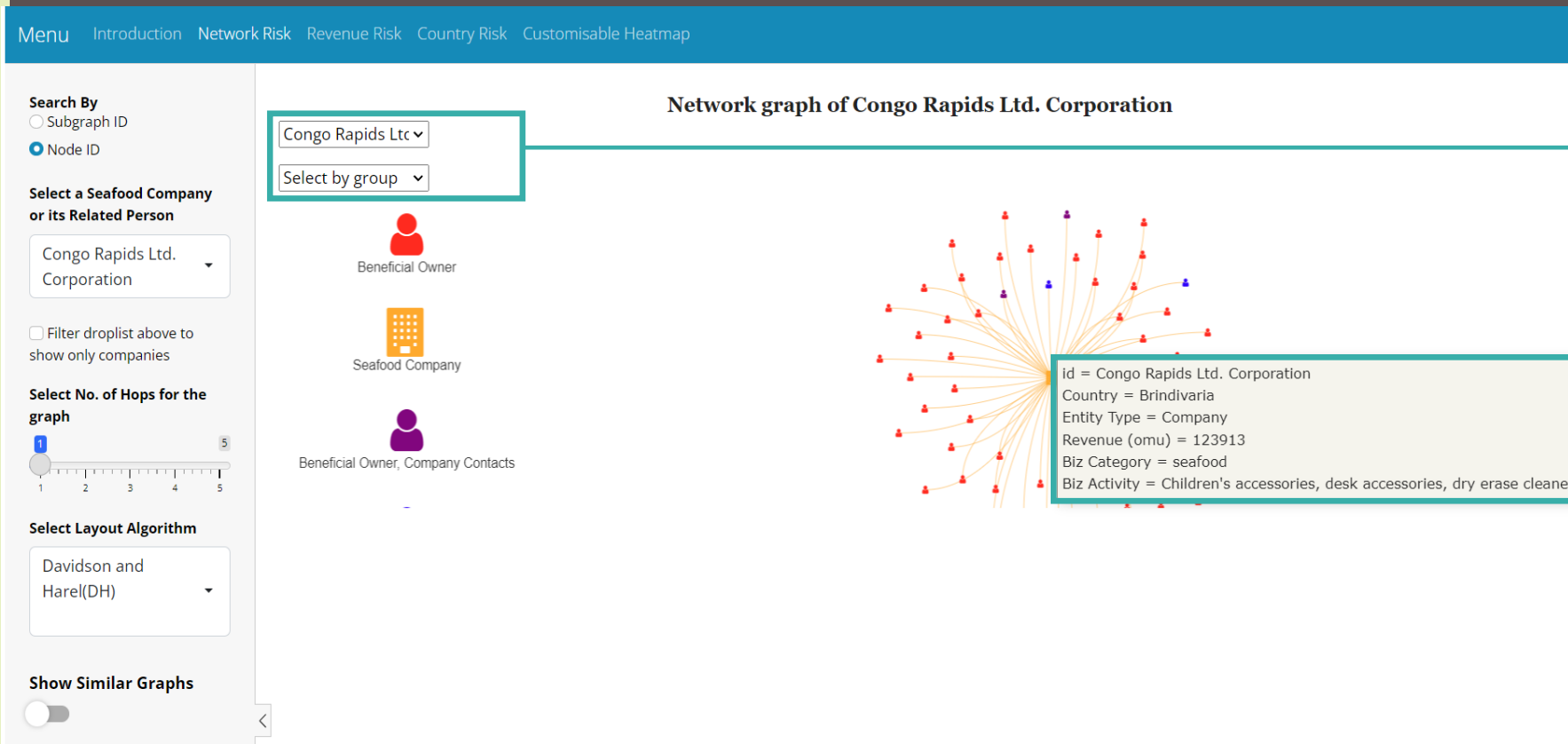
Beneficial Owner, Company Contact

Network graph of Congo Rapids Ltd. Corporation

Try this!

Check out Subgraph 102, what do you think of its structure?

# Network Risk – Exploring the plot



Use the Node ID filter or Class Type filter to locate the Entity or the Entity Class

Mouse over the entity to view their details

# Network Risk – Show Similar Graph Basic Features

## Try this!

Find a subgraph which is similar (in degree distribution ) to Subgraph 210

Switch to toggle on and off the hidden panel which contains selections to identify similar graphs for comparison

Button to select one of the 3 available methods to compare the subgraphs

Slider to select the number of top similar graphs to return

Button to perform the graph comparison and display a similar graph beside the current one

Menu

Introduction

Network Risk

Revenue Risk

Country Risk

Customisable Heatmap

Search By

☐ Subgraph ID

☒ Node ID

Select a Seafood Company or its Related Person

Congo Rapids Ltd. Corporation

☐ Filter droplist above to show only companies

Select No. of Hops for the graph

1

2

3

4

5

Select Layout Algorithm

Davidson and Harel(DH)

Select by id

Select by group

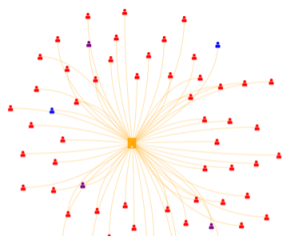
Beneficial Owner

Seafood Company

Beneficial Owner, Company Contacts

Company Contacts

Network graph of Congo Rapids Ltd. Corporation



Show Similar Graphs

☒

(Note: Hop count of the original subgraph is set to 1 to run this comparison.)

Compare By

☒ Structure

☐ Degree Distribution

☐ Graph Size

List Top N similar subgraphs

1

2

3

4

5

6

7

8

9

10

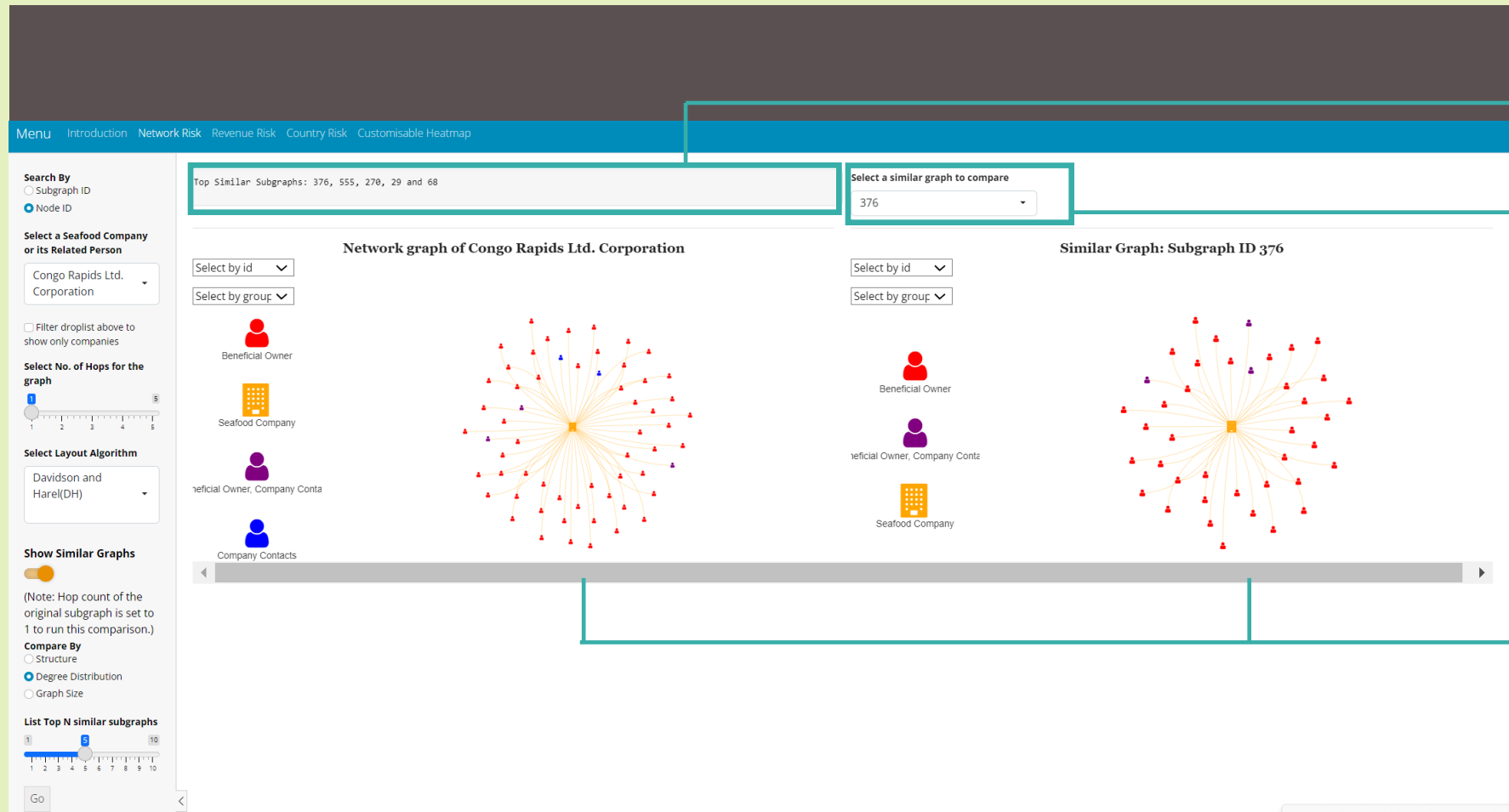
Go

## The Comparison Methods

Structure: Based on graphlet kernel which is used to measure the similarity between graphs by counting subgraphs with a limited size  $k^3$

Degree Distribution: The Jensen-Shannon divergence measure is used to compare degree distributions of graphs  
Graph Size: Calculates the overall similarity between the two subgraphs by taking an average of the node and edge similarities.

# Network Risk – Exploring the Comparative Plot



Displays the list of subgraphs which are like the current graph (left)

Option to select which similar subgraph to serve as the comparative

The current subgraph (left) and a similar subgraph (right) are displayed side-by-side to ease of comparison

# Financial Risk – Basic Features



Key statistics to help anchor the data exploration process, for easy comparisons

Quick selection of specific subgraph of interest, to understand where they stand vis-à-vis other subgraphs, across variables

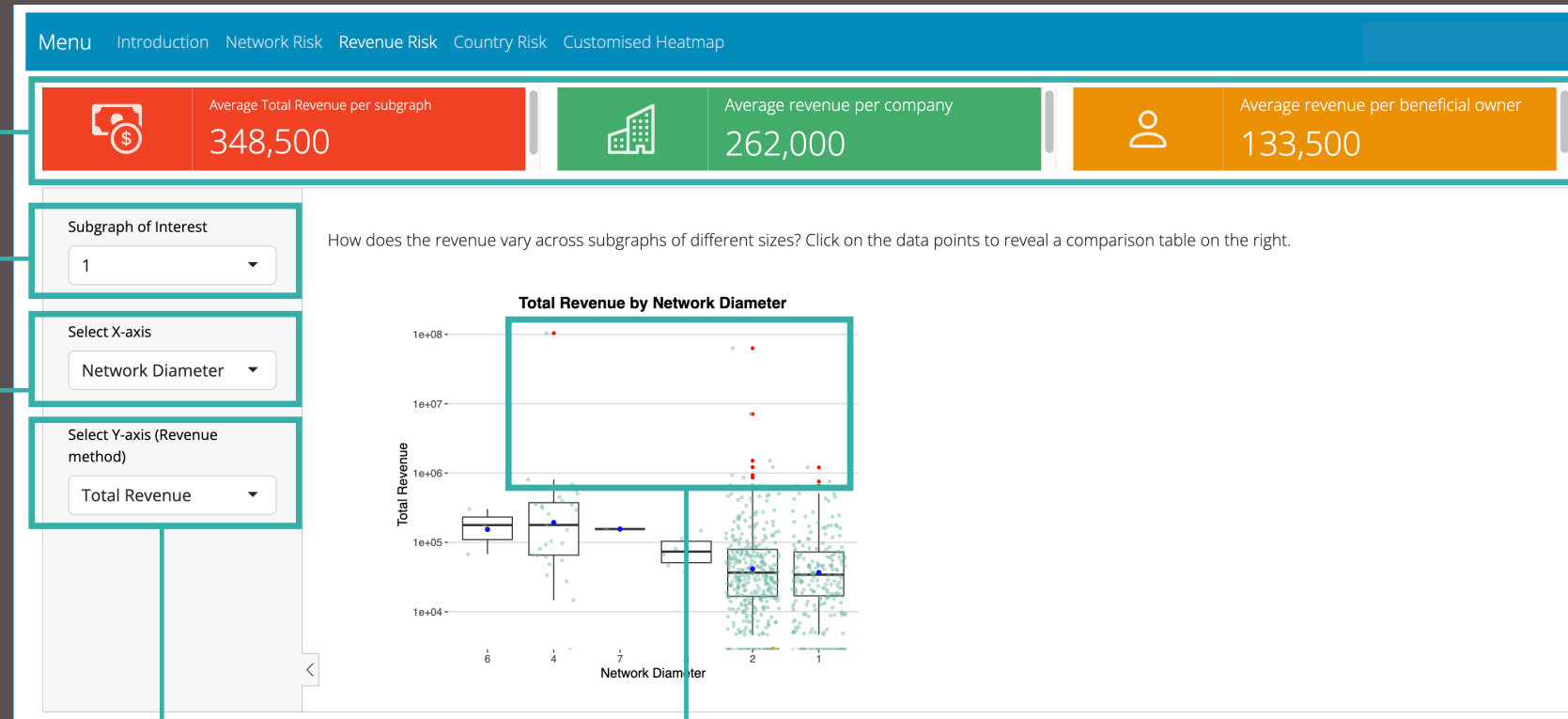
Options to customize the x-axis of the boxplot. Options include:

- Network Diameter
- Size of subgraph (number of nodes)

Try this!  
Checkout Subgraph 55. Does the revenue look “normal” to you?

Options to customize the y-axis of the boxplot. Options include:

- Total Revenue (per subgraph)
- Revenue per company
- Revenue per beneficial owner



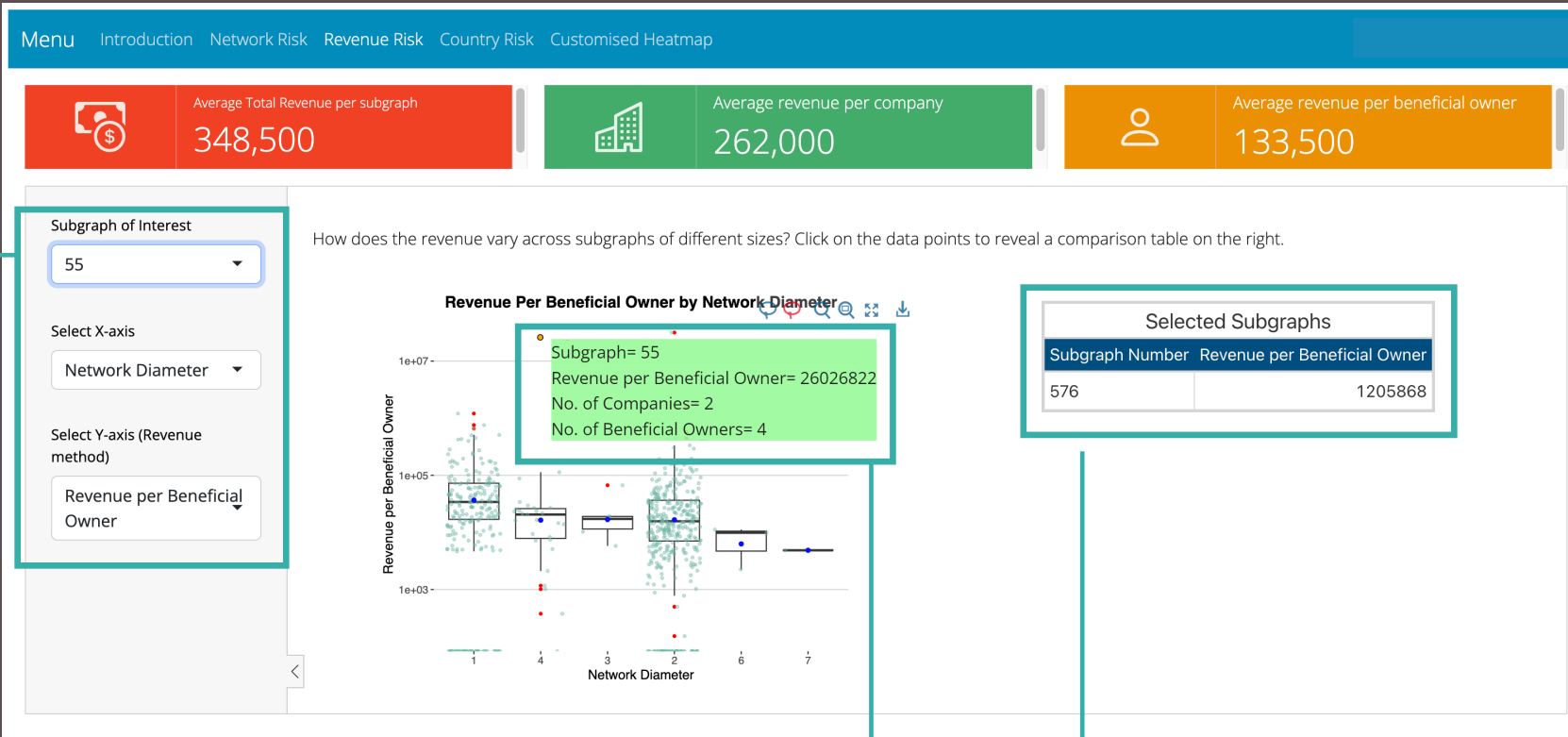
Outlier nodes are highlight in red.



# Financial Risk – Exploring subgraphs



Select boxplot parameters



Node of selected subgraph will appear in orange, with a tooltip

Users are also able to select other nodes by clicking on the boxplot, on top of subgraph 55 to make comparisons. Information of any additional selected will appear in table on the right.

# Country Risk – Basic Features



Try this!

Check out Subgraph 29. How is it special?

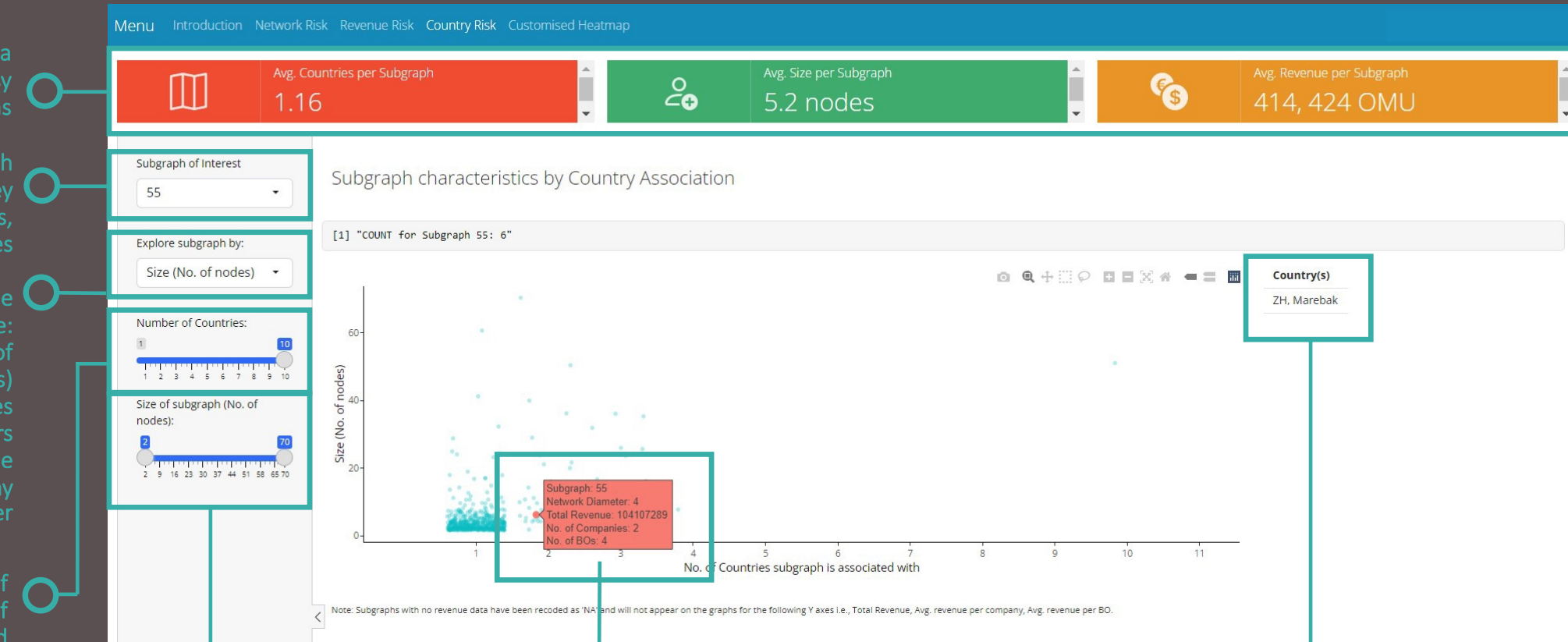
Key statistics to help anchor the data exploration process, for easy comparisons

Quick selection of specific subgraph of interest, to understand where they stand vis-à-vis other subgraphs, across variables

Options to customize the y-axis of the boxplot. Options include:

- Size of subgraph (number of nodes)
  - Number of companies
  - Number of beneficial owners
    - Total Revenue
    - Avg. revenue per company
    - Avg. revenue per beneficial owner

Option to filter x-axis range of scatterplot, based on number of countries subgraphs are associated with



Option to filter the data points in the scatterplot, based on size of subgraphs

Aligned with the financial risk module, selected subgraph of interest would be highlighted in orange, with relevant data presented in the tooltip

Country associated with subgraph presented in table

# Customisable Heatmap - Clustering

Navigate to optimization tab to fine tune the parameters in **orange** boxes

Customize x-axis by selecting indicators of interest. Options include:

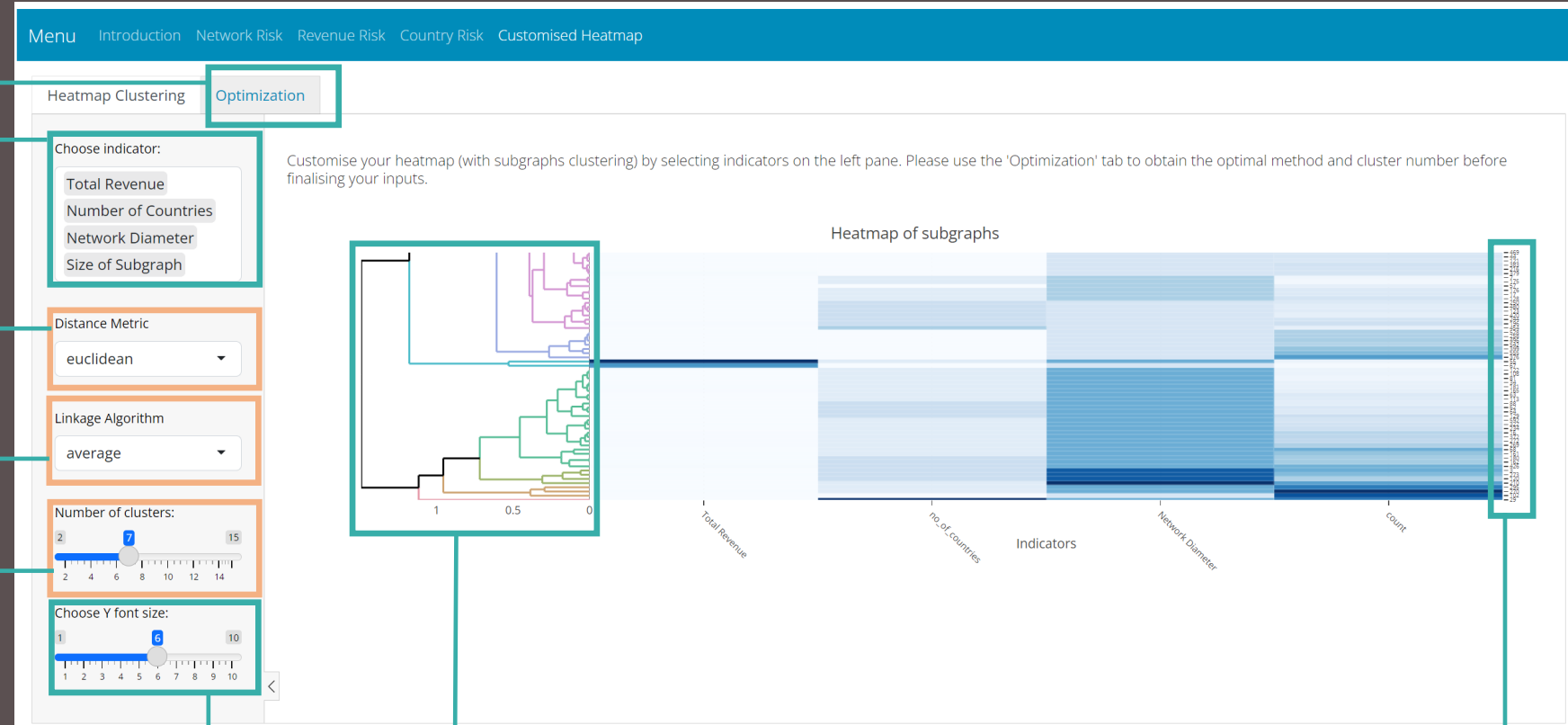
- Total Revenue
- Revenue per Company
- Revenue per Beneficial Owner
  - Number of Countries
  - Number of Companies
- Number of Beneficial Owners
  - Network Diameter
  - Size of Subgraph

Select method used to compute distance matrix. Options include:

- euclidean
- maximum
- manhattan
- canberra

Select hclust method. Options include:

- ward.D
- single
- complete
- average
- mcquitty
- median
- centroid



Specify number of clusters

Adjust size of Y axis labels

Dendrogram showing cluster hierarchy

Try this!

Which subgraphs are similar if we just look at their revenue and number of BOs

Subgraph IDs

# Customisable Heatmap - Optimisation

We are now in the Optimization tab.

- Step 1:  
Select indicators of interest. Options include:
- Total Revenue
  - Revenue per Company
  - Revenue per Beneficial Owner
  - Number of Countries
  - Number of Companies
  - Number of Beneficial Owners
  - Network Diameter
  - Size of Subgraph

Step 3: Select 'centroid'

MenuIntroductionNetwork RiskRevenue RiskCountry RiskCustomised Heatmap

Heatmap ClusteringOptimization

Choose indicator:

Revenue per Company

Number of Countries

Number of Companies

Number of Beneficial Owners

Network Diameter

Linkage Algorithm

centroid

Find highest optimal clustering method and optimal number of clusters

dist_methods	hclust_methods	optim
unknown	ward.D	0.43
unknown	ward.D2	0.74
unknown	single	0.86
unknown	complete	0.90
unknown	average	0.95
unknown	mcquitty	0.91
unknown	median	0.91
unknown	centroid	0.96

Estimating the number of clusters using average silhouette width

Number of clusters (k)	Average silhouette width
1	0.00
2	0.58
3	0.72
4	0.75
5	0.70
6	0.72
7	0.68
8	0.70
9	0.68
10	0.68

Step 2:  
'centroid' method should be used because it gave the highest optimum value

Step 4:  
Optimal number of clusters = 4