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

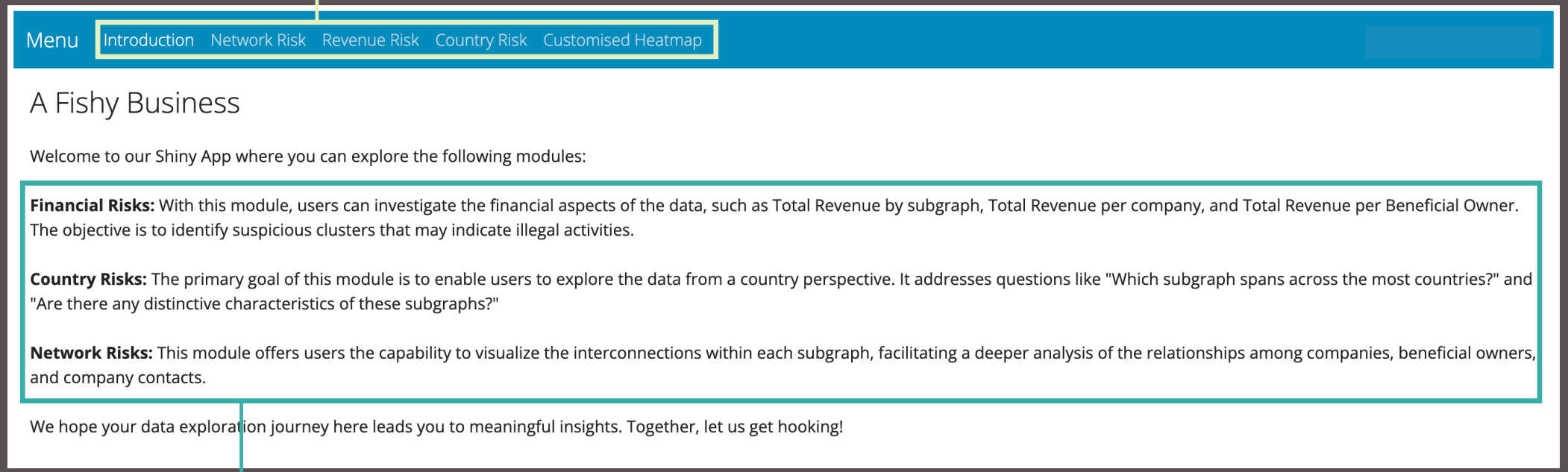
+

Country Risks

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

Overview – Landing Page

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



Brief explanation on what each module entails.

Network Risk – Basic Features

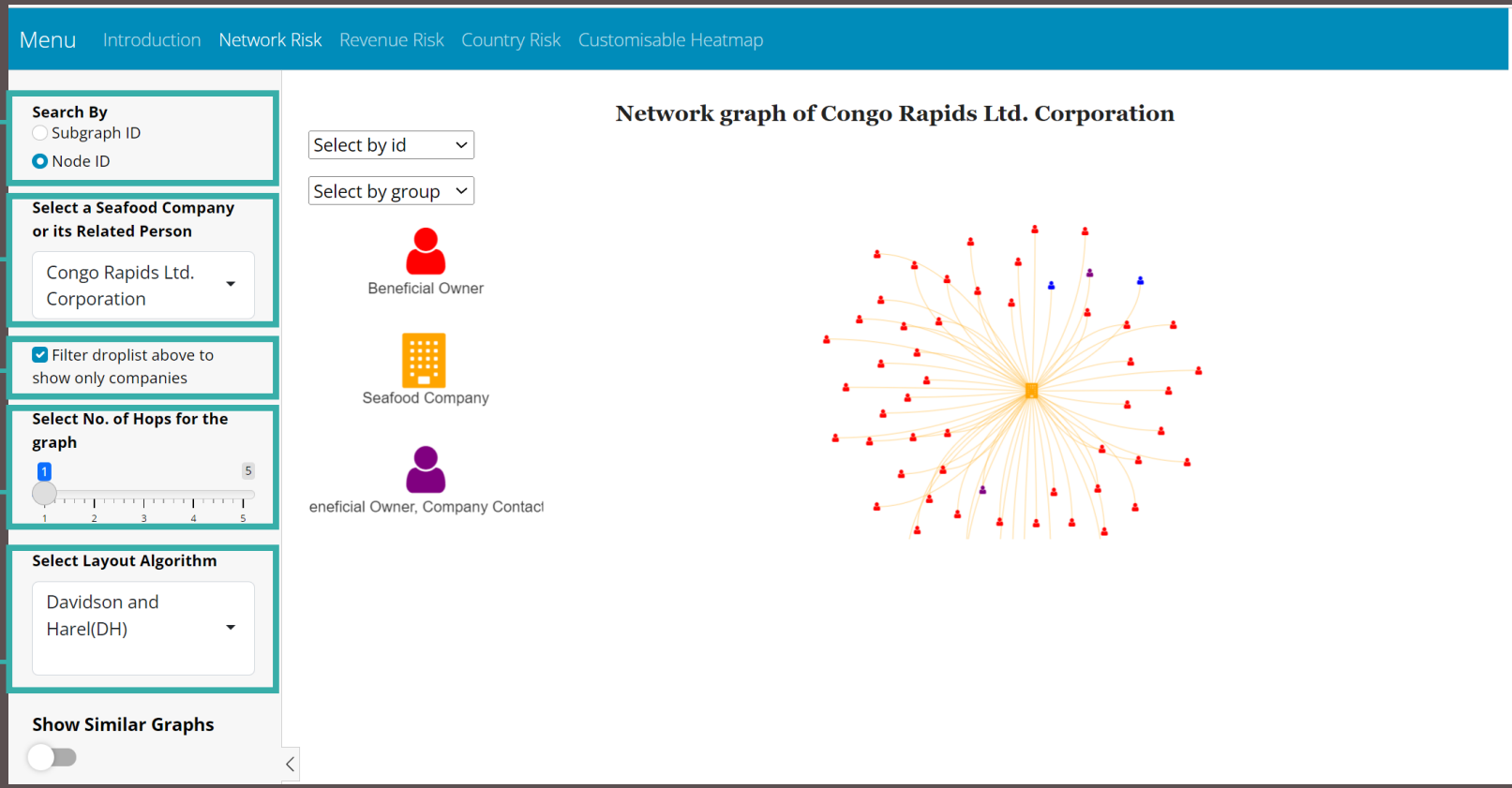
Option to search by either Subgraph ID or Node ID.
For the given data, the Node ID is the name of the entity

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

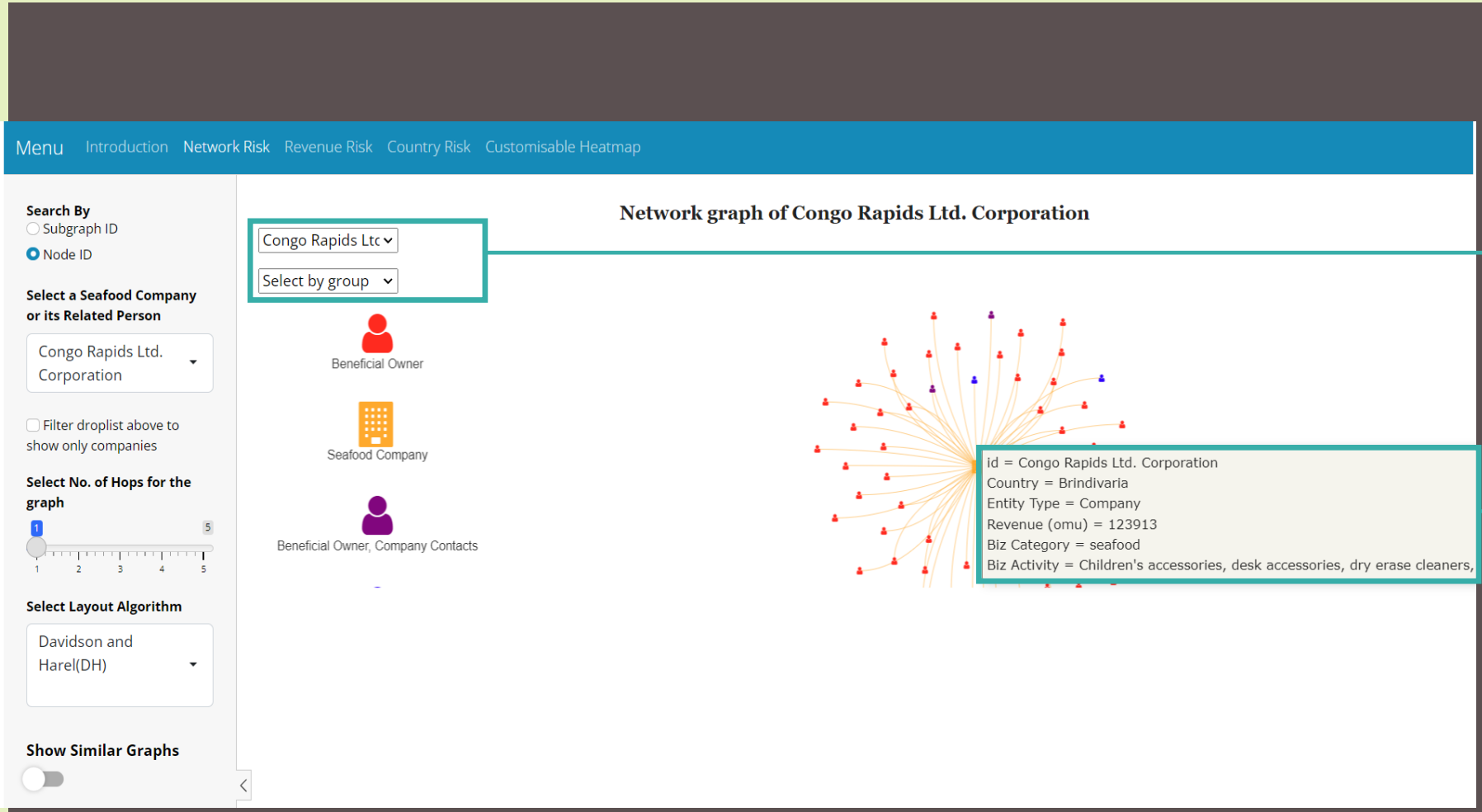
Select the hops to display for the graph entities. Larger hop number will result in more connected entities and relationships being displayed

Select the layout algorithm to display the network graph



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

Network Risk – Exploring the plot

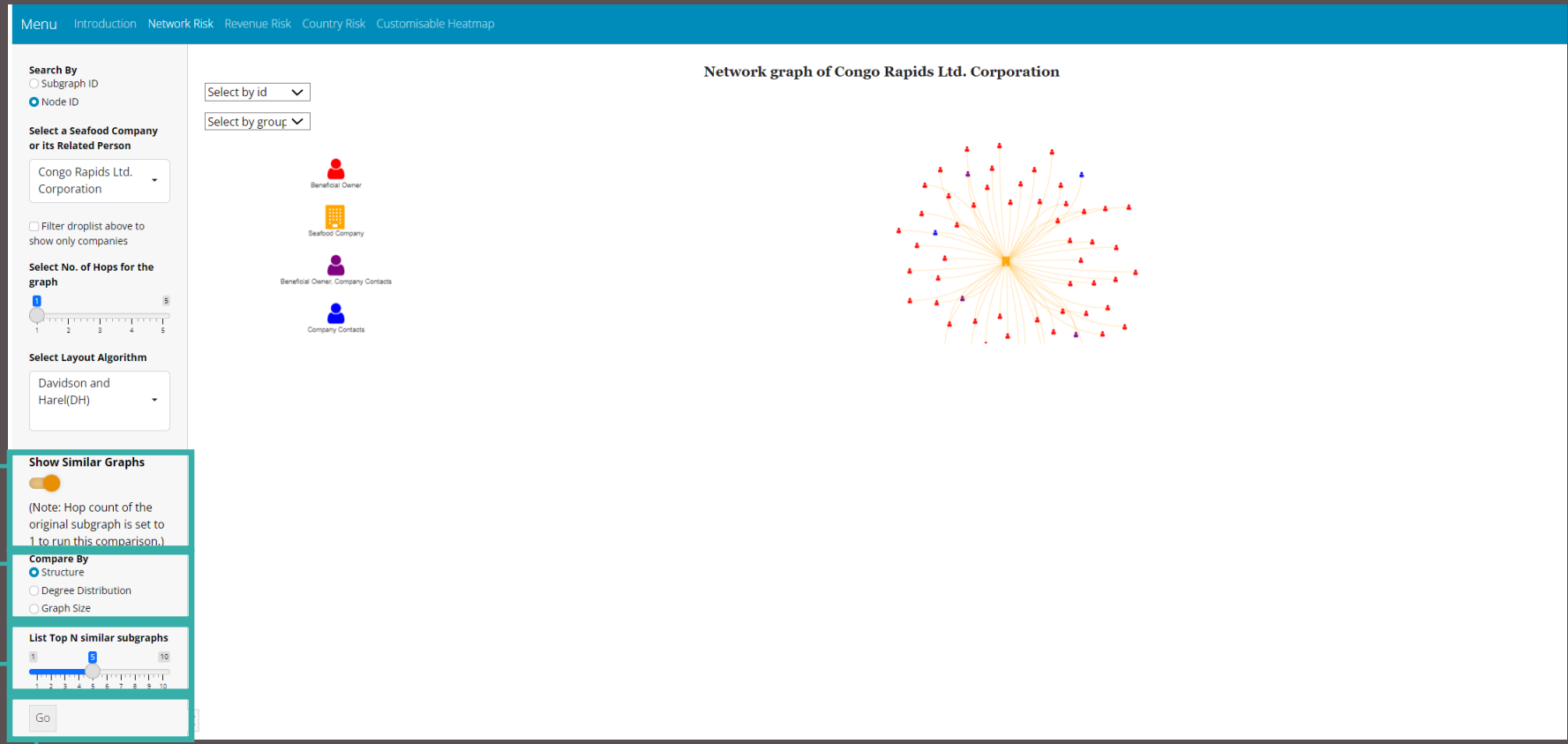


Use the Node ID filter or Class Type filter to locate the Entity or the Entity Class

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

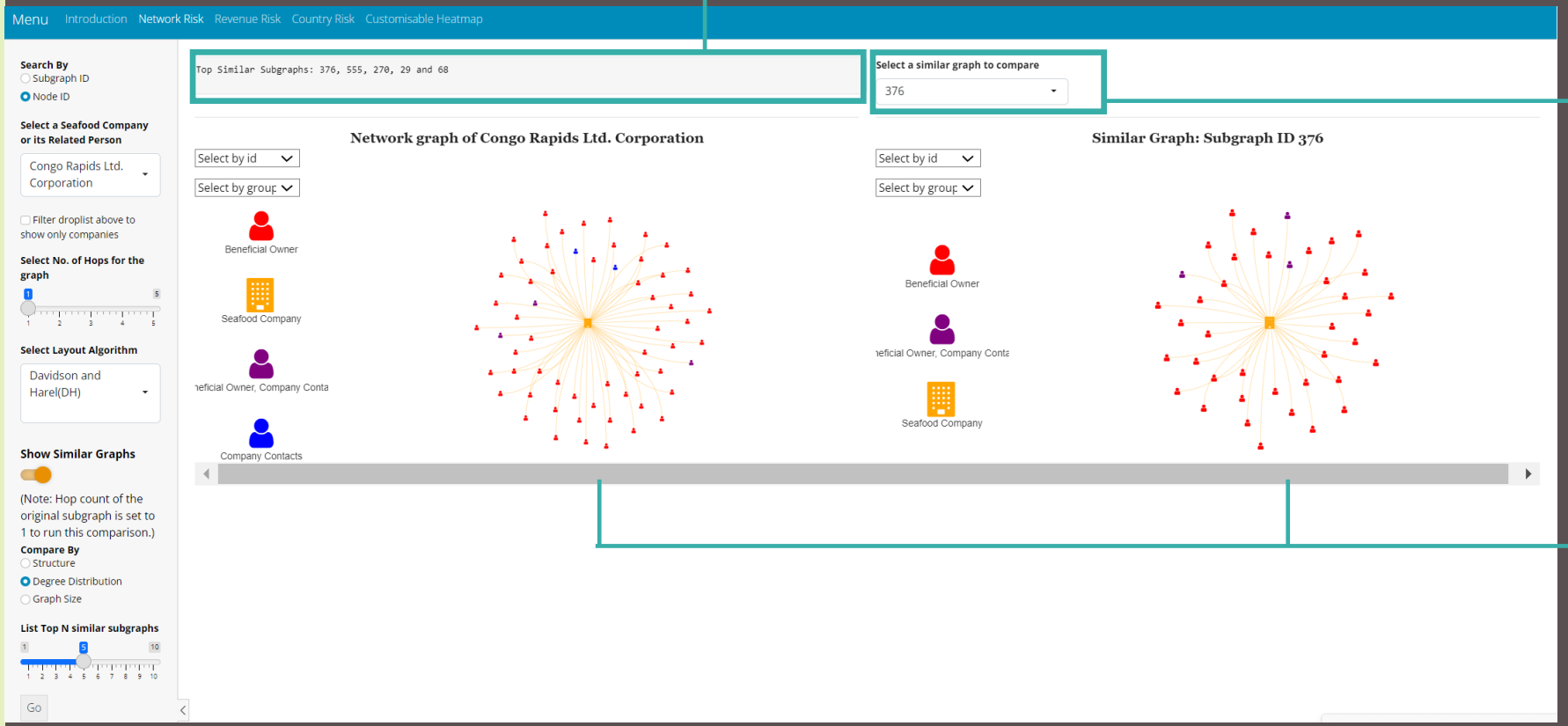
The Comparison Methods

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

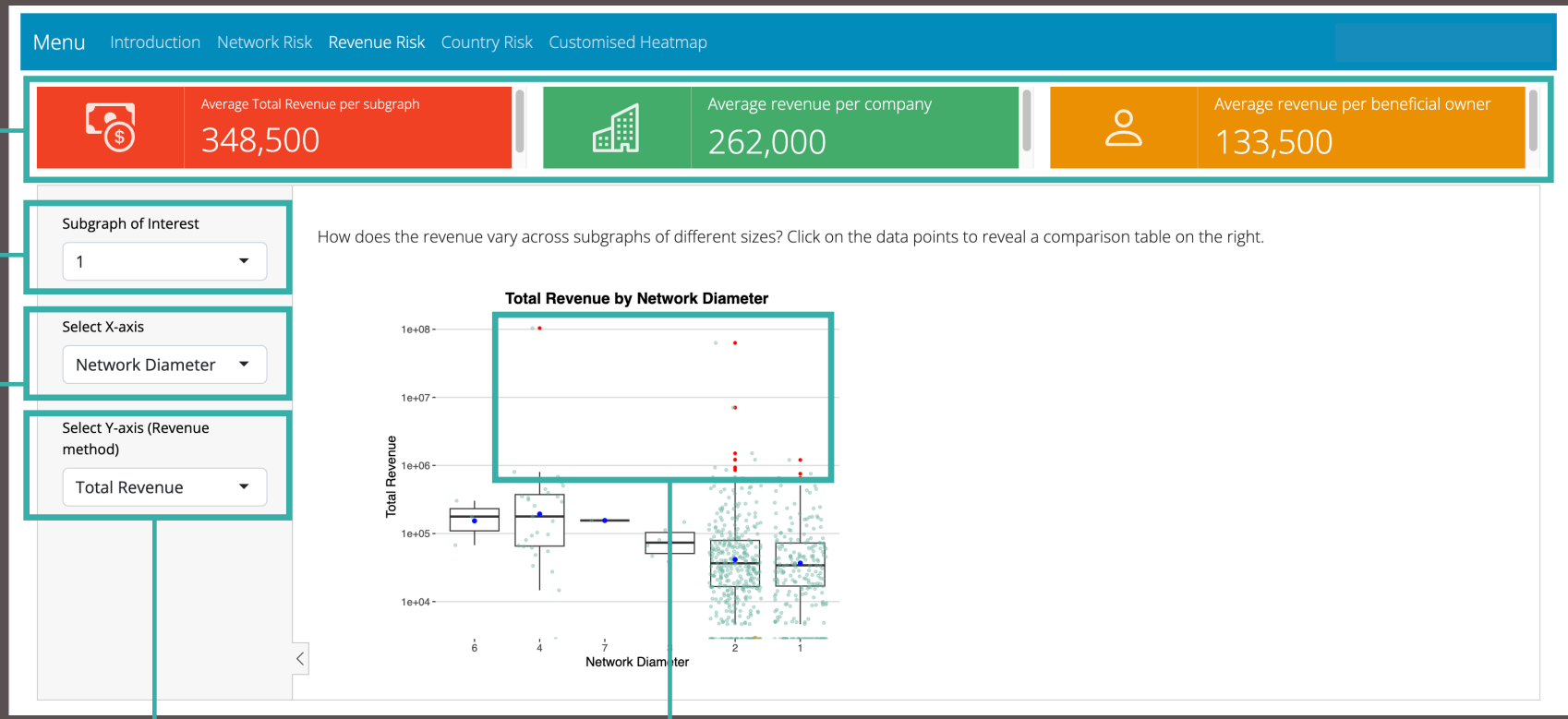
Financial Risk – Basic Features

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)

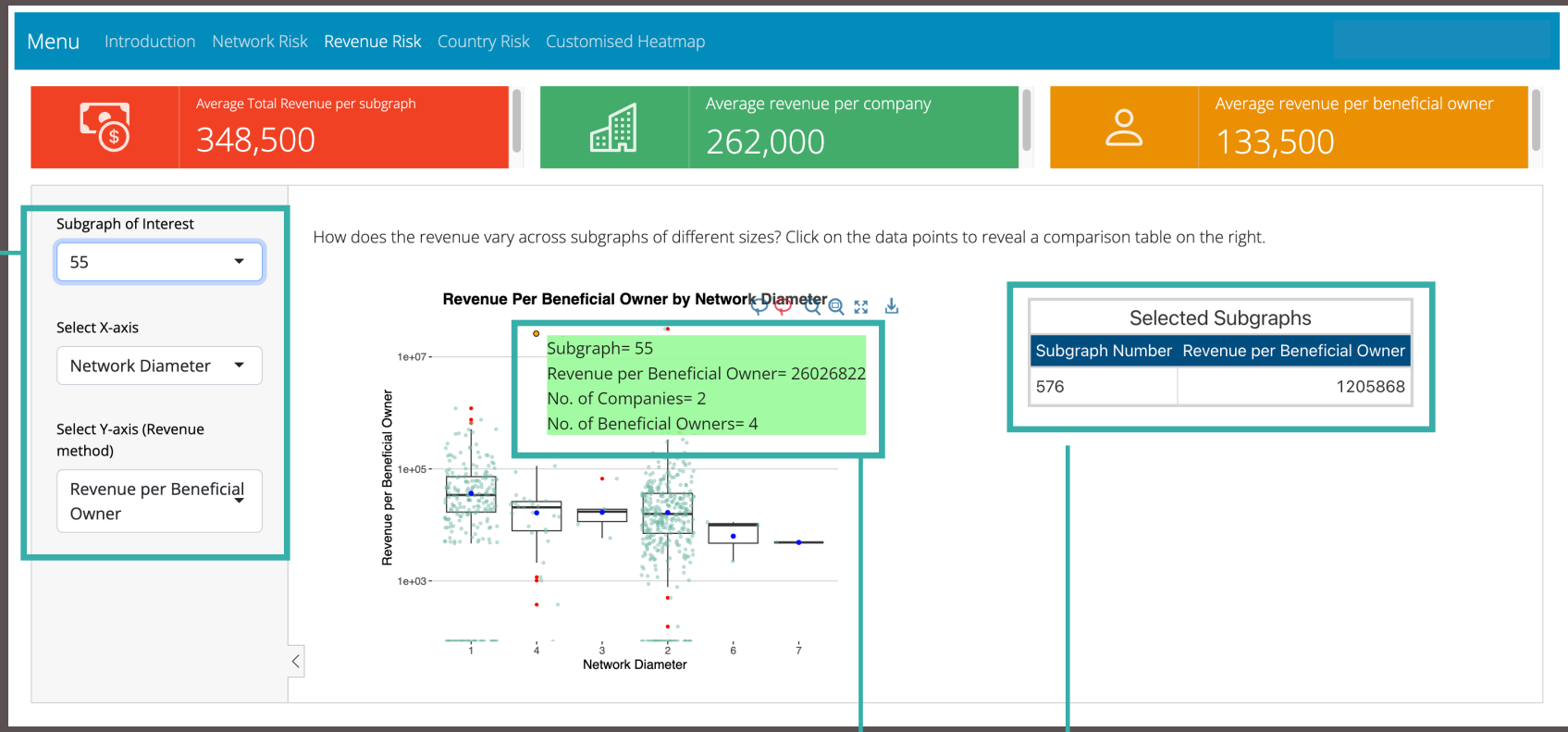
Try this!
Checkout Subgraph 55. Does the revenue look “normal” to you?



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

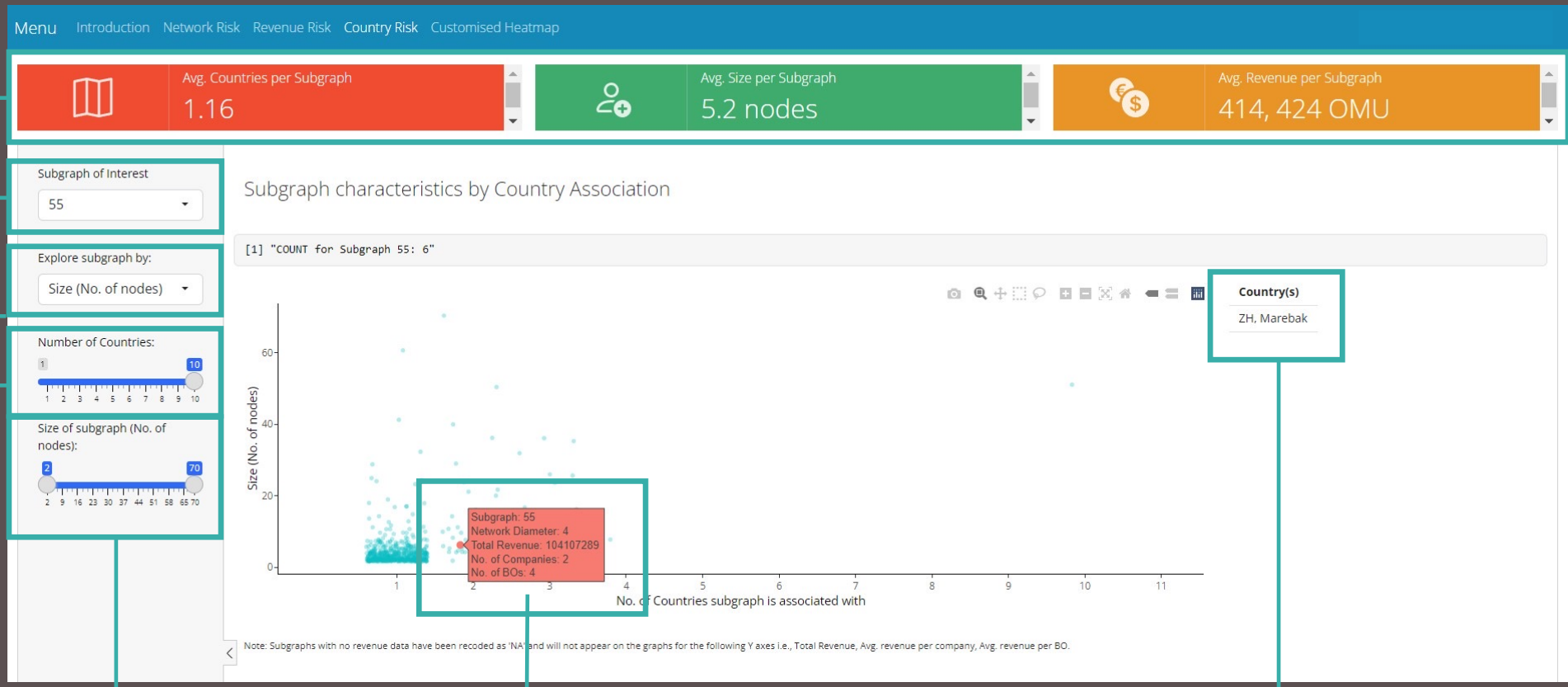
Try this!
Check out Subgraph 29. How is it special?

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 x-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

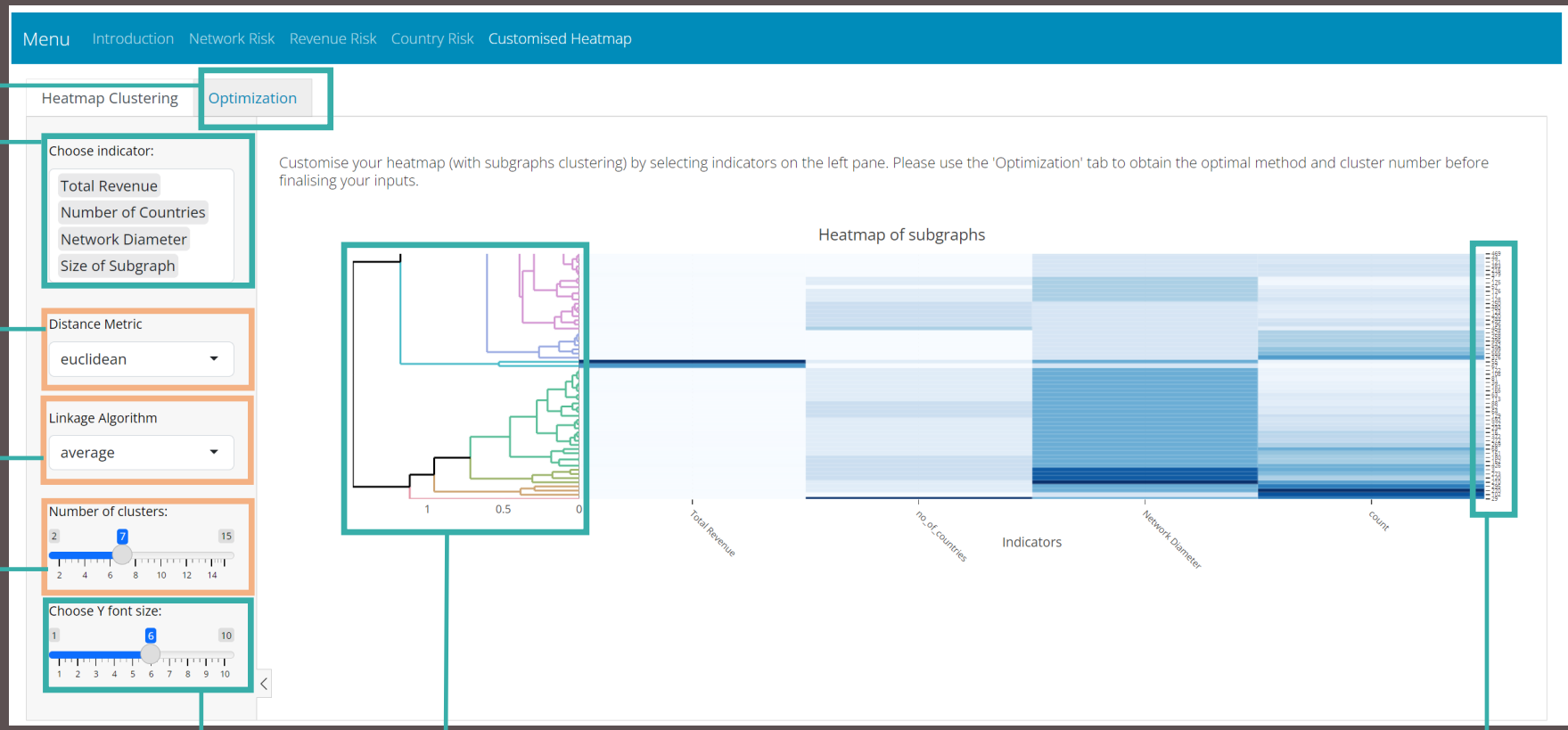
Customisable Heatmap - Clustering

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

Subgraph IDs

Try this!

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

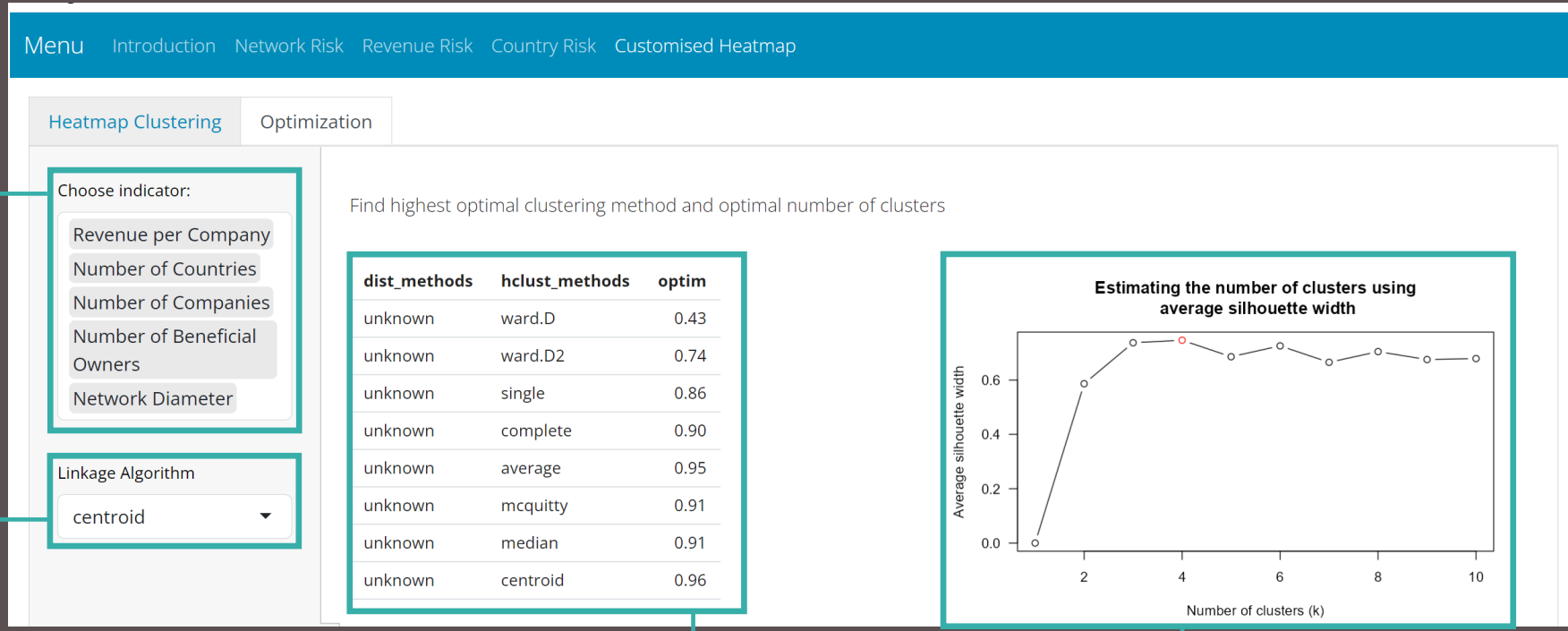
Customisable Heatmap - Optimisation

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'



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

Step 4:
Optimal number of clusters = 4