Network Visualization Project Specification

Visualization Project \ Specification - August 1, 2024

Section 1: Input Data

1. Available as a static CSV file attached here.
2. Filename: "Links-Contractor-300-073124-ds.csv"
3. 📎   Untitled Attachment
4.

Questions on input data

1. Some of the values in the input file are "dummy" values that will be updated over time. There will be more observations (rows) added as well. For the web hosted solution will you provide a location/approach for file updating?
2. Will the file live in a cloud SQL database where new versions of the file can be uploaded? Or is there another approach you propose?

Section 2: Visualization Type

1. **Network Graph:** https://d3-graph-gallery.com/network.html
   a. Reference Graph: https://observablehq.com/@d3/force-directed-graph-component
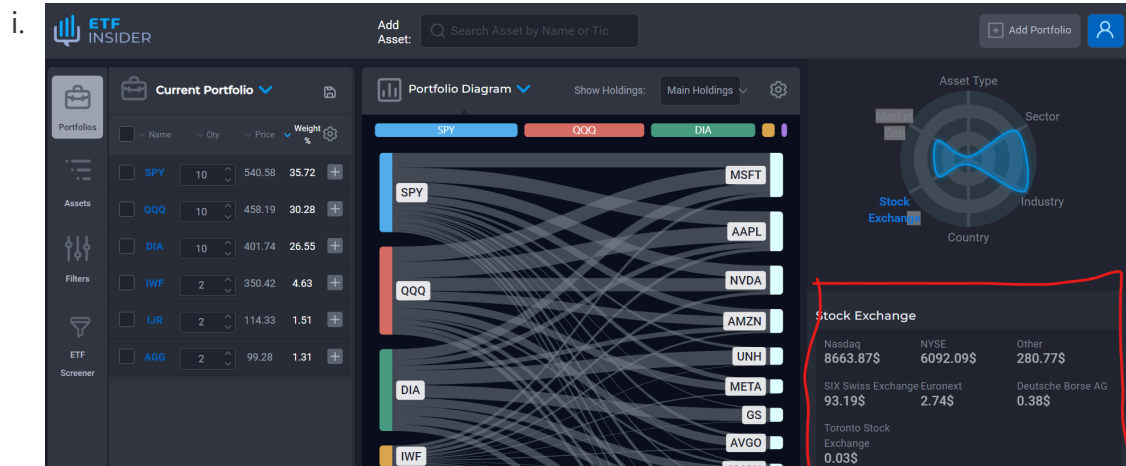
Section 3: Phase 1 Requirement

Default (Preset) Network Graph Variables: Below are the default network visualization rendering settings user would be on page load.

1. Primary Clustering Variable: "Social Determinant Category" (Column B)
2. Node Link Categories:
   a. Primary Link [render as a thicker line element versus others]: "Primary Determinant Analyzed" (COL C)
   b. Secondary Links [render as thinner line element versus a. above]: "Secondary Determinant Analyzed 1", "Secondary Determinant Analyzed 2", and "Secondary Determinant Analyzed 3"
3. Node Size: Cited By: (Column V)
   a. Allow user to select between:
      i. Cites: (Column U)
      ii. Cited By: (Column V)

iii. Research Method Population Scale (Column X)

iv. Harm Magnitude (Column Y)

v. Harm Population Impact (Column Z)

4. **Node Metadata Guidance**

   a. I am open for this to be a modal overlay when a user scrolls over a node OR displayed in the left rail (UI column). I am open to using the "UI shell" of this demo site: https://etfinsider.co/app/. The values for each of the variables below could be displayed on the right rail as show below:

     i.

     ii.

   b. For each node, please show the following fields

   c. List Heading: "Research Metadata"

     i. "Title" (Column G)

     ii. "Year Published" (Column H)

     iii. "Primary Author" (Column J)

     iv. "Number of Works Cited" (Column U)

     v. "Number of External Citations" (Column V)

     vi. "Author Total Works Published" (Column M)

     vii. "Research Location" (Column T)

     viii. "Research Type" (Column N)

     ix. "Research Method" (Column W)

     x. "Research Method Population Size" (Column X)

   d. List Heading: Racial Harm Metadata"

     i. "Social Determinant Category" (Column B)

     ii. Primary Determinant Analyzed" (Column C)

     iii. "Secondary Determinant Analyzed 1" (Column D)

     iv. "Secondary Determinant Analyzed 2" (Column E)

v. "Secondary Determinant Analyzed 3" (Column F)
vi. "Aligned SDG" (Column S)
vii. "Harm Magnitude"
viii. "Harm Population Impact"

Section 3: Phase 2 Specification

This would be an incremental phase after the above "base" visualization is complete.

This phase would include building filtering capabilities in a web UI. I am very open to using the "shell" of the visualization tool here: https://etfinsider.co/app/.

In the UI, the user can select different variable values for the following network characteristics (using the network graph outlined above):

1. **Primary Clustering Variable (Category):**
   a. Primary Determinant Analyzed" (Column C)
   b. "Secondary Determinant Analyzed"
      i. Inclusive of any unique value within Columns D, E and F ["Secondary Determinant Analyzed 1", "Secondary Determinant Analyzed 2", and "Secondary Determinant Analyzed 3"]
   c. "Research Method" (Column W)
   d. "Research Method Population Size" (Column X)
   e. "Harm Magnitude" (Column Y)
   f. "Harm Population Impacts" (Column Z)

2. **Node Size Variable:**
   a. "Number of Works Cited" (Column U)
   b. "Number of External Citations" (Column V)
   c. "Author Total Works Published" (Column M)
   d. "Year Published" (Column H)
   e. "Research Method Population Size" (Column X)
   f. "Harm Magnitude" (Column Y)
   g. "Harm Population Impacts" (Column Z)

3. **Node Links**
      For the options available to the user in the left sidebar, do not include option for user to select between "primary" or secondary "links"
   a. "Secondary Determinant Analyzed"

       i. Inclusive of any unique value within Columns D, E and F ["Secondary Determinant Analyzed 1", "Secondary Determinant Analyzed 2", and "Secondary Determinant Analyzed 3"]
  b. "Research Method Population Size" (Column X)
  c. "Harm Magnitude" (Column Y)
  d. "Harm Population Impacts" (Column Z)
  e. "Research Method" (Column W)

Section 4: Other Questions & Notes

1. **Force Direction Settings**: I'm open to using recommended / default settings
2. Can we use the variables above to configure a Sankey chart? (https://gist.github.com/d3noob/06e72deea99e7b4859841f305f63ba85)
3. Can we use the guidnace and varialbes above to configure as a Chord Chart (https://d3-graph-gallery.com/graph/chord_colors.html)