VAST 2023 MC3 Data Notes

Data dictionary and nodes for MC3:

Graph Description: The challenge problem KG is generated using

- 27,622 nodes
- 24,038 edges
- 7,794 connected components
- Possible node types include: {company and person}
- Possible node sub types include: {beneficial owner, company contacts}
- Possible edge types include: {person}
- Possible edge sub types include: {beneficial owner, company contacts}
- This is an undirected multi-graph
- The graph format is a json format intended to match d3's node-link format and be compatible with networkx.node_link_graph. At the root-level, it is a dictionary with graph-level properties specified as keys (directed, mulitgraph, graph). The nodes and links keys each provide a dictionary of the nodes and links respectively. The nodes entries that must include an id key that is unique for each node. The links entries include source and target keys that refer to node id values. All other keys provided in node and link dictionaries are attributes for that node or link.

Node Attributes:

- type – Type of node as defined above.
- country – Country associated with the entity. This can be a full country or a two-letter country code.
- product_services – Description of product services that the “id” node does.
- revenue_omu – Operating revenue of the “id” node in Oceanus Monetary Units.
- id – Identifier of the node is also the name of the entry.
- role – The subset of the “type” node, not in every node attribute.
- dataset – Always “MC3”.

Edge Attributes:

- type – Type of the edge as defined above.
- source – ID of the source node.
- target – ID of the target node.
- dataset – Always “MC3”.
- role - The subset of the “type” node, not in every edge attribute.