VAST 2023 Data Notes

Data dictionary and nodes:

Main Graph:

- 34552 nodes
- 5464092 directed edges (from shipper to receiver)
- This is a directed multi-graph, so multiple edges between the same two nodes are possible
- 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. A simple graph in this format is:

```
{'directed': True,
 'mulitgraph': True,
 'graph': {},
 'nodes': [{'id': 1, 'dataset': 'MC2'},
                  {'id': 2, 'dataset': 'MC2'}],
 'links': [{'source': 1,
            'target': 2,
            'datset': 'MC2'}]
}
```

Node Attributes:

- id -- Name of the company that originated (or received) the shipment
- shpcountry -- Country the company most often associated with when shipping
- rcvcountry -- Country the company most often associated with when receiving
- dataset -- Always 'MC2'

Edge Attributes:

- arrivaldate -- Date the shipment arrived at port in YYYY-MM-DD format.

- hscode -- Harmonized System code for the shipment. Can be joined with the hscodes table to get additional details.
- valueofgoods_omu -- Customs-declared value of the total shipment, in Oceanus Monetary Units (OMU)
- volumeteu -- The volume of the shipment in 'Twenty-foot equivalent units', roughly how many 20-foot standard containers would be required. (Actual number of containers may have been different as there are 20ft and 40ft standard containers and tankers that do not use containers)
- weightkg -- The weight of the shipment in kilograms (if known)
- dataset -- Always 'MC2'
- type -- Always 'shipment' for MC2
- generated_by -- Name of the program that generated the edge. (Only found on 'bundle' records.)

Note: Some data provided by Oceanus was anonymized leading to some shipper and receiver names/countries being omitted. These are represented by numerical names in the graph.

Bundles

Each bundle represents the output of an AI program for link inference. Each bundle represents a list of potential edges to add to the main graph. The challenge is to decide WHICH of these bundles should be used (and present evidence for why the selected bundles should be used).

The bundles are:

- carp
- catfish
- chub_mackerel
- cod2
- herring
- lichen
- mackerel
- pollock
- salmon
- salmon_wgl
- shark
- tuna