Data Science Take Home Exercise

Graph Exploration
Expected Duration: 2 hours

Background

Graph

A graph is a set of objects (nodes) and links between those objects (edges)that can be used to describe relationships between things and ideas.

Neighborhood

The structure of any graph depends on the relationship between its underlying nodes. "Neighborhoods" exist when collections of nodes have higher interdependence amongst themselves than they do with other nodes.

Exercise

Summary

You have been provided a graph of related topics in the entertainment industry, where nodes can represent things like actors, characters, shows, movies, etc. Natural neighborhoods exist in this graph as certain actors/directors frequently work together (Wes Anderson/Bill Murray) and characters frequently appear together in TV shows and movie franchises. **You are tasked with identifying the best N neighborhoods in the graph**. All other considerations (what N is, how neighborhoods are identified, exhaustiveness, overlap, density/cohesion, etc.) are up to you.

We encourage you not to take longer than two hours working on this exercise. While no time limit will actually be enforced, we recognize the value of your time and the stress that overly-invasive take home exercises can bring.

Data

You have been provided the graph data in JSON format. The graph is directed, so each edge/link in the data has a direction flowing from the source node to the target node.

Nodes

Each node contains the following properties:
- **name**: A description of the node.
- **prevalence**: The relative prevalence of the node in the underlying inventory.
- **x**: A projected x-coordinate from a force-directed simulation in a Cartesian space.
- **y**: A projected y-coordinate from a force-directed simulation in a Cartesian space.

Links

Each link contains the following properties:
- **source**: Index of the source node in the node list.
- **target**: Index of the target/destination node in the node list.
- **value**: The weight of the edge connecting the source and target nodes.