Fund2Vec: Mutual Funds Similarity Using Graph Machine Learning

3/8/21 4:25 PM
Speaker: Dhagash Mehta (The Vanguard Group)
Date: 03/04/2021

Introduction: Product Similarity

Q1: Given a product, what are other similar products?
Q2: How similar are the given two products?
Product Similarity is known as "peer analysis', 'competitors analysis', etc.

Multiple applications

- Sales and marketing: knowing a customer has a competitor's fund, proactively convince the customer to switch to a similar home-grown product
- Alternative portfolio construction: for a given portfolio of funds consisting of competitors' funds, construct an alternative portfolio with the same risk-return profile but consists of only home-grown funds
- Portfolio diversification: two or more similar funds in a portfolio may unintentionally reduce diversification
- Similar fund with different theme
- Competitors' analysis
- Tax loss harvesting: move from one fund to another similar one for tax-loss harvesting
- Launching new products: launch a new fund similar to one popular in specific markets

Current approaches and problems

- Third-party categorization: e.g. Morningstar/Lipper categories
  - known to partly rely on qualitative approach, partially a black-box and sometimes irreproducible process. More importantly, no ranking
- Compute the overlap between two portfolios (with the Jaccard index, weighted Jaccard index, etc.)
  - Captures the bigger picture but need to be careful if granular details are needed
- Compute the Euclidean distance between pairs of portfolios in the chosen variables-space
  - Captures linear relationship
- Compute the cosine similarities between vectors corresponding to different portfolios in the chosen variables-space
  - Captures linear relationship
- Many other unsupervised machine learning techniques such as unsupervised clustering
  - Usually linear relationship, or doesn't scale well

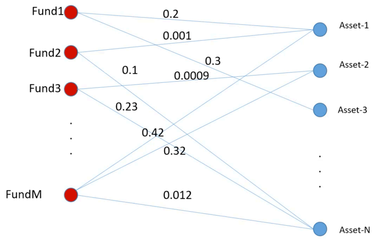
Our idea

reformulate the data of mutual funds and assets as a network, use a graph neural network to identify the embedded representation of the data, and compute similarity in the learned lower dimensional representation

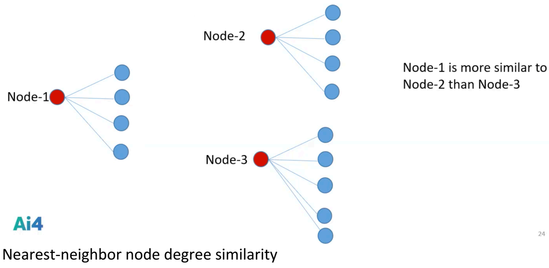
Funds-Assets Network

Funds/Assets	Asset-1	Asset-2	Asset-3	Asset-4	...
Fund-1	0.2	0	0.3	0	...
Fund-2	0.001	0	0	0.1	...
Fund-3	0	0.0009	0	0.23	...
...	###	###	###	###	###

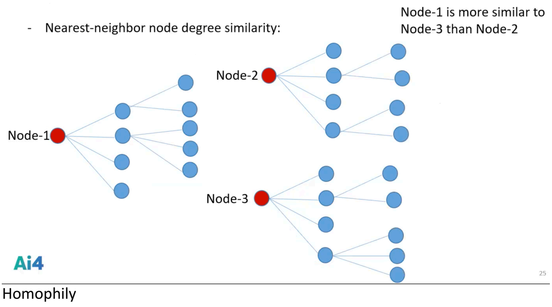
- Entries are weights of asset in given fund
- Turn into a bipartite network (two distinct types of nodes):



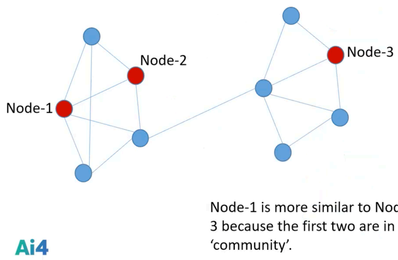
- Now the problem is transformed to finding similar nodes on the network
- There are many ways nodes can be similar to one another on a network
  - Node degree similarity



  - Nearest-neighbor node degree similarity



  - Homophily



Product Similarity: Node2Vec

Node2Vec follows multiple random walks from all the nodes, while interpolating between breath-first search and depth-first search using different hyperparameters