

Assignment 1

Instructions

Please remember the good scholarly practice requirements of the University regarding work for credit. You can find guidance at the School page

<https://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct>

This also has links to the relevant University pages.

You are **not** allowed to collaborate with other students on this assignment or to ask or answer questions about the contents of the assignment. If you do not understand a specific question, ask Valerio and Ogy on Piazza or in person.

Submitting Instructions

For this coursework, you will be asked to analyse a network of investors and stock, using network analysis techniques. All the analysis must be done Python. You can either use a notebook or an IDE of your choice, but you must submit all your code in a **single zip file** (even if it is just one Jupyter Notebook file, this must be zipped). You also must submit a **pdf** (very important) in which you report the results and discuss them. Submission will be available on **Learn**.

General Instructions

In this coursework, you will analyse a real-world temporal network based on what you have learned in class. You will also have a dataset with stock prices required only for task 6. Many exercises will require you to discuss the results of your analysis, some other will leave you the choice of which algorithm to use for a particular task. This is by design because this coursework assesses whether you understand network science and whether you can apply it to real-world networks. For this reason, if you realise you need to make assumptions to answer a question, do so and always, always motivate your assumptions and answers!

Warning: Some network metrics might require some time to compute. Please consider this when doing the coursework and allow enough time to perform the required computations. Also remember that you can use the School's DICE machines, which can be let to run!

Introduction

As the newest data analyst at DBBA Capital, your first project centers on analysing relationships between major stocks in the market, with a specific focus on Microsoft (MSFT). Your mission is to investigate how investor behaviour forms connections between MSFT and other significant stocks through the lens of shared investors.

You will analyse the quarterly stockholding patterns of different investors from Q1 2016 to Q2 2023 to understand how shifts in investor sentiment and strategy influence the movement and relationships involving MSFT. For example, do investors who hold MSFT also frequently invest in WFC or BX? What patterns emerge in how investors allocate capital across sectors, particularly between technology and financial services, and how does MSFT fit into this landscape?

This analysis will focus on a broader temporal dataset covering various stocks across different sectors, but MSFT will be the focal point. You will track how changes in investor portfolios during critical periods, such as the pandemic, impact MSFT's performance and its connection with other stocks. By performing this analysis, you can help DBBA Capital identify trends and potential influences on MSFT that may inform smarter, more strategic investment decisions for the future.

Task 1 – Constructing a Bipartite Network

Using the provided temporal datasets spanning multiple time periods from 2016 to 2023, construct a bipartite network that connects all unique investors (across all time periods) on one side with all unique investments (stocks) on the other. This network will incorporate data from all available temporal datasets (not stock prices), giving us a comprehensive representation of investor behavior over time.

In this network, the weight of each edge between an investor and a stock will indicate the total number of different time periods in which the investor has made investments in that particular stock. Therefore, if an investor has invested in the same company across multiple periods, this will be reflected in the edge weight, allowing us to understand the extent of their investment activity.

Task 1.1 – Visualising the Network (3 marks)

Plot the resulting weighted bipartite network in a legible manner, ensuring that both investors and investments are clearly visible and distinguishable. Include this plot in your report, along with a brief description of how you created the network and what it represents.

Task 1.2 – Bipartite Ego Network (3 marks)

DBBA Capital is particularly interested in Microsoft (MSFT) as a potential investment. To better understand the landscape of investors in this stock, create and plot the ego network of MSFT. Plot the ego network of MSFT, ensuring that all investors connected to MSFT are clearly labeled. Provide a brief explanation of the investors involved and how often they have invested in MSFT.

Task 2 – Network Projections

Task 2.1 – Creating a Stock-Side Network Projection (5 marks)

Next, you need to create a *network projection* on the stock side of the bipartite network.

A network projection refers to a transformation process used in bipartite networks (or two-mode networks) to create a simpler, one-mode network. In a simple bipartite network, there are two types of nodes, and connections only exist between nodes of different types. A network projection reduces this bipartite structure to focus on one type of node, connecting them based on shared associations with the second type of node. You can find more information on network projections and the respective coding ¹.

You need to create a projection on the stock side of the network, where nodes represent stocks and edges indicate that two stocks share at least one common investor. The weight of the edge between any two stocks should reflect the number of shared investors.

After creating this projection, plot the network and discuss the connections between stocks, paying special attention to which stocks are most closely linked through shared investors. In your report, explain what these connections reveal about investor behaviour and market trends, and how DBBA Capital might use this information to inform investment decisions.

Task 2.2 – Network Comparison (6 marks)

Discuss the key differences between the bipartite network and the projection network created in the previous tasks. Explain the importance of each of these two network types in understanding investment patterns. Discuss how insights gained from both network types can inform strategic investment decisions for DBBA Capital.

Task 2.3 – Ego Network Comparison (4 marks)

Using the stock-side network projection from Task 2.1, create and plot the unipartite ego network for Microsoft (MSFT). Include the plot in your report. Discuss the difference between the ego network in the bipartite vs unipartite case. Does it have more or less neighbours than the bipartite ego network? Why? Are they strongly connected? What do the neighbours represent?

Task 3 – Basic Network Analysis**Task 3.1 – Network Statistics (14 marks)**

Now, that you know how to build the network projection on the stock side, create a separate projection for each quarter (each data file). For both the whole network projection and the projection ego-network for MSFT, produce a table with the summary statistics (i.e. mean, max, min, and standard deviation) of the following network quantities:

- Number of nodes
- Number of links
- Density

¹Opsahl, T. (n.d.). Two-mode networks: Projections. Retrieved October 2, 2024, from <https://toreopsahl.com/tnet/two-mode-networks/projection/>

- Average clustering coefficient
- Average degrees
- Average strength
- Assortativity

If you need to make any assumption or decision regarding the metric to use to compute any of these quantities, clearly motivate it.

Task 3.2 – Discussion (10 marks)

Comment on the statistics you computed above and what information they give you about the network of stocks and the MSFT stock specifically. Briefly discuss how the ego network statistics differ from the statistics obtained for the whole network, explaining whether the differences or similarities are expected or not. Motivate your answers.

Task 4 – Changes of the network statistics during the pandemic

Task 4.1 – Temporal Evolution of Statistics (10 marks)

Plot the temporal evolution of the quantities you computed in Task 3.1 for the ego network and the whole network and compare the difference between the networks. For each quantity, discuss if it can be used for analysing the importance of the MSFT stock over time. Based on your discussion, choose the quantities that you find important. What information you can draw about the change of those network statistics during the pandemic?

Task 4.2 – Centrality (10 marks)

Choose a suitable centrality measure that would give us important information about the nodes in the whole network, and clearly motivate your choice. Use this measure to find the 3 most central nodes for each quarter. Compare the centrality of MSFT overtime with that of the most central nodes. What can you conclude from this?

Task 5 – Clustering and Modularity (10 marks)

Find the communities in each quarter in the stock-side network projection. To do so, use an algorithm of your choice, and justify your decision. Analyse how the communities evolve overtime, focussing on the membership of MSFT. Does this node fall in the same community with the same stocks across different quarters? What conclusions can you draw from this?

Task 6 – Analysing Stock Correlation Through a Spanning Tree

In this task, you will analyze the relationships between stocks based on their daily closing prices for the past six years and a half focusing on Microsoft (MSFT) and 50 other assets. A stock correlation network is a graphical representation where stocks are represented as nodes, and the correlations between them serve as edges, indicating the strength of their relationships.

Your task is divided into three parts:

Task 6.1 – Construct a Stock Correlation Network (2 marks)

Use the provided closing stock prices dataset to construct a stock correlation network, where each stock is a node, and the edges between them are weighted based on the correlations between the monthly returns of each pair of stocks.

Task 6.2 – Build a Spanning Tree (5 marks)

Using the correlation network you constructed, build a spanning tree that represents the relationships among the stocks. Select an appropriate measure to define the link weights between stocks and justify your choice.

Task 6.3 – Analyse MSFT’s Role in the Spanning Tree (8 marks)

Analyse Microsoft’s position within the spanning tree by discussing its importance in terms of centrality and its connections to other stocks. How does Microsoft’s role in the spanning tree compare to its role in the stock-side network projection from earlier tasks?

Investigate the behaviour of the spanning tree around the time of the pandemic. Describe changes in Microsoft’s role in the spanning tree before (March, 2020), during (between March, 2020 - March, 2021), and after (March, 2021) the pandemic period. Comment on observations such as changes in connections and centrality. Use quantitative metrics whenever possible to support your discussion.

Task 7 – Result Discussion (10 marks)

Write a concise report of approximately 250 words summarizing your observations from this assignment. In your report, evaluate whether you recommend DBBA Capital to invest in Microsoft (MSFT). Justify your recommendation by referencing results from your analyses conducted throughout this assignment, as well as any relevant external sources (please provide citations for any external references used).