Coursework

Instructions

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

informatics.ed.ac.uk/taught-students/all-students/your-studies/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/or Yanpei (Andy) on Piazza or in person.

Submitting Instructions

For this coursework, you will be asked to analyse a network of investors and stocks, using network analysis techniques. All the analysis must be done in 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. You also must submit a pdf (in the zip)** in which you report the results (include all the required figures and tables) and discuss them. Include all the relevant discussion of results in the pdf report. Your final submission should consist of your pdf report and the supporting code. 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 both 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 completing 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

You have been appointed as the new data analyst at FINET Capital! Your initial project focuses on exploring relationships between major stocks and examining how investors' holdings create connections between them. FINET Capital is specifically interested in the performance of Amazon (AMZN), as it is a leading company in both the technology and retail sectors, and its trading activity often reflects broader market trends. Your tasks will include constructing networks of shared investors, analysing the structure of these networks, and comparing AMZN's connections to other significant stocks to discover patterns in the market and investment strategies.

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 AMZN. For example, do investors who hold AMZN 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 AMZN fit into this landscape?

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

Task 1 – Investor - Stock Bipartite Network

Task 1.1 – Constructing a Bipartite Network (3 marks)

Using the provided temporal datasets spanning a time period from Q1 2016 to Q2 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 1.2 – Visualising the Network (3 marks)

Plot the resulting weighted bipartite network in a legible manner, making sure that both investors and stocks 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 2 – Network Projections and Filtering

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

Construct a projection of the bipartite network you created in Task 1 onto the stock side, where each node represents a stock and an edge is created between two stocks if they share at least one common investor. You should select a projection algorithm that you consider most appropriate and clearly justify your choice with reference to the data and relevant literature. The weight assigned to each edge must be determined by the chosen projection method, and you should explain why this method is suitable for capturing relationships between stocks. After generating the projected network, you should report the number of nodes, the number of edges, the average strength, the density, and the average path length in the projected network.

Task 2.2 – Network Filtering (6 marks)

Projected networks often contain many weak or noisy connections that obscure the underlying structure. Therefore, you should apply a filtering algorithm of your choice to the projected stock network in order to identify its most important edges. You should justify the choice of algorithm by referring to relevant literature and explain why it is suitable for the type of data you are working with.

Specify and motivate the choice of the appropriate parameters for the algorithm of your choice. After obtaining the filtered network, report the number of nodes, the number of edges, the average strength, the density, and the average path length of the filtered network. Compare these metrics with those of the projected network from Task 2.1, discussing how they changed and whether the changes were what you expected. Identify which metric exhibited the largest change and explain what this indicates about how the structure of the network was affected by the filtering process.

Task 2.3 – Bipartite vs. Unipartite Network Comparison (4 marks)

You should create a clear and easy-to-read visualisation of the filtered stock network, and include it in your report. The layout, choice of colors, and representation of edge weights should be carefully designed to make sure that the most important structures and relationships are visible.

After visualising the network, discuss the key differences between the original bipartite network you created in Taks 1 and the projected, filtered network you just visualised. 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 FINET Capital.

Task 3 – Basic Network Analysis

Task 3.1 – Network Statistics (10 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). You need to choose a suitable projection algorithm and justify your choice. You do not need to filter these individual projections per quarter. Using the stock-side network projection create an ego network for Amazon (AMZN) with degree of separation of one.

Calculate the following summary statistics for each network across each quarter:

- Number of nodes
- Number of links
- Density
- Average clustering coefficient
- Average degrees
- Average strength
- Assortativity

Perform this separately for the full stock projection network and for the AMZN ego network with a degree of separation of one. Present the results in two tables: one showing the summary statistics for the full stock projection network and another showing the summary statistics for the AMZN ego network. Each row of these tables should correspond to a network quantity (i.e. you need 7 rows) and the columns should display the mean, maximum, minimum, and standard deviation of this metric across all quarters (i.e. you need 4 columns).

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 – Full Network Statistics Interpretation (5 marks)

Comment on the summary statistics you computed for the full stock network projection in Task 3.1. Explain what these statistics reveal about the overall connectivity, clustering, and distribution of relationships between stocks. Discuss what the mean, maximum, minimum, and standard deviation of the network quantities indicate about general patterns of co-investment and investor behaviour across all stocks. Highlight any particularly central or highly connected stocks and how the network structure reflects overall market interactions.

Task 3.3 – AMZN Ego Network Interpretation (5 marks)

Comment on the summary statistics you computed for the AMZN ego network with degree of separation one. Compare these statistics to those of the full network and explain what they reveal about Amazon's position and importance in the network. Discuss differences in clustering, degree, and strength relative to the overall network. Justify whether these differences or similarities are expected based on the structure of investor co-investments and Amazon's role in the market.

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 with degree of separation one 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 AMZN stock over time, especially during the pandemic period. Based on your discussion, choose the quantities that you find important. What information can you draw about the

change of those network statistics during the pandemic? Feel free to use additional literature or online resources for this discussion.

Task 4.2 – Centrality Measure and Most Central Nodes (5 marks)

Choose a suitable centrality measure that provides meaningful information about the importance or influence of nodes in the whole stock-side filtered network projection. Clearly justify your choice, explaining why it is appropriate for analysing co-investment patterns. Using this measure, compute the centrality values for all nodes in each quarter and identify the three most central stocks.

Task 4.3 – Comparison with AMZN (5 marks)

Compare the centrality of Amazon (AMZN) over time with that of the most central nodes identified in Task 4.2. Discuss whether AMZN is consistently central, increasing, or decreasing in importance relative to other top stocks. Draw conclusions about AMZN's role in the network and investor behaviour over the pandemic period.

Task 5 – Clustering and Modularity

Task 5.1 – Community Detection (5 marks)

Find the communities in each quarter within the stock-side filtered network projection. Choose a suitable community detection algorithm and clearly justify your selection, explaining why it is appropriate for analysing investor co-holdings. Compute the communities for all nodes in each quarter.

Task 5.2 – Evolution of Communities and AMZN Membership (5 marks)

Analyse how the communities evolve over time, with a particular focus on the membership of Amazon (AMZN). Determine whether AMZN remains in the same community with the same stocks across different quarters. Discuss the patterns you observe and draw conclusions about AMZN's relationships with other stocks. You may reference additional literature to support your discussion.

Task 6 – Analysing Stock Correlation Through a Spanning Tree

In this task, you will analyse the relationships between stocks based on their daily closing prices for six years and a half (Jan 2018 to Sep 2024) focusing on Amazon (AMZN) 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 four 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 AMZN's Role in the Spanning Tree (8 marks)

Analyse Amazon's position within the spanning tree by discussing its importance in terms of centrality and its connections to other stocks. How does Amazon'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 how Amazon's role in the spanning tree evolved across the pandemic, specifying the periods as follows: before the pandemic (1 January 2018 – 29 February 2020), during the pandemic (1 March 2020 – 28 February 2021), and after the pandemic (1 March 2021 – 30 September 2024). Comment on observations such as changes in connections and centrality. Use quantitative metrics whenever possible to support your discussion.

Task 6.4 – Investment Decisions Related to AMZN (4 marks)

Based on your analyses of the stock correlation network, spanning tree, and AMZN's role within these structures, identify which stocks you would recommend FINET Capital to invest in alongside Amazon and which stocks you would advise against. Justify your choices by considering the strength of correlations, network centrality, clustering patterns, and potential risk diversification benefits. Use quantitative evidence from your analyses to support your recommendations.

Task 7 – Result Discussion

Write a concise report of approximately 350–400 words summarising your observations from this assignment by answering the tasks below.

Task 7.1 – Summary of Observations (3 marks)

Summarise the main findings from your analyses, focusing on the temporal evolution of the stock network, the AMZN ego network, and any notable changes during the pandemic period. Highlight the key network statistics that were most informative.

Task 7.2 – Justification and External References (5 marks)

Justify your recommendation using evidence from the network statistics, centrality measures, community structure, or other analyses conducted in the assignment. Include relevant external sources, such as market reports or academic studies, with proper citations.

Task 7.3 – Investment Recommendation (2 marks)

Provide a clear recommendation on whether FINET Capital should invest in Amazon (AMZN), based on your analyses.