DBBA Coursework 1

s2714494

October 2024

Introduction

In this coursework, we will be taking the role of a data analyst at DBBA Capital, where we were given a project of analysing relationships between major stocks in the market, specifically focusing on Microsoft (MSFT). Our aim is to observe how investor behaviour affects connections between MSFT and other stocks through shared investors.

We will also observe MSFT's performance and its connection with other stocks through different periods of the COVID-19 pandemic. By performing this analysis, we aim to provide meaningful insights for DBBA Capital, enabling them to make strategic investment decisions for the future.

Task 1 – Constructing a Bipartite Network

A bipartite graph is a graph where the vertices can be divided into two disjoint sets such as that all edges connect a vertex in one set to a vertex in another set. There are no edges between vertices in the disjoint sets (Metcalf and Casey 2016). In our case, we construct the bipartite network to visualize the interactions between investors and the stocks they invested in, with one set representing investors and the other representing stocks. This bipartite network allows us to analyze investment patterns, where edges between investors and stocks represent investments, and their weights indicate the frequency of these investments across multiple time periods.

Task 1.1 – Visualizing the Network

Our dataset consists of quarterly investments made by different investors, ranging from 2016 to 2023. Each node in this network represents either an investor or a stock, and an edge indicates an investment relationship.

To construct the bipartite network, we went through these steps:

1. Data preprocessing: We used temporal datasets for each quarter from 2016 to 2023, where each dataset consists of investors and the stocks they invested in. For each dataset, we reshaped the data to clearly list each investor-stock pair. This allowed us to aggregate investor-stock relationships across multiple quarters.
2. Nodes definition: In our network, each node is either an investor or a stock.
3. Edges definition: In our network, an edge exists between an investor and a stock if the investor has invested in that stock in a specific quarter. Edge weights are determined by the number of distinct quarters in which the investment was made.
4. Graph creation: We created the bipartite network by defining that investors form one set of nodes, and stocks form the other. As stated above, each edge represents an investment, with weights representing the number of distinct quarters in which an investment was made.

Figure 1 shows the bipartite network that illustrates the investment relationships between investors and stocks. In this network, investors are positioned on the left, while stocks are arranged on the right side of the graph. Each edge connecting an investor to a stock represents an investment, with the color of the edge indicating the frequency of the investment: blue edges represent less frequent investments, while red edges highlight more frequent investments.

This visualization provides an overview of investment behavior over multiple time periods. By examining the network, we can identify which investors who demonstrate consistent investments over multiple time periods, indicating a long-term strategic investments and confidence in the stocks' performance. The weighted edges, based on the frequency of investments, allow us to identify the intensity of investor-stock relationship. At a glance, we can identify significant investor-stock relationships, indicated by the color red.

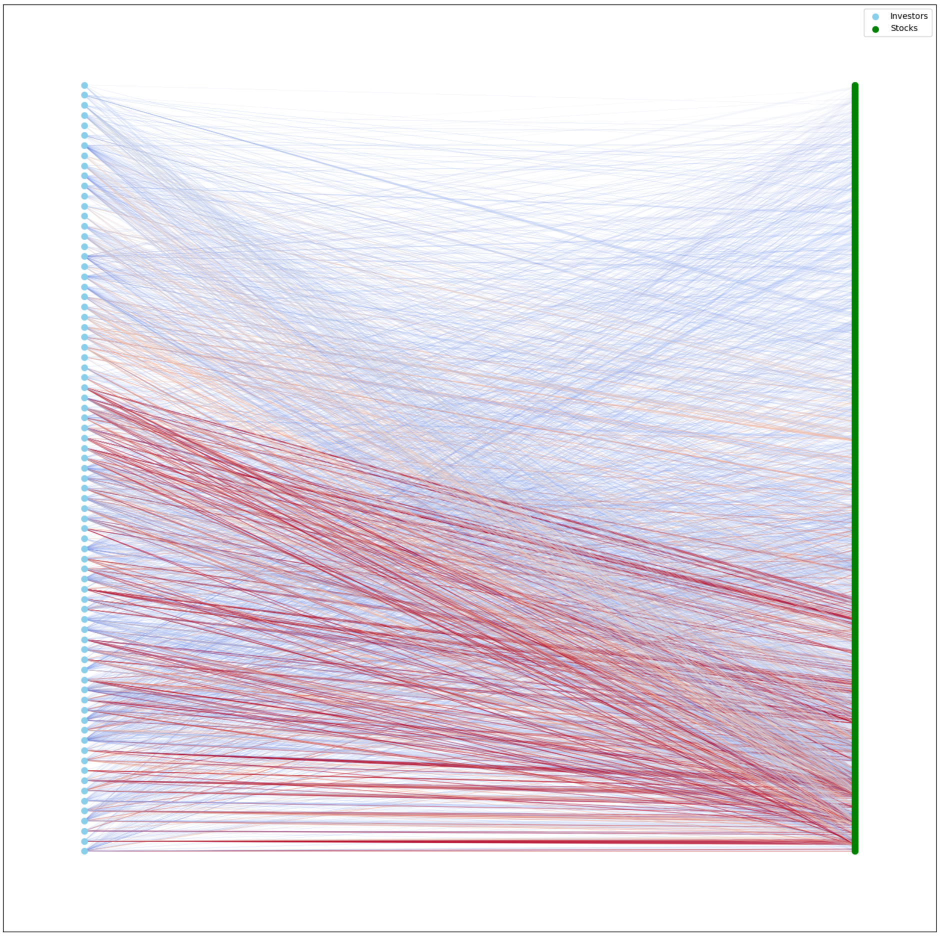


Figure 1: Visualization of the weighted bipartite network.

Given the large size of the network, interpreting investor behaviour directly is challenging. For better interpretability, alternative methods can be employed, such as grouping stocks by industry type, filtering the network to retain only the most significant relationships, or focus on an ego-network for a specific stock. For our next analysis, we will be focusing on Microsoft ('MSFT') by plotting its ego-network.

Task 1.2 – Bipartite Ego Network

Ego network is a special type of network built around a particular unit designated as ego, and includes all nodes directly connected to it (Freeman 1982). In this analysis, 'MSFT' acts as the ego node, with the investors directly linked to it as the alter nodes. By focusing on this ego network, we aim to identify investors who have a significant relationship with 'MSFT', potentially indicating consistent, long-term strategic investment in the company.

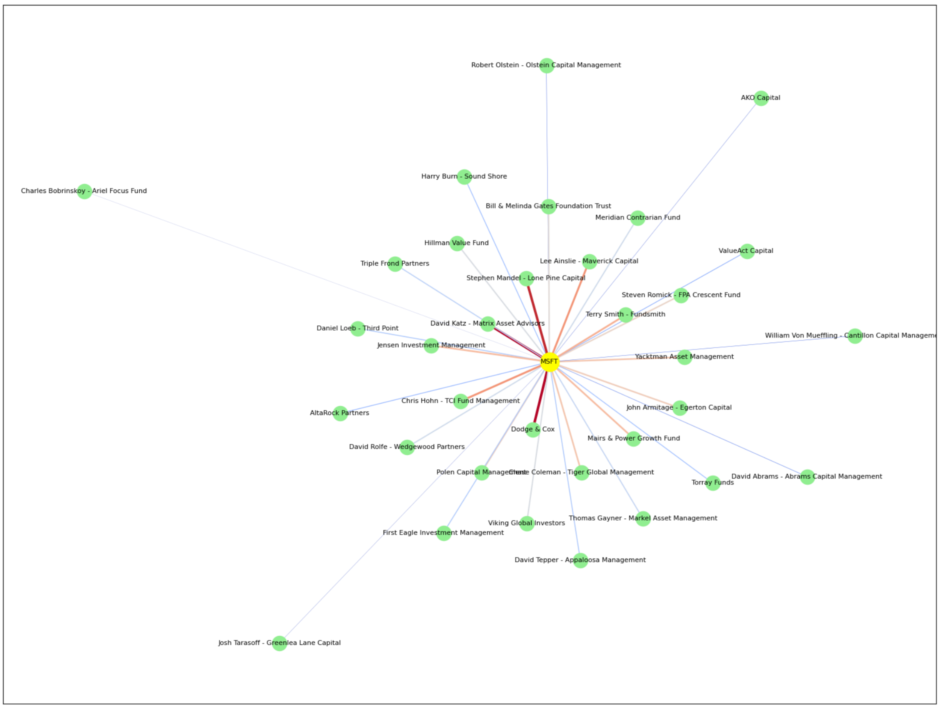
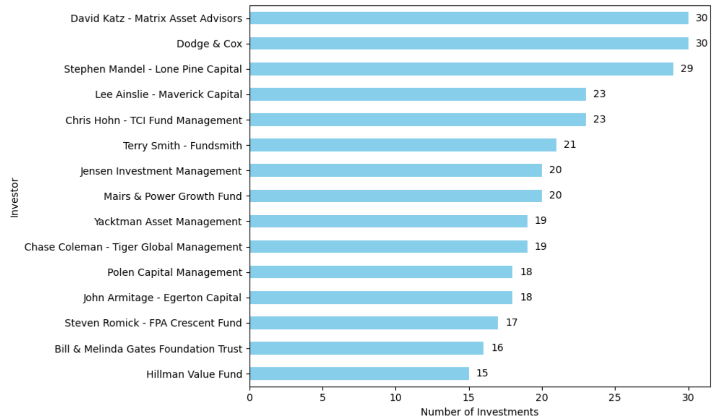


Figure 2: Visualization of the bipartite ego network.

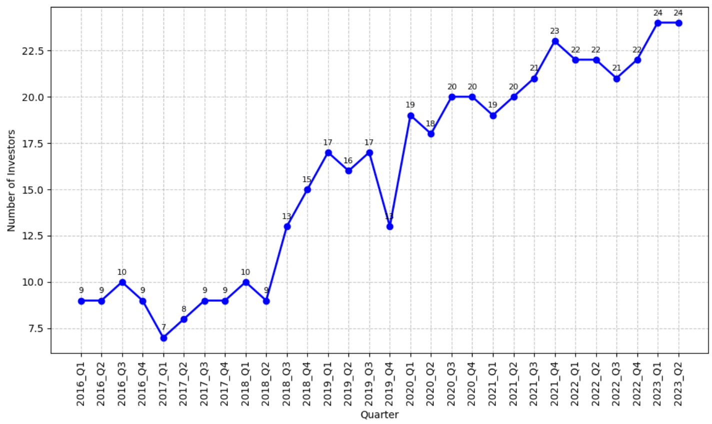
Figure 2 shows the ego network for Microsoft ('MSFT'), displaying all direct connections between MSFT and the investors who have invested in this stock across different time periods. This network is focused to show only MSFT at the center (highlighted in yellow), with other connected nodes representing investors. In this graph, edge colours indicate the frequency of investment activity: lighter edges represent less frequent investments, while darker edges signify more frequent investments. This ego network analysis offer an insight into MSFT's popularity among investors, as we can identify key investors who maintain consistent interest in the stock.

From the ego network, we examined that the investors who frequently invested in MSFT (indicated by the darker edge colour) are David Katz - Matrix Asset Advisors, Dodge & Cox, and Stephen Mandel - Lone Pine Capital. As illustrated in Figure 3a, both David Katz - Matrix Asset Advisors and Dodge & Cox made investments to MSFT in 30 different time periods, while Stephen Mandel - Lone Pine Capital invested in 29 periods. These frequent investments might indicate confidence in MSFT's performance and long-term potential. Such repeat investments can also have a reassuring effect on other market participants, signaling MSFT as a stable and attractive investment choice. Additionally, if these frequent investors were to reduce their investments in MSFT in the future, it could serve as a cautionary signal to the market. Figure 3b provides a temporal view of MSFT's investors from 2016 to 2023, showing changes in the number of unique investors each quarter. This trend shows an overall increase in unique investors over time, with some fluctuations. This increase in investor interest could be related to the

COVID-19 pandemic, in which the demand for technological advancements accelerated. MSFT, as a key player in the technology sector, likely attracted a number of investors seeking the innovations created by the company.



(a) Top Investors in MSFT



(b) Quarterly Investor Count for MSFT

Figure 3: MSFT Investment Analysis: (a) Top investors based on the frequency of investments in MSFT, and (b) Quarterly trend of investors in MSFT from 2016 to 2023.

Task 2 – Network Projections

In this section, we analyze the stock network by creating a network projection on the stock side of the bipartite network. A network projection is a transformation process that simplifies a bipartite network into a one-mode network, focusing on a single type of node. This approach reduces the bipartite structure by connecting nodes of one type based on their shared associations with nodes of the second type.

Task 2.1 – Creating a Stock-Side Network Projection

In our case, stocks are connected if they share at least one common investor, with the weight of each edge representing the number of shared investors, allowing us to study the relationships and co-investment patterns between stocks. Figure 4 shows the illustration of the projection network which contains **1,529 nodes**, **108,956 edges**, with a **density of 0.0932**. This relatively low density suggests that only a few stock pairs share common investors, reflecting distinct investment patterns. Rather than indicating a highly interconnecting market where investors frequently invest in multiple stocks, this density might indicate that certain stocks only appeal to specific groups of investors.

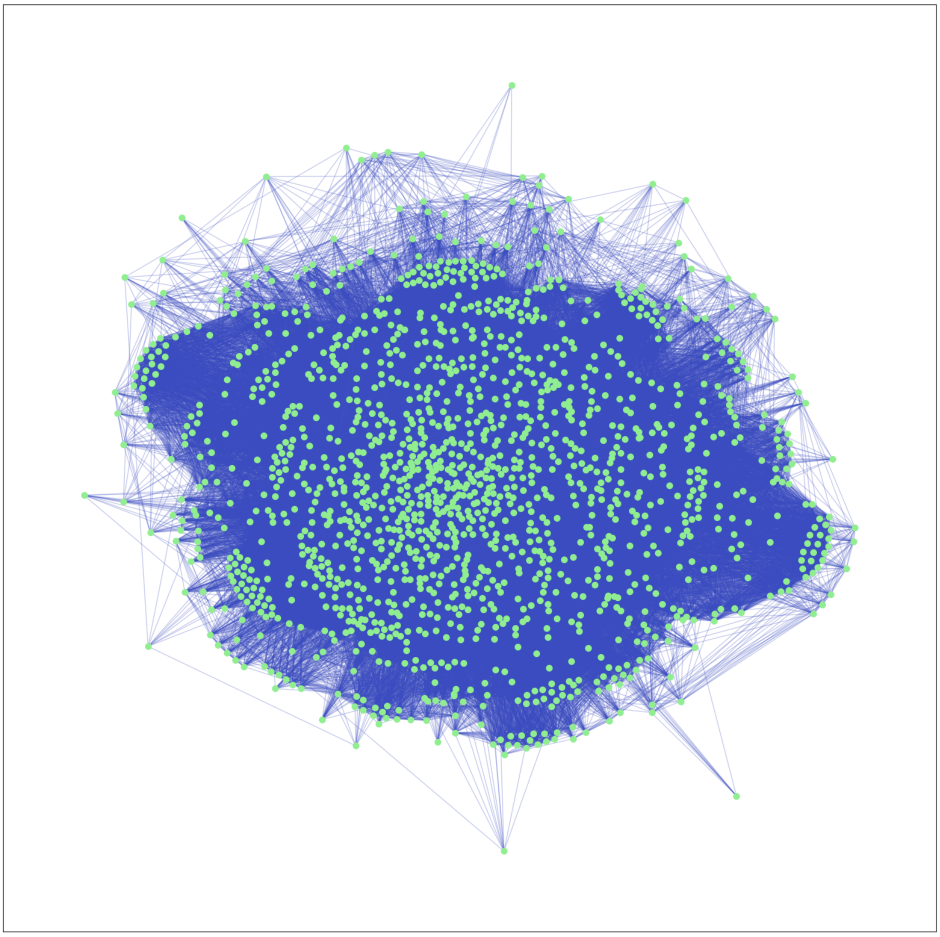


Figure 4: Visualization of the projection network.

Due to the large size of the projection network, we applied a threshold-based filtering method to enhance the interpretability of the projection network, focusing specifically on the most significant connections. Only edges representing more than 10 shared investors were retained in the filtered projection network.

Figure 5 illustrates the filtered projection network, where only 95 edges representing more than 10 shared investors were retained. Interestingly, this filtered projection network shows that MSFT is closely linked with Google ('GOOG'/'GOOGL'), Amazon ('AMZN'), Meta ('META'), and Visa ('V'), primarily companies within the technology sector. This structure indicates potential groupings by sector or industry, reflecting investor preferences for diversifying within specific sectors. For instance, an investor who holds MSFT might also invest in GOOG as a strategy to broaden their exposure in the technology industry.

By analyzing this pattern, DBBA Capital could gain valuable insights into market trends and investor behaviors. By focusing on significant connections within the network, we can identify technology-related stocks that share many investors, suggesting a sector-based preference. Investors who hold stocks in the tech sector may tend to invest in other companies within the same sector. Additionally, strong connections between stocks in different sectors also exist. For instance, MSFT is connected to UNH, which belongs to the healthcare sector, indicating that some investors diversify their portfolios across multiple resilient sectors. This pattern highlights a balanced strategy, where investors spread their investments both within and across sectors to manage risk.

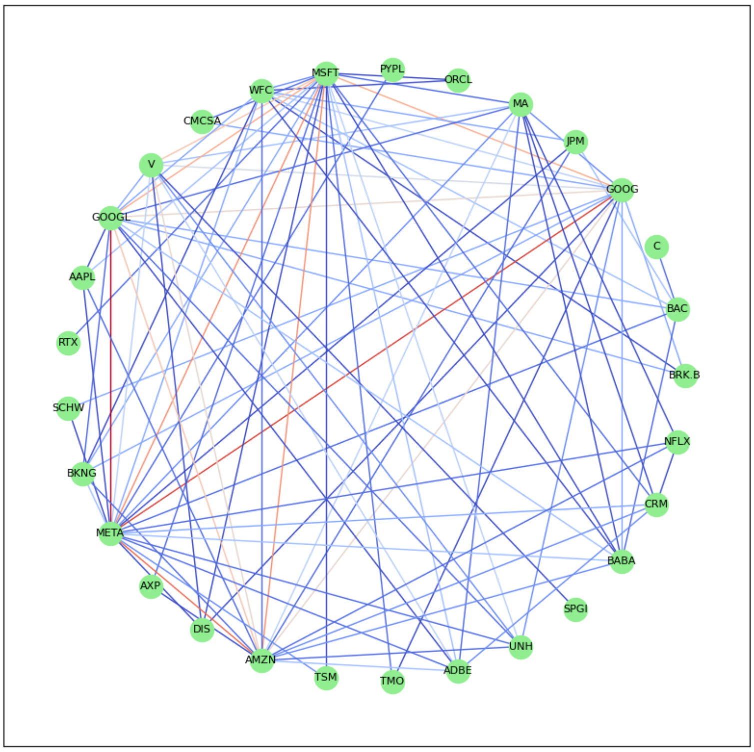


Figure 5: Filtered projection network.

Task 2.2 – Network Comparison

In previous sections, we analyzed the stock network using bipartite network and projection network, where each network provides different insights that can be used to define investment strategies for DBBA Capital.

1. **Bipartite network** connects two different set of nodes-in this case, investors and stocks-showing direct relationships between investors and their respective investments. Each edge represents an investment link, enabling us to examine which investors invested in which stocks across distinct time periods. This network helps to identify frequent investors and key stakeholders who may send signals or market sentiments. DBBA Capital can leverage this information to analyze individual investment behaviours, identify high-interest stocks, and assess market signals through the investment activities of key stakeholders in each stock. In our analysis, MSFT's investments demonstrate an overall increase in the number of investors over time, signaling high demand for technology innovations, especially driven by the COVID-19 pandemic. Key stakeholders in MSFT include David Katz of Matrix Asset Advisors, Dodge & Cox, and Stephen Mandel of Lone Pine Capital, whose sustained investment could indicate long-term confidence in MSFT's market potential.
2. **Projection network** reduces the dimension of bipartite network by focusing on one type of node (stocks), creating connections between stocks based on shared investors. Therefore, unlike bipartite network, the connections in the projection network are indirect. In the projection network, an edge

exists when two stocks share one or more investors, where the weight of each edge represents the number of shared investors. This network structure offers a higher-level view of the stock network, where we can examine co-investment patterns and clusters based on sectors or similar industries. DBBA Capital can analyze sector-based investment preferences and identify groups of stocks that attract similar investors. In our analysis, MSFT tend to form clusters with other stocks in the technology sector, such as GOOG/GOOGL, META, and AMZN, which indicates diversification strategies within the sector. However, it is also notable that MSFT shares 15 investors with UNH (UnitedHealth), indicating that diversification across multiple resilient sectorscan also be a strategic consideration to avoid sector specific risks.

The comparison between bipartite network and projection network is summarized in Table 1. By combining insights from both networks, DBBA Capital can formulate targeted strategies and risk management before investing in MSFT.

Aspect	**Bipartite Network**	**Projection Network**
Type of Nodes	Two types: investors and stocks	One type: stocks
Edge Representation	Direct connections representing investments by individual investors	Indirect connections representing shared investors between stocks
Edge Weights	Number of investments made by distinct investors	Number of shared investors
Insights Gained	Individual investment behavior	Stock clusters, sector-based diversification

Table 1: Comparison of Bipartite and Projection Networks

Task 2.3 – Ego Network Comparison

To further analyze the relationships between stocks closely linked with MSFT in terms of shared investors, we constructed a unipartite ego network, which retained only the stocks directly linked with MSFT. Figure 6 illustrates the unipartite ego network, which has 870 nodes, 70,850 edges, with a density of 0.187. This ego network structure shows a different perspective compared to the bipartite ego network that we constructed in Task 1.2.

The bipartite ego network shows direct connections between MSFT and its investors, while the unipartite ego network illustrates indirect connections between MSFT and other stocks based on shared investors. Therefore, we can examine a significant difference in terms of the number of neighbours which MSFT is connected to. In the bipartite ego network, MSFT is linked with 33 investors, whereas in the unipartite ego network it is linked with 869 different stocks. This suggests diversification strategies, where investors are more likely to invest in a broad variety of stocks. Another difference between the two networks is the possibility of connections between neighbours. Since the bipartite ego network represents direct connections between MSFT and its investors, connections between investors are not present in the network. On the other hand, in the unipartite ego network we examined the indirect connections between MSFT and other stocks, hence connections between MSFT's neighbours are present.

To further analyze the connections in the unipartite ego network, we calculated the average clustering coefficient of the network structure. The clustering coefficient is a tendency where a large number of networks show a tendency for link formation between neighbouring vertices, i.e., the network topology deviates from uncorrelated random networks in which triangles are sparse. This reflects the clustering of edges into tightly connected neihbourhoods (Saramäki et al. 2007). The average clustering coefficient for the unipartite ego network is 0.741, suggesting that nodes in this network tend to form tightly connected groups. If two stocks are both connected to MSFT (i.e., they share investors with MSFT), they are also likely to be connected to each other (share other common investors), indicating a high degree of interconnectedness. Practically, this suggests that investors in MSFT often invest in similar sets of stocks, following a co-investment strategy across multiple stocks in the network.

Figure 7 displays the filtered network from Task 2.1, where only stronger edges (representing more than 10 shared investors between stocks) are retained. This reveals that MSFT primarily clusters with

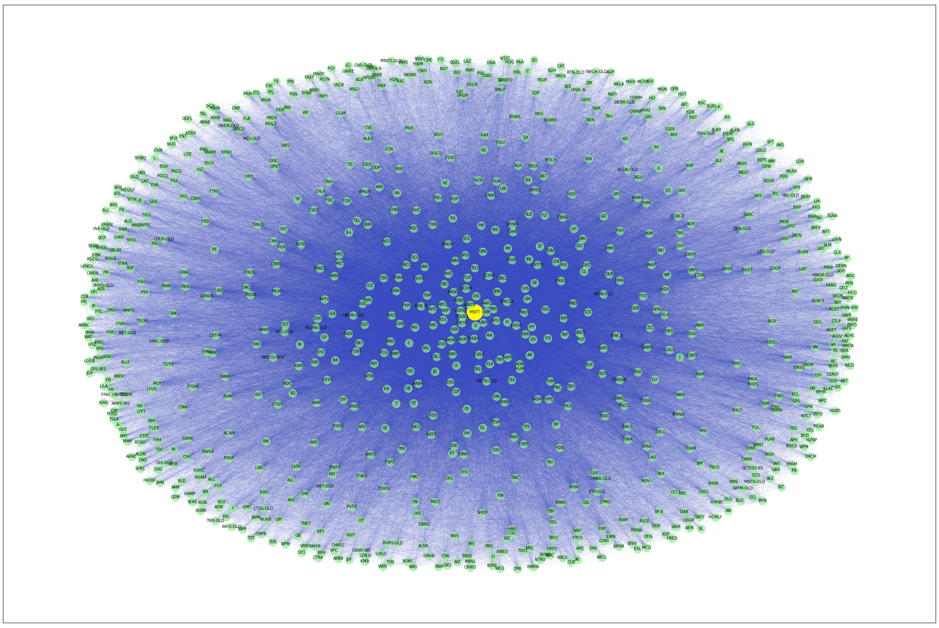


Figure 6: Visualization of the unipartite ego network

other technology stocks, such as GOOG/GOOGL, META, AMZN, AAPL, ADBE, and ORCL. However, MSFT also maintains connections with stocks from other sectors, including UNH (healthcare), WFC (financial services), and CMCSA (telecommunications), indicating that some investors follow a more balanced strategy, diversifying across resilient sectors.

Task 3 – Basic Network Analysis

In this section, we will compare both the projection network and the projection ego network based on their statistics.

Task 3.1 – Network Statistics

To analyze the network statistics, we computed the quarterly metrics for both the projection network and the projection ego network. The key metrics computed include:

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

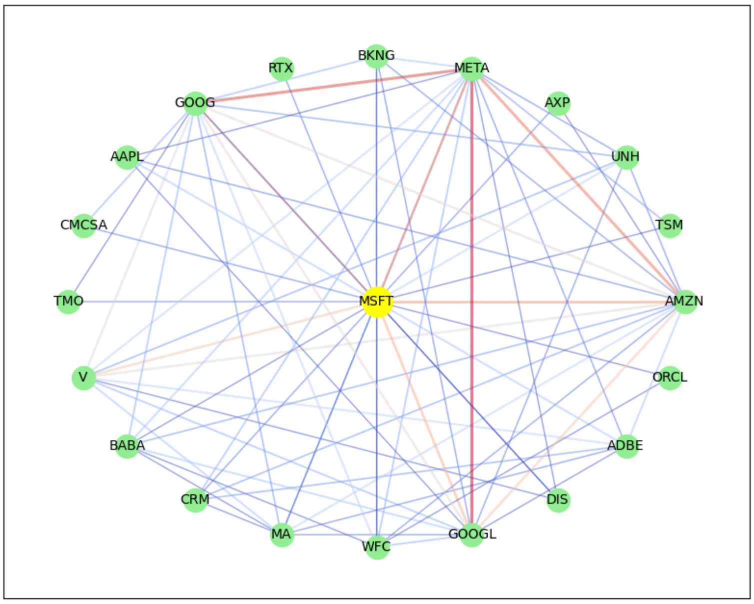


Figure 7: Visualization of the filtered unipartite ego network

For further analysis of clustering, we are particularly interested in examining the presence of clusters formed between stocks (i.e. whether edge triangles exist); thus, we have chosen to use the unweighted clustering coefficient. The weighted clustering coefficient might disproportionately reflect the clustering behaviour of a few highly connected stocks while most of the edges have relatively low weights, potentially skewing the overall clustering pattern of the network. Figure 8 shows the distribution of edge weights within the projection network, showing that most edges have low weights (few shared investors), with only a small number of edges reflecting many shared investors.

Table 2 summarizes the statistics for both the projection network and projection ego network, providing an overview of node connectivity, clustering, and assortativity patterns. We will further discuss these statistics in the next subsection.

Task 3.2 – Discussion

As mentioned in the previous subsection, Table 2 provides an overview of statistics for both the projection network and the projection ego network. As shown in Table 2a, the projection network has a low average density (0.0509), indicating a sparsely connected network. This suggests that investor behavior may be selective, with certain stocks attracting only specific investor groups, likely reflecting sector-based strategies. The high average unweighted clustering coefficient (0.8225) suggests that stocks in the projection network tend to form clusters, indicating that they are frequently co-invested by the same investor groups, likely within similar industries. The slightly negative assortativity (-0.0272) suggests a weak tendency for stocks with different degrees to connect, but given the near-zero value, it may imply that higher-degree stocks attract both low- and high-degree connections without a clear preference.

Table 2b provides statistics for the projection ego network, which focuses only on stocks directly connected to MSFT through shared investors. Thus, the number of nodes and edges is expected to be significantly lower compared to the whole projection network. The higher density (0.1723) in the ego network reflects MSFT's connections with a diverse set of stocks, as well as the likelihood of these stocks sharing additional common investors between them, thus creating more links. This denser structure is expected in an ego network since it centers around a highly connected node like MSFT. While

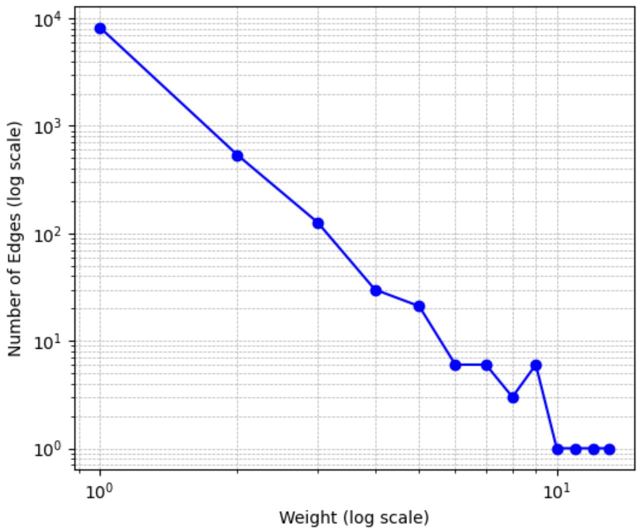


Figure 8: Visualization of the projection network weight distribution

Table 2: Statistics for Projection Network and Projection Ego Network

(a) Statistics for Projection Network

	mean	**max**	**min**	**std**
Number of nodes	586.90	648.00	496.00	60.02
Number of edges	8720.10	9823.00	6948.00	1185.91
Density	0.0509	0.0575	0.0462	0.0037
Avg clustering coefficient (unweighted)	0.8225	0.8405	0.8078	0.0077
Avg clustering coefficient (weighted)	0.1028	0.1740	0.0561	0.0399
Avg degree	29.61	31.64	27.42	1.17
Avg strength	1.1045	1.1491	1.0608	0.0303
Assortativity	-0.0272	0.0232	-0.0677	0.0309

(b) Statistics for Projection Ego Network

	mean	**max**	**min**	**std**
Number of nodes	187.97	240.00	104.00	43.30
Number of edges	2998.93	4100.00	1305.00	960.82
Density	0.1723	0.2437	0.1356	0.0266
Avg clustering coefficient (unweighted)	0.7720	0.8257	0.7339	0.0228
Avg clustering coefficient (weighted)	0.1050	0.1782	0.0579	0.0395
Avg degree	31.17	36.13	25.10	3.69
Avg strength	1.2298	1.3177	1.1299	0.0613
Assortativity	-0.1185	-0.0475	-0.1770	0.0461

slightly lower, the average unweighted clustering coefficient is still relatively high (0.7720), suggesting that co-investments by the same investor groups are still notably present even after filtering the projection network. This slight drop compared to the whole projection network may be due to exclusion of some clusters that do not include MSFT, likely within different sectors or industries. Additionally, these clusters could reflect sector-based patterns, as investors might combine MSFT with other technology stocks. The slightly negative assortativity (-0.1185) in the ego network suggests a tendency for high-degree stocks, like MSFT, to connect with lower-degree stocks. However, the value is relatively small, indicating that this tendency is not particularly strong or significant.

Task 4 – Changes of the network statistics during the pandemic

Task 4.1 – Temporal Evolution of Statistics

1. **Number of nodes**
   In Figure 9, we can observe that the number of nodes in the projection network gradually increased over the observed period, with the most noticeable growth occurring between 2018 Q1 and 2019 Q4 (pre-pandemic). This trend then became relatively stagnant during the COVID-19 period, with minor fluctuations within the 630–648 range. The MSFT ego network began with approximately 135 nodes and exhibited slight fluctuations until 2018 Q2, after which it rose to 201 nodes by 2019 Q3. Following a dip in 2019 Q4, the ego network fluctuated around 220–240 nodes during the COVID-19 period before stabilizing.

   These trends indicate that the most significant increase in the number of nodes for both networks occurred between 2018 Q1 and 2019 Q3 (pre-pandemic), likely reflecting a diversification trend or market expansion, as more stocks became interconnected through shared investors. During the COVID-19 pandemic, the relatively stable node count suggests that investors did not significantly alter the stocks in their portfolios, even during the pandemic uncertainty.

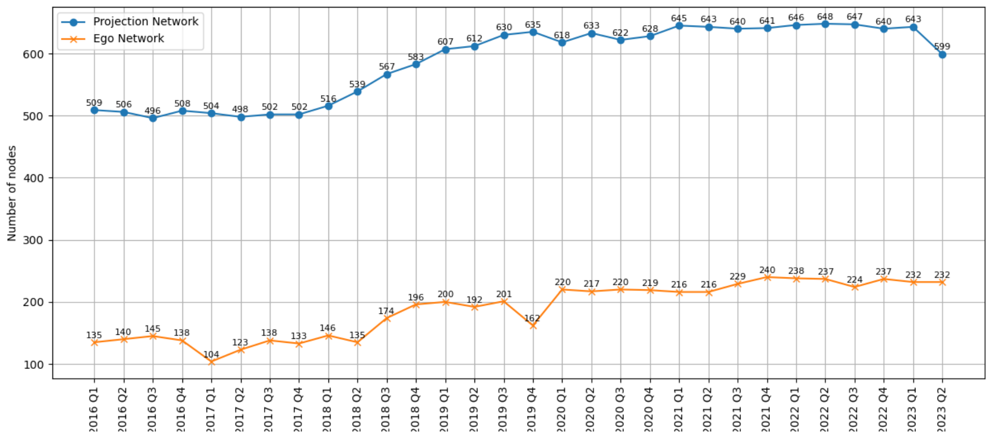


Figure 9: Temporal evolution of number of nodes

2. **Number of edges**
   As seen in Figure 10, the number of edges in the projection network remained steady around 7000 until 2017 Q3, after which it showed a noticeable increase between 2018 Q1 and 2020 Q1. During the COVID-19 pandemic, this trend stabilized with minor fluctuations in the 9500–9800 range. A slight decline is observed in 2023 Q2, with the number of edges dropping to 9005, which may indicate a reduction in co-investments post-pandemic. The MSFT ego network followed a similar pattern from 2016 Q1 to 2019 Q3, before dipping in 2019 Q4. During the beginning COVID-19 pandemic, the number of edges in the MSFT ego network saw a rise, but then relatively stabilized.

   From these trends, we can observe a pre-pandemic growth period between 2016 Q1 and 2019 Q3, with the number of edges in both networks increasing steadily. During the COVID-19 pandemic, however, we did not see a significant change in the trend as both networks showed a relatively stable trend.

3. **Density**
   In Figure 11, we observe that the density of the projection network remained relatively low and stable throughout the observed period, fluctuating around 0.050. This low and stable density suggests that the broader network remained sparsely connected, indicating limited overlap in investor holdings across the entire market. The stability, even during the pandemic, implies that while the number of edges generally increased (as shown in Figure 10), it was proportional to the increase in nodes (Figure 9), maintaining an overall sparse structure.

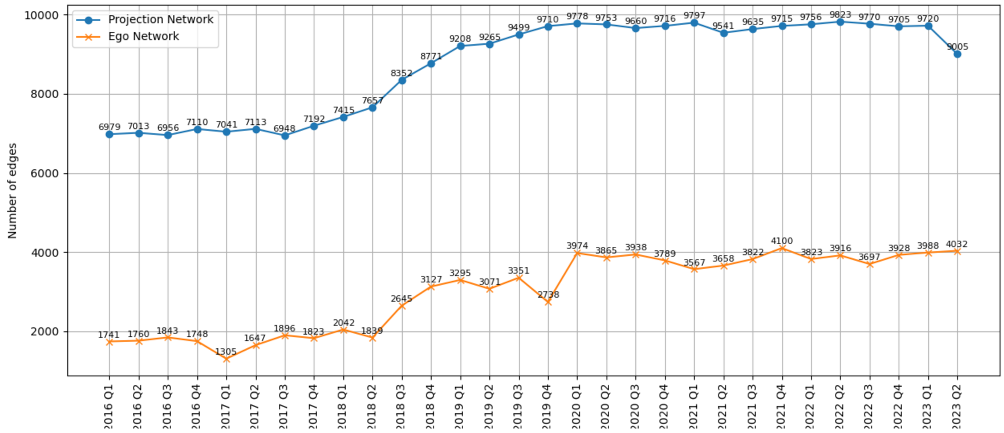


Figure 10: Temporal evolution of number of edges

In contrast, the MSFT ego network displayed a higher density than the projection network, with more notable fluctuations, particularly in 2017 Q1 and 2019 Q4. During the COVID-19 pandemic, the density of the ego network showed a slight decline, reaching 0.154 by the end of the pandemic. This gradual decline suggests that investors may have begun diversifying their portfolios, leading to a reduced concentration of connections centered around MSFT. The slight decrease in density could reflect an increased variety of stocks in investor portfolios, contributing to a more distributed pattern of co-investments.

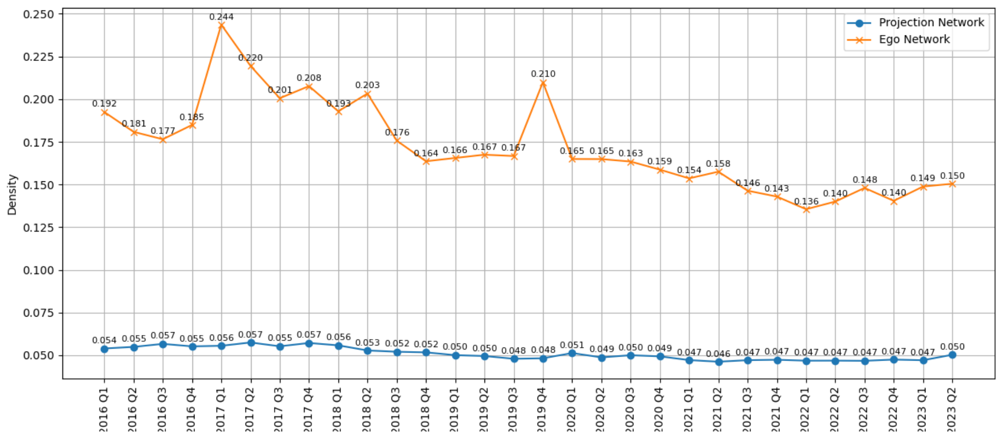


Figure 11: Temporal evolution of density

4. **Average clustering coefficient**

   In Figure 12, we observe that the average clustering coefficient for the entire projection network remains consistently high, ranging from approximately 0.810 to 0.835. This high and relatively stable clustering coefficient in the projection network indicates a strong tendency for clustering, meaning that groups of stocks are frequently co-invested by multiple investors. The stability of this clustering coefficient suggests that overall investor behavior did not shift significantly, even during the COVID-19 pandemic.

   In contrast, the MSFT ego network displays a more variable clustering coefficient, beginning around 0.801 in 2016 Q1 and generally trending downward. However, during the COVID-19 pandemic, the average clustering coefficient in the MSFT ego network started to increase, reaching 0.777 in 2021 Q1. The downward trend from 2016 Q1 to 2019 Q4 suggests that MSFT's connections became less clustered, potentially reflecting increased diversification among MSFT's neighbours, which reduced interconnections between them. The rise in clustering during the COVID-19 pandemic suggests that investors began to concentrate their investments around MSFT and similar stocks, possibly viewing them as more resilient or stable choices amid market uncertainty.

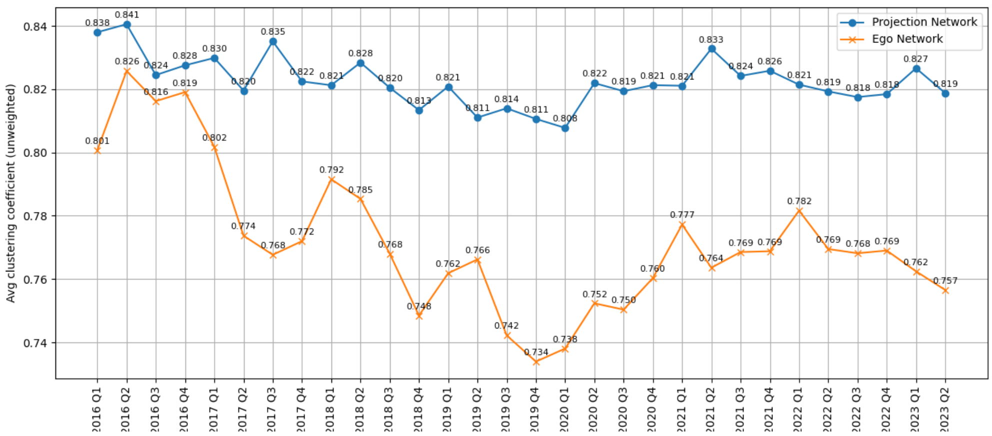


Figure 12: Temporal evolution of average clustering coefficient

5. **Average degree**
   The average degree of both the projection network and the MSFT ego network can be observed in Figure 13. For the entire projection network, the average degree showed a gradual increase from 27.4 in 2016 Q1 to a peak of 31.6 in 2020 Q1. This reflects a period of growing connectivity among stocks, possibly driven by market expansion and co-investment patterns. After 2020 Q1, the trend fluctuated slightly between 30.0 and 30.3, indicating stability within the overall network, even during the COVID-19 pandemic.

   In the MSFT ego network, the average degree also showed an upward trend, reaching its peak at 36.1 in 2020 Q1 as the COVID-19 pandemic began. This pre-pandemic increase suggests an intensification of connections around MSFT, potentially reflecting investors' heightened focus on MSFT and similar technology stocks. Following the peak in 2020 Q1, the MSFT ego network's average degree decreased slightly, which may suggest a shift in investor behavior as market conditions became more uncertain.

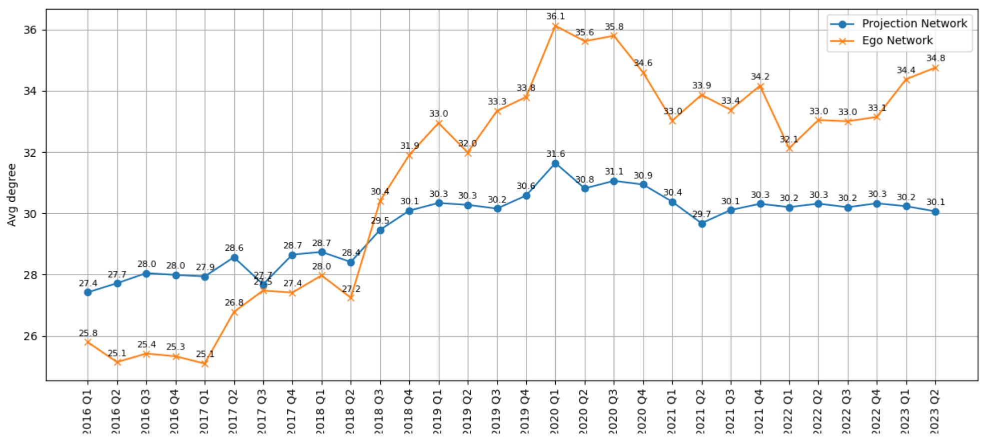


Figure 13: Temporal evolution of average degree

6. **Average strength**
   The average strength measures the average weight of connections in the network, where each weight represents the number of shared investors between stock pairs. In Figure 14, we observe a general increase in the average strength of the entire projection network from 2016 Q1, peaking at 1.149 in 2020 Q3. This rising trend suggests a growing concentration of investment interest, with

more investors co-investing in particular pairs of stocks, potentially reflecting focused investment strategies.

For the MSFT ego network, the average strength also shows a rising trend from 2016 Q1 to 2020 Q2, though with some fluctuations. This indicates that co-investment ties between MSFT and related stocks were strengthening, reinforcing MSFT's central role in investor portfolios. However, from 2020 Q2 to 2022 Q4, the average strength gradually decreased, possibly reflecting a shift in investor behavior due to the COVID-19 pandemic. Investors may have diversified their portfolios more during this period, reducing the concentration of co-investments around MSFT and responding to broader market uncertainties.

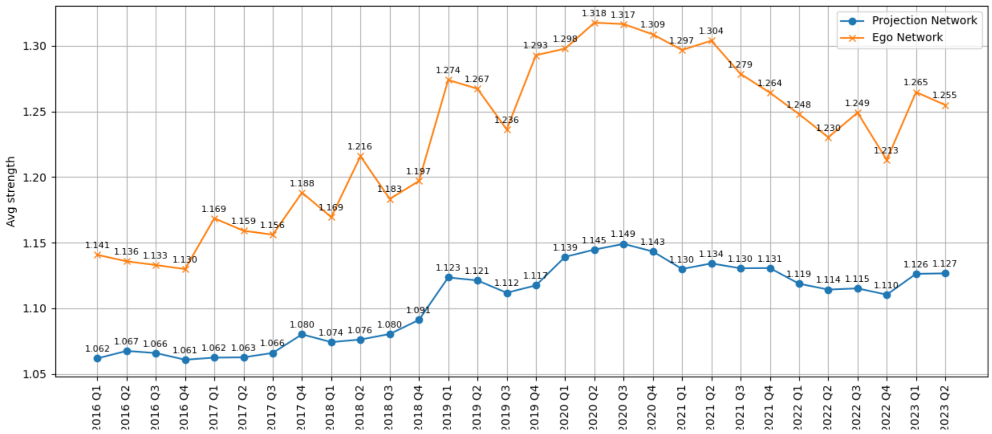


Figure 14: Temporal evolution of average strength

7. **Assortativity**
   Figure 15, we observe that in the whole projection network, assortativity shifted from an initial positive value to a minor negative value over time. Between 2016 Q1 and 2020 Q1, assortativity steadily decreased, reaching -0.061. This trend suggests that high-degree stocks increasingly connected with low-degree stocks, indicating the emergence of central hubs within the network that linked diverse nodes. During the COVID-19 pandemic, this pattern began to stabilize with only slight fluctuations, possibly reflecting the market's respond to economic uncertainty, where investors tend to invest in safer "hubs" or well-connected stocks, rather than creating assortative connections.

   In the MSFT ego network, assortativity started at a slightly negative value of -0.063 in 2016 Q1 and continued to decline, reaching a low in 2020 Q4. This trend indicates that both before and during the pandemic, MSFT consistently connected with a broad range of stocks, including many low-degree stocks, suggesting co-investment strategies. This negative assortativity points to MSFT's role as a central hub with diverse connections, potentially reflecting a strategic approach to risk mitigation and diversification across the network during periods of market instability.

To summarize, these are some statistics that showed notable shifts during the pandemic period:

- **Average clustering coefficient**: Pre-pandemic, the average clustering coefficient showed a decreasing trend, indicating that investments were becoming less concentrated around MSFT-linked stocks. This suggests that investors were increasingly open to diversifying across multiple sectors. During the COVID-19 pandemic, the MSFT ego-network became more clustered, possibly reflecting a preference for sector-based investments. This shift suggests that MSFT, as a prominent tech company, demonstrated resilience against the adverse effects of the COVID-19 pandemic and may have even benefited from increased demand for technology solutions during lockdown periods.

- **Average degree**: Before the COVID-19 pandemic, the average degree in the MSFT ego network showed an upward trend, indicating that MSFT was consistently linking to more stocks over time and becoming a focal point for investor interest. These connections could have been within the

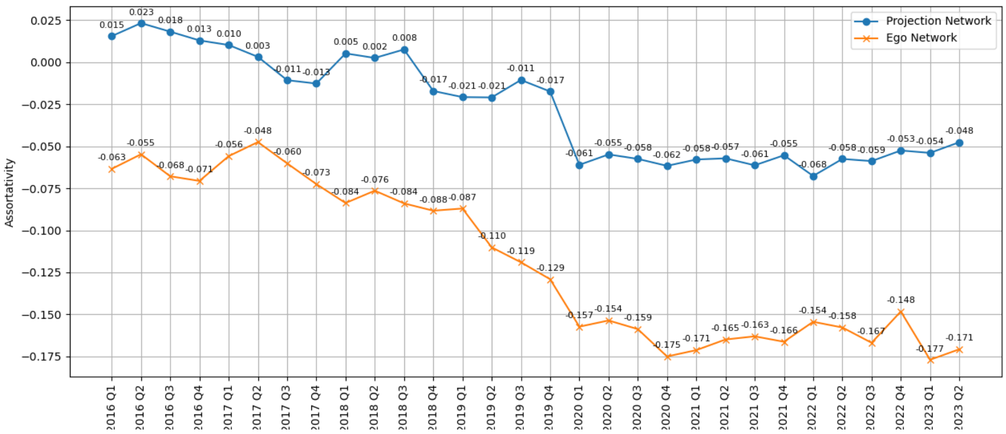


Figure 15: Temporal evolution of assortativity

technology sector or extended across other sectors. During the pandemic, however, this trend slightly decreased, suggesting that investors were not expanding their portfolios to include a wide variety of new connections. Instead, they may have concentrated their holdings on a smaller set of stocks.

Task 4.2 – Centrality

In this subsection, we analyze the centrality of nodes in the entire network to identify which stocks are most central in each quarter from 2016 to 2023. For this analysis, we use eigenvector centrality to compare MSFT's influence in the network relative to other stocks.

Eigenvector centrality is a measure of a node's importance in a connected network. Unlike simpler centrality measures, such as degree centrality, which only considers the number of direct connections a node has, eigenvector centrality assigns scores based on the concept that connections to highly influential nodes contribute more significantly to a node's centrality than connections to nodes with lower scores (Bonacich 1972). In other words, a high eigenvector centrality score indicates that a node is not only well-connected but also connected to other influential nodes. Therefore, this measure considers both the quantity and quality of a node's connections.

In the case of our stock network, applying eigenvector centrality allows us to assess whether a stock is not only interconnected with many other stocks (in terms of shared investors) but also linked to highly significant, well-established stocks in the market. A higher eigenvector centrality indicates that a stock is frequently co-invested with other influential stocks, suggesting that investors perceive it as a "safe" investment. During periods of market instability, such as the COVID-19 pandemic, stocks with high eigenvector centrality may better retain their value, as investors tend to hold onto these anchor stocks. For MSFT, we are particularly interested in observing how its eigenvector centrality evolved over time to determine whether investors consider MSFT a stable, essential stock in their portfolios, especially during periods of uncertainty.

To conduct this analysis, we computed the top 3 stocks with the highest eigenvector centrality for each quarter and analyzed MSFT's position relative to them.

In Figure 16, we observe MSFT's position relative to the top 3 most central nodes based on eigenvector centrality. While MSFT's eigenvector centrality initially fluctuated between 2016 Q1 and 2018 Q3, it showed a consistent upward trend afterward, reaching its peak in 2023 Q3. This upward trajectory suggests that MSFT's influence within the network has strengthened over time, likely due to a growing concentration of investors around MSFT and increased interconnections with other central stocks such as META, GOOG/GOOGL, AMZN, and V.

During the COVID-19 pandemic, MSFT's centrality remained close to or within the range of the top 3

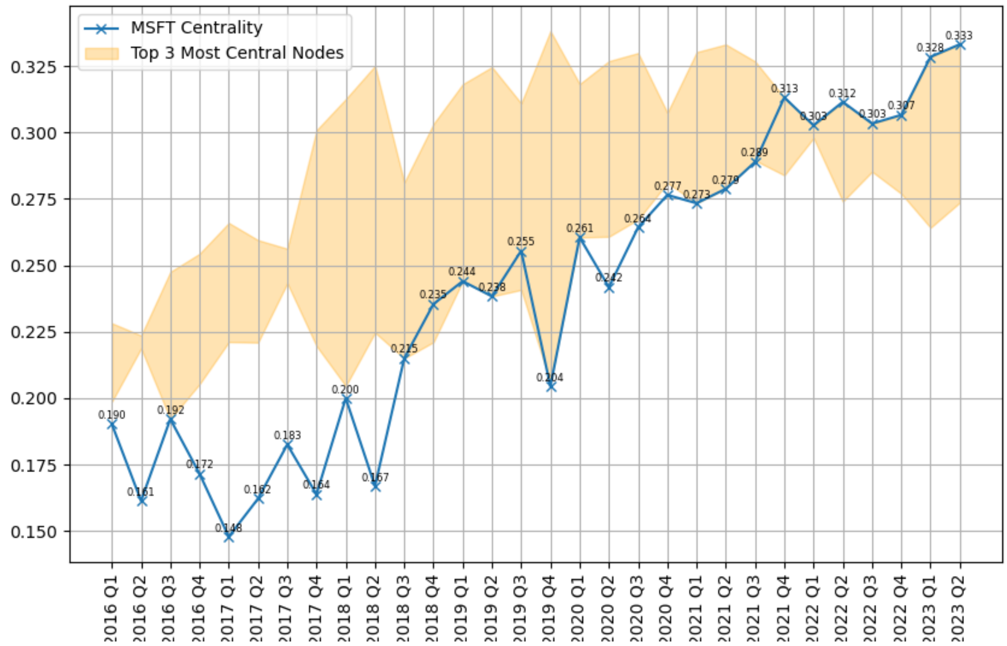


Figure 16: Temporal evolution of MSFT's eigenvector centrality

most central stocks, as shown by the shaded area in the graph. This stability may reflect a shift toward digital innovations and technology solutions driven by the pandemic, reinforcing MSFT's central role. As MSFT's centrality continued to rise, it reached the upper boundary of the top 3 range, indicating that MSFT has become increasingly consistent as one of the central stocks in the network.

In Table 3, we observe the most central nodes for each quarter from 2016 Q1 to 2023 Q2. In the pre-pandemic period, specifically from 2016 Q1 to 2018 Q3, the central nodes in the stock network were occupied by stocks from a more diverse range of sectors, such as Bank of America ('BAC'), Berkshire Hathaway Inc. ('BRK-B'), and Wells Fargo & Co. ('WFC'). Starting from 2018 Q4, however, a shift in investment preferences led to tech-based companies occupying the central positions in the network, with MSFT beginning its overall rise as a more central node. This trend continued throughout the COVID-19 pandemic period, with stocks from the technology sector—or stocks that heavily leverage technological innovations—primarily GOOG, META, MSFT, and V—consistently competing for central positions. This shift is likely due to the pandemic, particularly during lockdown periods, which drove a general move towards digital innovations. For example, Visa ('V') continued to retain centrality during this period due to its role in providing digital payment solutions, making it resilient and less susceptible to catastrophic events like the pandemic. A similar trend can be observed with MSFT, whose focus on services such as cloud computing and AI has likely strengthened its centrality within the stock network.

Overall, MSFT's rising eigenvector centrality points to a wide network influence, especially due to the fact that eigenvector centrality emphasizes the importance of connections with other important nodes. For DBBA Capital, this suggests that MSFT could be considered a reliable investment (anchor stock), especially during periods of market uncertainty when it becomes even more central within the network, likely due to the demand of the services it provides.

Task 5 – Clustering and Modularity

In this section, we will identify and analyze MSFT's communities in the stock-side network projection. To conduct this analysis, we considered two methods of community detection, the Girvan-Newman method and Louvain method.

Quarter	Most Central		2nd Most Central		3rd Most Central	
	Node	Centrality	Node	Centrality	Node	Centrality
2016 Q1	BAC	0.2283	GOOG	0.2208	BRK.B	0.1987
2016 Q2	BRK.B	0.2236	WFC	0.2191	BAC	0.2187
2016 Q3	WFC	0.2476	BAC	0.2185	MSFT	0.1921
2016 Q4	BAC	0.2544	WFC	0.2499	BRK.B	0.2053
2017 Q1	WFC	0.2660	BAC	0.2537	BRK.B	0.2212
2017 Q2	WFC	0.2595	BAC	0.2472	BRK.B	0.2210
2017 Q3	GOOG	0.2563	BAC	0.2482	WFC	0.2433
2017 Q4	GOOG	0.3009	WFC	0.2644	BRK.B	0.2196
2018 Q1	GOOG	0.3126	BAC	0.2185	WFC	0.2046
2018 Q2	GOOG	0.3251	META	0.2534	WFC	0.2247
2018 Q3	GOOG	0.2812	BAC	0.2181	MSFT	0.2150
2018 Q4	GOOG	0.3033	MSFT	0.2354	META	0.2211
2019 Q1	GOOG	0.3182	META	0.2524	MSFT	0.2441
2019 Q2	GOOG	0.3247	META	0.2666	MSFT	0.2384
2019 Q3	GOOG	0.3110	MSFT	0.2555	V	0.2408
2019 Q4	GOOG	0.3382	META	0.2210	V	0.2060
2020 Q1	GOOG	0.3183	META	0.2831	MSFT	0.2605
2020 Q2	GOOG	0.3269	META	0.2973	AMZN	0.2608
2020 Q3	GOOG	0.3299	META	0.3081	V	0.2673
2020 Q4	GOOG	0.3075	META	0.2994	V	0.2811
2021 Q1	META	0.3303	GOOG	0.3047	MSFT	0.2734
2021 Q2	META	0.3332	GOOG	0.3042	MSFT	0.2788
2021 Q3	META	0.3268	GOOG	0.3035	V	0.2895
2021 Q4	MSFT	0.3132	GOOG	0.2878	V	0.2839
2022 Q1	MSFT	0.3027	GOOG	0.2993	META	0.2980
2022 Q2	MSFT	0.3115	GOOG	0.2816	V	0.2740
2022 Q3	MSFT	0.3034	AMZN	0.2947	META	0.2853
2022 Q4	MSFT	0.3066	V	0.2868	GOOG	0.2770
2023 Q1	MSFT	0.3285	V	0.2728	GOOG	0.2642
2023 Q2	MSFT	0.3332	AMZN	0.2866	GOOG	0.2736

Table 3: Top 3 Central Nodes by Eigenvector Centrality per Quarter (2016 Q1 - 2023 Q2)

The Girvan-Newman method detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. The Louvain method finds small communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node, and the step is repeated.

Since the Girvan-Newman algorithm relies on removing edges with high betweenness, it might be more suitable if applied to networks which have distances as the edge weights, aligned with its emphasize on calculating the betweenness centrality in each iteration. Additionally, the Girvan-Newman method might struggle to find communities within large networks, due to the fact that the initial structure of large-scale networks are likely to be unknown.

The Louvain method detects communities through a modularity optimization process. Initially, it groups nodes into small communities by locally optimizing modularity, then groups these small communities into single nodes and repeats the process. This iterative approach continues until no further modularity improvements are possible. Since the Louvain method prioritizes modularity maximization, it is highly effective for large networks and can detect communities based on strong internal connections.

Given the structure of our network, where edge weights represent the number of shared investors (indicating strong connections between stocks), the Louvain method is better suited for our analysis. Its focus on maximizing modularity aligns well with our goal of identifying cohesive groups of stocks based on shared investors. Consequently, we will use the Louvain method for the subsequent analysis.

Figure 17 shows the results of applying the Louvain method to detect communities within the stock projection network. Modularity measures the strength of a network's division into modules (communi-

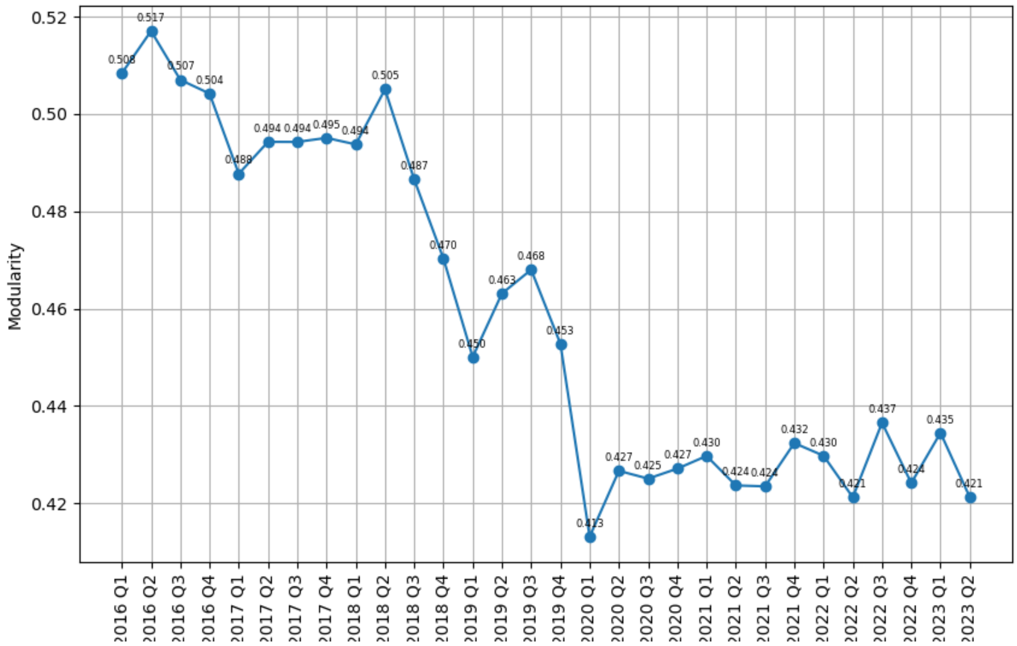


Figure 17: Temporal evolution of modularity

ties), with high modularity indicating dense connections within modules and sparse connections between modules. From the graph, we observe that from 2016 Q1 to 2019 Q4, the modularity value displayed an overall downward trend, suggesting that the network became harder to divide into distinct communities. This trend implies that during this period, broader, more diverse investment strategies and speculations emerged in the market, leading to investments that were less concentrated in specific sectors. However, at the start of the COVID-19 pandemic, this trend reversed, with modularity values fluctuating around 0.420 - 0.435 through the end of the observation period. This suggests that community divergence halted, likely because investors became more cautious and concentrated their investments in specific sectors during the pandemic and recovery periods.

Figure 18 illustrates the number of stocks in MSFT's community over time. The number of stocks initially shows an overall upward trend from 2016 Q1 to 2020 Q1, followed by a period of fluctuations until the end of the observation period. This pattern indicates an initial phase of market expansion and diversification, with an increasing number of co-investments leading to a wider variety of stocks being grouped with MSFT. During the COVID-19 pandemic, the trend became more variable, suggesting that investors adopted a cautious approach to portfolio expansion, potentially reflecting a wait-and-see strategy. Despite these fluctuations, there is no clear downward trend, indicating that MSFT remained a relatively stable investment choice, consistently included in investor portfolios.

Figure 19 illustrates the dynamic evolution of stocks that clustered with MSFT over time, filtered to show the top 50 stocks with the most occurrences in MSFT's community. The blue colour indicates that the stock was a member of MSFT's community in the specific period. From the graph, we observe that technology-focused stocks such as Netflix ('NFLX'), Amazon ('AMZN'), Google ('GOOG'), Adobe ('ADBE'), and Meta ('META') were frequently grouped with MSFT, especially in more recent periods, indicating a high level of co-investment among them. This trend is likely driven by the shift in market sentiment during and after the pandemic, as investors increasingly favored stocks that heavily leverage technology. Another interesting observation is that MSFT is consistently clustered with UnitedHealth Group Inc. ('UNH'), a healthcare stock. This implies that both the technology and healthcare sectors are seen as strong and resilient, even during periods of market uncertainty, prompting investors to diversify towards leaders in these sectors. Additionally, the COVID-19 pandemic accelerated the integration of technology innovations within the healthcare sector, such as data analytics and telemedicine, thereby attracting investors interested in the convergence of technology and healthcare.

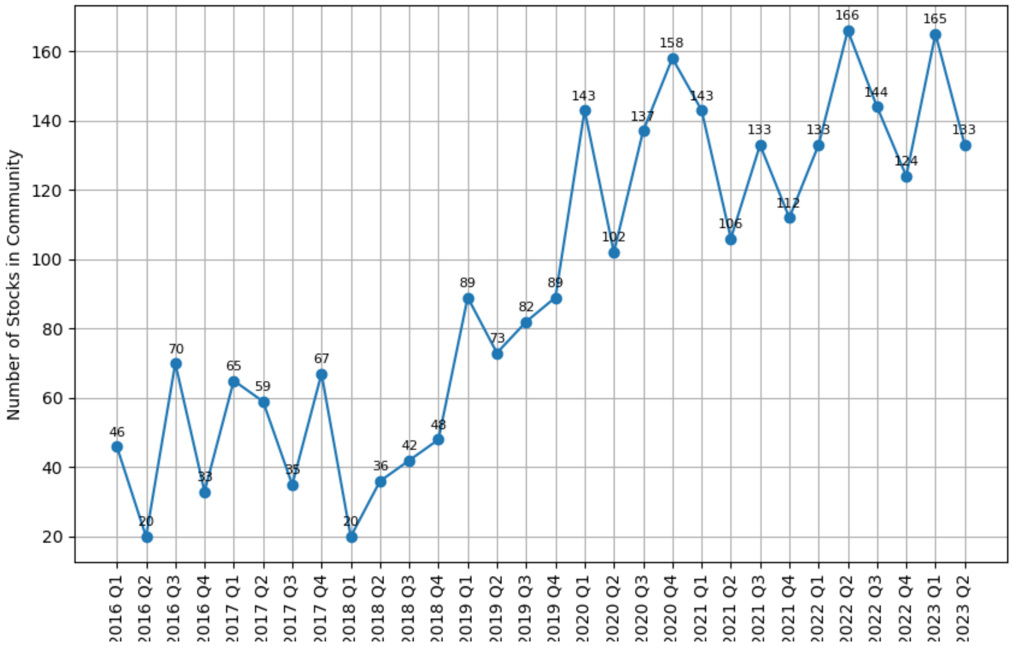


Figure 18: Number of stocks in MSFT community

In conclusion, the Louvain community detection method frequently clustered MSFT with stocks that leverage technology, indicating a sector-specific investment trend. However, some diversification strategies were also observed, as evidenced by UNH's consistent appearance in the same community. DBBA Capital can leverage these insights to formulate market-entry strategies, where a focused approach in the technology sector, combined with moderate diversification into resilient sectors (such as healthcare), may be suitable.

Task 6 – Analyzing Stock Correlation Through a Spanning Tree

In this section, we will analyze MSFT's role through different periods (pre, during, and post pandemic) using the spanning tree method.

Task 6.1 – Construct a Stock Correlation Network

To construct the stock correlation network, we used the daily stock prices data, ranging from January 2018 to September 2024. We then transformed the data by first retaining the end-of-month stock prices and calculating the monthly returns for each stock. Next, we calculated the correlations between each stock and built a network with edge weights representing the correlations between the stocks.

Figure 20 illustrates the stock correlation network, with a total of 49 nodes representing stocks and 1,176 edges representing their correlations. To focus more on MSFT's correlation with other stocks, in Figure 21 we can observe the top 15 stocks most correlated with MSFT. Many of these highly correlated stocks belong to the technology sector, including ADBE, AMZN, AAPL, GOOG, and NVDA, reflecting shared market influences, similar business models, or overlapping investor interests in this sector. Interestingly, we also find the Canadian National Railway ('CNI') among the top correlations. This strong connection may stem from the Bill & Melinda Gates Foundation Trust, which holds significant shares in CNI.

This analysis suggests that while MSFT is naturally correlated with other technology firms due to sectoral similarity, it also shows meaningful correlations with firms outside of the sector, likely indicating

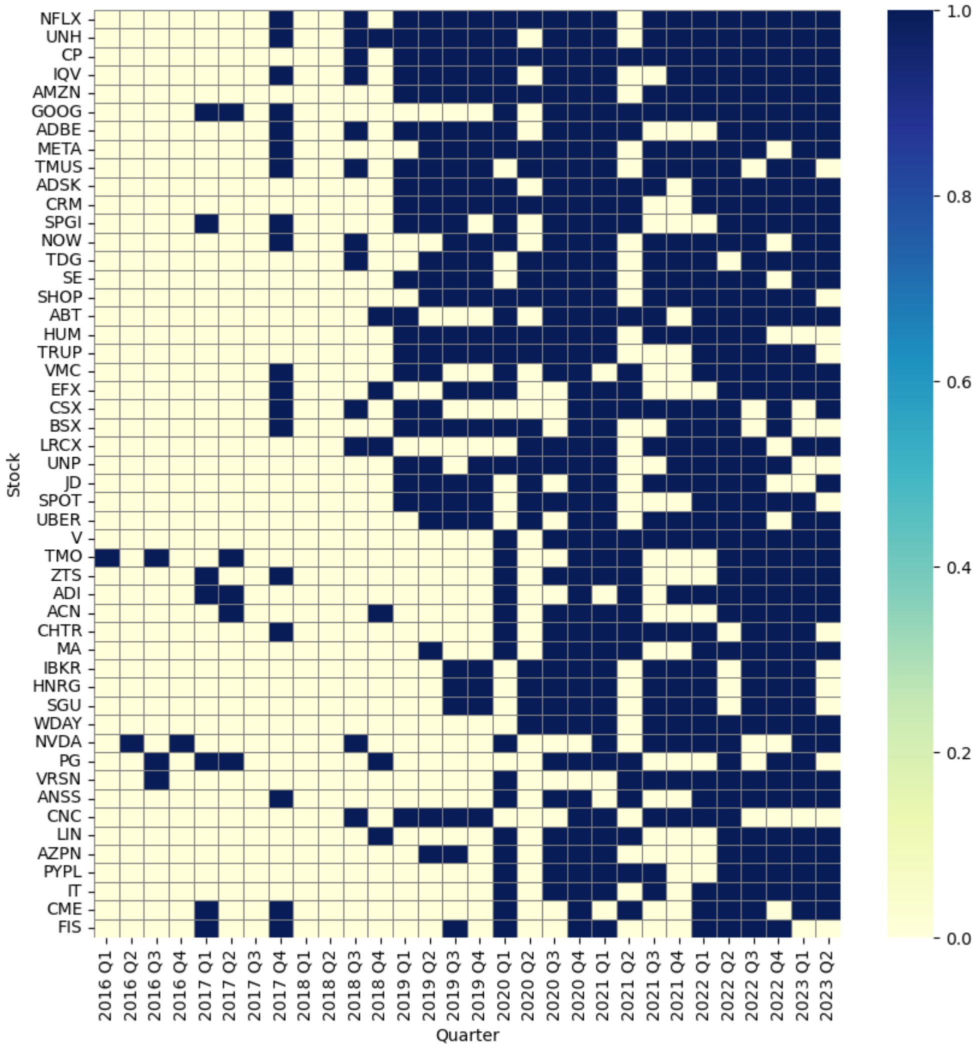


Figure 19: MSFT community members over time (Top 50 stocks based on occurences)

potential diversification strategies where investors see MSFT as a "core-holding", while also investing in other resilient industries, such as transportation.

Task 6.2 – Build a Spanning Tree

To further analyze our stock correlation network, we constructed a minimum spanning tree based on the distances between stocks. Distance is derived from the correlation value, where a smaller distance implies a stronger relationship between stocks, indicating that their prices tend to move together. We used distance as a measure because the minimum spanning tree algorithm efficiently connects all nodes with the smallest possible sum of edge weights (distances), emphasizing the most significant relationships in the network. In financial terms, shorter distances suggest that stocks might respond similarly to market conditions or macroeconomic factors, providing valuable insights for diversification strategies.

Additionally, using distance allows us to analyze the network using distance-based centrality measures, which include closeness centrality and betweenness centrality. Closeness centrality measures how close a node is to all other nodes in the network based on average shortest path distances. A node with high

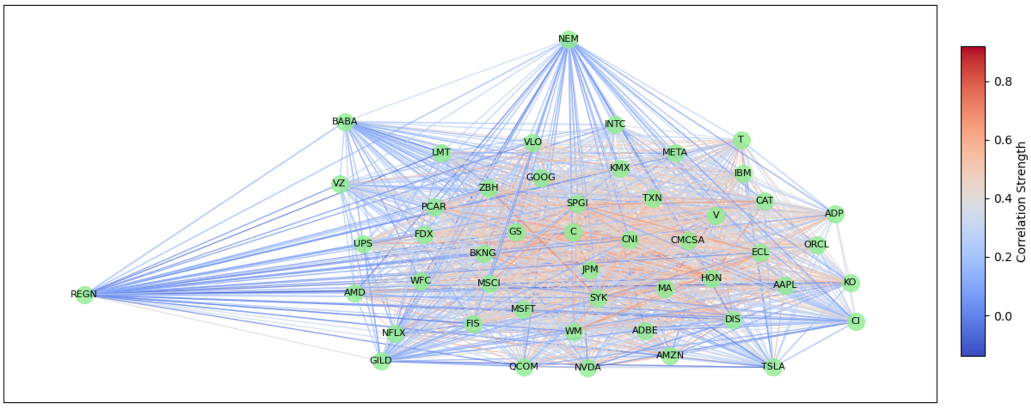


Figure 20: Stock correlation network

closeness centrality has relatively shorter paths to all other nodes in the network, suggesting that it is central or accessible in the network (Sabidussi 1966). Betweenness centrality measures the extent in which a node lies on the shortest paths between other nodes. Nodes with high betweenness centrality act as bridges or intermediary in the network (Freeman 1977).

Task 6.3 – Analyze MSFT's Role in the Spanning Tree

Figure 22 illustrates the minimum spanning tree of the stock network, where we can observe that MSFT clusters around other technology stocks like Google ('GOOG'), Adobe ('ADBE'), Amazon ('AMZN'), Meta ('META'), Intel ('INTC'), Apple ('AAPL'), and Nvidia ('NVDA'), suggesting a sectoral influence within the spanning tree. In this structure, MSFT serves as a hub for the technology cluster while also acting as a bridge connecting this cluster with other parts of the tree. In terms of betweenness centrality, MSFT ranks relatively high (0.4043), indicating that it lies on many shortest paths between other pairs of nodes. This aligns with MSFT's role as a bridge, linking the technology cluster to other parts of the tree. However, MSFT's moderate closeness centrality (0.2612) suggests that it is closely connected within the technology cluster but is relatively distant from other parts of the network. Consequently, stocks like Visa (0.3768) and Honeywell (0.3574) are more central in terms of market influence, as they are more "connected" across the network, allowing them to act as broader market indicators. Table 4 provides an overview of MSFT's position in the network.

Compared to the previous analysis, MSFT was found to be more influential in the co-investment network due to its high eigenvector centrality, often ranking alongside other dominant tech stocks. This may be because the technology sector was perceived as a safer option by a wide variety of investors looking to diversify their portfolios.

Next, we investigate the behavior of the spanning tree across different periods of the pandemic (Figure 23). In Figure 23a, we observe that before the pandemic (before March 2020), MSFT was more centrally positioned, with a high betweenness centrality of 0.479 (3^{rd} highest) and a closeness centrality of 0.466 (2^{nd} highest). This central position allowed MSFT to act as a bridge, effectively connecting different sectors within the network. Additionally, the relatively distributed structure of the tree, with fairly spread-out clusters, suggests that stock correlations were less sector-based during this period.

In Figure 23b (March 2020 - March 2021), we observe MSFT's role as one of the bridges in the network, with a betweenness centrality of 0.438 (8^{th} highest). However, its position shifted away from the center of the tree, as indicated by its lower closeness centrality of 0.279 (10^{th} highest), suggesting that MSFT was less connected to the overall network. Additionally, technology stocks, including MSFT, formed a more distinct cluster. This indicates that MSFT and these technology stocks became more interconnected, possibly reflecting investor sentiment during the pandemic, when technology companies were seen as

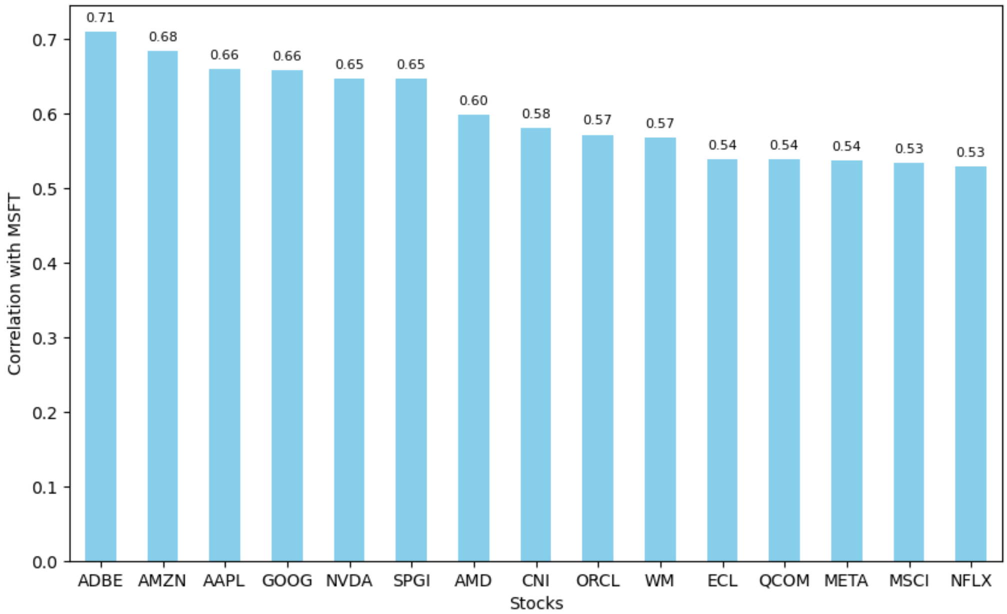


Figure 21: Top 15 stocks correlated with MSFT

essential.

In Figure 23c (after March 2021), the technology cluster observed previously remained prominent. The tightly connected cluster suggests that the prices of these stocks were highly interdependent, likely due to similar market trends, investor sentiment, or macroeconomic factors. MSFT exhibited more direct connections with other leading technology stocks such as NVDA, ADBE, GOOG, and META, establishing itself as a hub within the technology sector. Moreover, MSFT acted as a bridge between the tech sector and other parts of the tree, helping to propagate information through the network (0.474 betweenness centrality, 5^{th} highest).

Task 7 – Discussion

In this assignment, we presented two different viewpoints to analyze MSFT's role in the stock network. In the **stock projection network (co-investment analysis)**, MSFT's high centrality (calculated using eigenvector centrality) suggests that it is a popular and reliable anchor stock, consistently included in investors' portfolios. This implies continued investor confidence in MSFT's potential, especially given its strategic focus on artificial intelligence (AI), cloud computing, and other technological innovations that align with the growing global demand for digital transformation. For DBBA Capital, this signals that MSFT is a reliable stock trusted across a wide range of portfolios.

In the **minimum spanning tree network (return correlation analysis)**, MSFT's position and relatively high betweenness centrality (0.404) indicate that it acts as a hub within the technology sector, bridging it to other sectors. However, MSFT's moderate closeness centrality (0.261) suggests that it is not as immediately connected to the entire network, meaning its performance may be more influenced by sector-specific trends. By comparison, stocks like Visa ('V'), which has the highest closeness centrality in the spanning tree, may reflect market trends more quickly, potentially making them better market indicators than MSFT.

To conclude, both analyses highlight MSFT's strong and consistent position in the technology sector, suggesting that it could be a stable investment if DBBA Capital considers adding it to its portfolio. MSFT's high eigenvector centrality in the co-investment analysis, combined with its relatively high

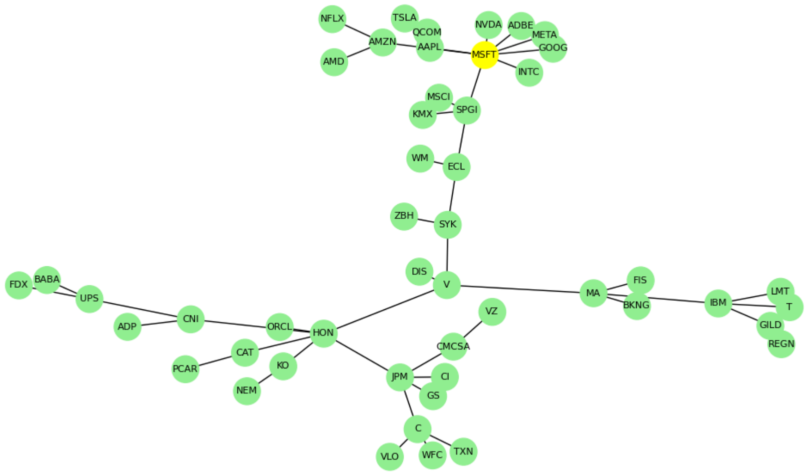


Figure 22: Minimum spanning tree using distance

betweenness centrality in the stock return analysis, confirms its resilience. However, given MSFT's strong integration within the technology sector, DBBA Capital should also consider a diversified approach by including stocks from resilient sectors such as financials or healthcare. A diverse portfolio would enable DBBA Capital to mitigate sector-specific risks while benefiting from MSFT's position in the technology sector.

References

Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1):113–120, 1972.

Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, March 1977. ISSN 00380431. doi: 10.2307/3033543. URL `http://dx.doi.org/10.2307/3033543`.

Linton C. Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304, 1982. ISSN 0165-4896. doi: https://doi.org/10.1016/0165-4896(82)90076-2. URL `https://www.sciencedirect.com/science/article/pii/0165489682900762`.

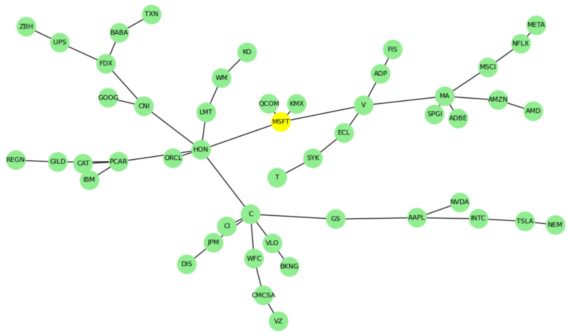
Leigh Metcalf and William Casey. Chapter 5 - graph theory. In Leigh Metcalf and William Casey, editors, *Cybersecurity and Applied Mathematics*, pages 67–94. Syngress, Boston, 2016. ISBN 978-0-12-804452-0. doi: https://doi.org/10.1016/B978-0-12-804452-0.00005-1. URL `https://www.sciencedirect.com/science/article/pii/B9780128044520000051`.

Gert Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, December 1966. ISSN 1860-0980. doi: 10.1007/BF02289527. URL `https://doi.org/10.1007/BF02289527`.

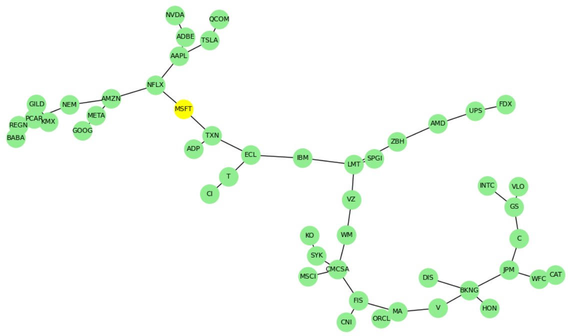
Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and János Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2), 2007. ISSN 1550-2376. doi: 10.1103/physreve.75.027105. URL `http://dx.doi.org/10.1103/PhysRevE.75.027105`.

Rank	Degree Centrality		Betweenness Centrality		Closeness Centrality	
	Node	**Value**	**Node**	**Value**	**Node**	**Value**
1	**MSFT**	**0.167**	V	0.655	V	0.377
2	HON	0.125	HON	0.598	HON	0.357
3	JPM	0.104	SYK	0.494	SYK	0.354
4	IBM	0.083	ECL	0.467	MA	0.341
5	SPGI	0.083	SPGI	0.444	ECL	0.326
6	C	0.083	**MSFT**	**0.404**	JPM	0.305
7	V	0.083	JPM	0.302	SPGI	0.295
8	MA	0.083	MA	0.264	CNI	0.291
9	AAPL	0.063	IBM	0.160	DIS	0.288
10	UPS	0.063	CNI	0.159	ZBH	0.282
11	CNI	0.063	C	0.122	CAT	0.281
12	SYK	0.063	AAPL	0.082	KO	0.275
13	AMZN	0.063	UPS	0.082	ORCL	0.274
14	ECL	0.063	AMZN	0.082	FIS	0.273
15	CAT	0.042	CAT	0.042	IBM	0.272
16	GILD	0.042	GILD	0.042	BKNG	0.271
17	CMCSA	0.042	CMCSA	0.042	C	0.265
18	KO	0.042	KO	0.042	**MSFT**	**0.261**
19	KMX	0.021	KMX	0.000	GS	0.259
20	WM	0.021	WM	0.000	WM	0.258

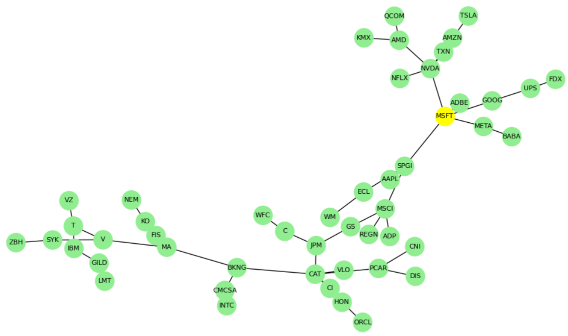
Table 4: Centrality Measures for Selected Nodes



(a) Pre-Pandemic



(b) During Pandemic



(c) Post-Pandemic

Figure 23: MSFT's role in the network across different periods