A Bipartite Network Analysis of Global Investment Markets

COMP0123: Complex Networks & Online Social Networks

Minoo Kim
Student ID: 25194663
Candidate Number: XJYH6

8 January 2026

Abstract

The true value and stability of a company are no longer determined solely by internal financial metrics, but by its structural position within the complex world of global investments. This study uses a network science perspective to analyze this interconnectedness, modeling the relationship between institutional investors and publicly traded companies as a weighted bipartite graph. Utilizing a large scale dataset derived from September 2025 SEC 13F filings, we study a "blind" structural analysis to test whether the topology of capital allocation, independent of financial history, can accurately identify market leaders and systemic risks. To achieve this, HITS, PageRank, betweenness centrality, and bipartite clustering coefficients are used to identify core companies acting as critical structural bridges. The results suggest that the 2025 financial landscape is highly centralized where the structural diversity of investors are a vital indicator of a company's success.

Contents

1 Introduction

1.1 Context

With modern financial markets becoming increasingly more complex and interconnected, it is crucial to understand a company's position within a larger framework. For example, looking at a company in isolation is no longer enough to understand its true value, even with internally large financial profit or innovation. To get a complete picture of a firm's standing within this landscape, we must look at where it sits within the broader picture of global investments. A company's true standing, therefore, can be measured by the strength and quality of the connections it maintains with major institutional investors.

Thus, this project is motivated by the belief that we can identify industry leaders and predict success simply by analyzing these connection patterns. By looking at the static investor-company network as a bipartite graph, we can identify network structures that could potentially define a company's "success".

Unlike standard "unipartite" social networks, this bipartite network consists of two distinct sets of nodes: Institutional Investors (e.g., hedge funds, mutual funds, and banks) and Publicly Traded Companies. An edge exists between an investor and a company whenever a holding is reported in the SEC 13F filings, a mandatory quarterly report for large institutional investment managers. Additionally, these edges are weighted by the market value of holdings (how much an investor invests into a company). The specific network data and connections are further explored in the Methodology section.

Therefore, by performing a "blind" analysis, this study tests whether the shape of the network alone can reveal the most influential players in the economy, even without knowing their names or financial history beforehand. The primary research question I then focus on is the following:

Q1: To what extent can the structural position of a company within a bipartite investment network serve as a reliable proxy for its financial success and market influence, even in the absence of traditional financial data?

1.2 Motivations

While traditional financial research often views institutional ownership as a simple percentage, I argue that the network topology of that ownership is more informative. Researchers usually look for correlations between the percentage of institutional ownership and future stock returns, or use these investment features as part of a classification or prediction problem. This project is unique, however, because it treats the market as a connected network of companies and investors that rely on each other. Thus, it preserves a pure bipartite topology, allowing for a direct analysis of the relationship between investors and companies, as well as conducting a blind analysis of investor-company relations.

The answer to this research question would be valuable for both investors and companies, as it can identify "Hubs" of capital and "Authorities" that may not be obvious from company financial statements. It also removes all identifiers and financial histories (in a "blind" analysis), testing if pure network structures can independently imply company "success".

Unlike studies that simplify the data, this project is suitable for network-science analysis as it analyzes the raw bipartite structure of the investor-company network. By doing so, it recognizes that an investor's influence is a function of the companies they hold, and a company's value is a function of the investors who hold it (this can be captured using the HITS algorithm). Usually, researchers also assume that companies are independent (as in, they are treated as indepedent observations/samples/features). Thus, network science is required here because the nodes are actually interdependent (e.g., the "Hub" score of an investor affects the "Authority" score of every company in its portfolio).

In addition to the primary research question, I also answers the following specific subquestions derived from the bipartite network's architecture:

Q2: Can we identify company success before it appears in future SEC quarterly earnings reports?

Q3: Can we identify a company's downfall through changes in its network connectivity or investor 'clustering'?

2 Literature Survey

There are many research works related to investor networks and firm performance. To better understand these current research endeavors and how they motivate the "blind" analysis of this project, I introduce three recent works that use bipartite network structures:

1. *Investor networks and social innovation* [1]

2. *Institutional investor networks and firm performance* [2]

3. *Better to stay apart: asset commonality, bipartite network centrality, and investment strategies* [3]

2.1 *Investor networks and social innovation*

Del Giudice et al. [1] leverages Stakeholder Network Theory to investigate how the structural position of investors within a network influences the success of Social Impact Bonds (SIBs). SIBs are collaborative financial instruments where private capital is used to fund social programs, with returns based on achieving specific social outcomes. The paper analyzes an investor-stakeholder network consisting of institutional investors, philanthropists, and impact investors by building a bipartite graph and projecting it into a unipartite investor network. They find that higher betweenness centrality correlates to greater influence, where "influence" is linked to improved project performance and stronger social impact outcomes. It is also noted that private companies act as key hubs that bridge information gaps between different types of stakeholders.

The authors' approach of focusing on examining global properties before turning to individual node metrics are useful, specifically using PageRank to assess the "quality" of connections and using violin plots to show how different investor types occupy different network positions. They conclude by suggesting a success prediction framework using network centrality. I will similarly explore network importance via PageRank.

Although the paper finds that investors with higher betweenness centrality have a greater influence, there are key limitations. Namely, the dataset they analyze contains only 347 investors and 293 stakeholders, which suggests the network may not be large enough to show the true scale-free behavior they describe. The success metric also relies on social impact, which is often subjective. Finally, by projecting a bipartite network into a unipartite one, direct investor-company influence information can be lost.

2.2 *Institutional investor networks and firm performance*

Bajo et al. [2] provides research findings most relevant to the motivations of this project. They argue that the network centrality of institutional blockholders, or investors owning more than 5% of a firm, impacts the market value of the companies they hold. This paper also uses data from SEC filings and constructs a bipartite network of active investors and firms. Similar to the previous paper, this network is projected into a one-mode investor network where two investors are linked if they share a blockholding in at least one firm. The paper concludes that there is a strong positive correlation between an investor's centrality and a firm's value, mostly due to the "Certification Effect," where a central investor acts as a signal of investment quality to the market.

Although on the surface [2] suggests research endeavors overlapping with this project, there are several divergent areas to identify. Firstly, even though their data comes from the same SEC filings, it only goes up to 2013. Using 2025 data is not only more representative of current relationships, but it is also more centralized and influenced by large companies (e.g., BlackRock, Vanguard). This temporal gap is significant because the modern landscape has shifted toward a state of high investor density that older datasets cannot capture. Furthermore, the methodology in [2] manually excludes passive investors and builds a unipartite graph. This projection may remove distinct relationships between investors and a specific company. By contrast, this project preserves the bipartite topology, allowing for a simultaneous analysis of investor hubs and company authorities, which ensures that the influence of high quality capital flows are fully accounted for in our success metric.

A key point we can borrow is the "Certification Effect" hypothesis. If high authority companies can be certified by the hub investors who own them, our use of HITS is further supported because being linked to other central nodes adds value to the firm.

2.3 *Better to stay apart: asset commonality, bipartite network centrality, and investment strategies*

In this paper, Flori et al. [3] dives into the relationship between the performance of U.S. mutual funds and their respective portfolio holdings. They represent this using a bipartite network with nodes being funds and assets, and links connecting the two. The main indicator proposed is the Average Commonality Coefficient (ACC) which measures the average popularity of assets within a fund's portfolio. If the ACC is low, it means a fund invested in niche assets, while a high ACC means a fund invested in popular assets commonly held in many other portfolios. Flori et al. finds that funds with low ACC generally outperformed those with high ACC, even after correcting for standard risk factors.

This paper similarly uses a bipartite network and utilizes its structure to identify important patterns. It also holds the idea that a node's structural position is an indicator for financial success. A major takeaway is the evidence provided for why certain network structures are more stable than others; for example, low ACC funds were less affected by the 2007-08 financial crisis.

It should be noted, however, that the data used is limited to U.S. mutual funds from 2004 to 2010, so the data is less recent and limited to the U.S. market. Also, because the ACC takes an average of an asset's neighbor values, more complex connections between funds and their assets may be lost.

3 Methodology

3.1 Data Source

To construct the network for this project, the first step involves transforming raw financial disclosures into a bipartite structure. The primary data source is the Layline Institutional Holding Reports (from September 2025) found on Kaggle [4] [5], which are derived from SEC 13F filings. A 13F filing is a mandatory quarterly report required by the Securities and Exchange Commission for all institutional investment managers with at least $100 million in assets. I use this data because it provides a transparent, standardized, and accurate representation of where the world's largest areas of capital are currently allocated. Thus, we can capture a global view of investments in specific public companies.

Formally, the network is defined as a bipartite graph $G = (U, V, E)$, where the sets U and V represent institutional investors and publicly traded companies, respectively. An undirected edge $(u, v) \in E$ exists if investor u reports a holding in company v, with the edge weight w_{uv} corresponding to the total market value of that specific position.

Implementation involved loading the dataset into a Pandas DataFrame and performing data cleaning. I merged duplicate company entries using CUSIP (company id) identifiers and filtered the data to the top 5,000 companies by total market value to ensure that the network captured the most influential nodes as well as to reduce network size, resulting in the bipartite graph G, consisting of approximately 11,800 nodes built using the NetworkX Python package [8]. Network size was reduced to 5,000 companies because calculating metrics like betweenness centrality has a high computational cost as we move through each "pair" of nodes. Thus, for a full graph, it would be too computationally expensive. G was then used to create the degree distributions and calculate metrics. A filtered (top 100) visualization of G is shown in Figure 1.

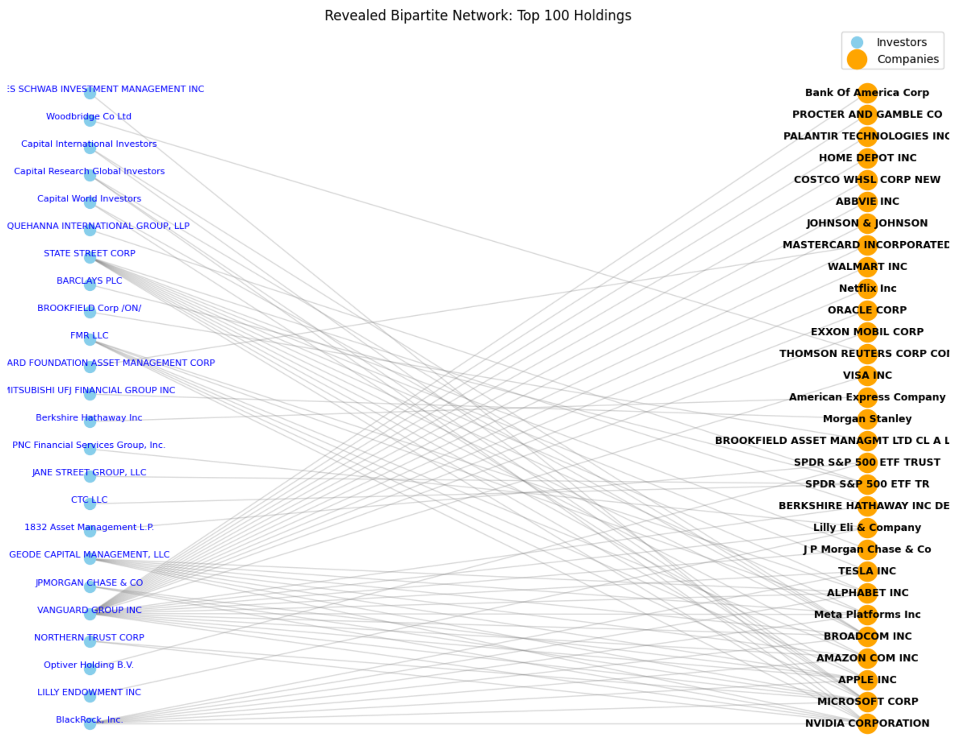

Figure 1: Bipartite graph with investors (left), companies invested in (right), and edges weighted by market value.

3.2 Methods & Algorithms

3.2.1 Degree Distribution

I first calculated the degree of each node to characterize the global architecture. For a company, degree represents the total number of unique institutional holder, so I plotted the degree distribution (Figure 2) using a log-log scale to verify a broad-scale or heavy-tailed pattern. This may confirm the existence of "hub" companies with massive institutional backing versus a long tail of less connected firms.

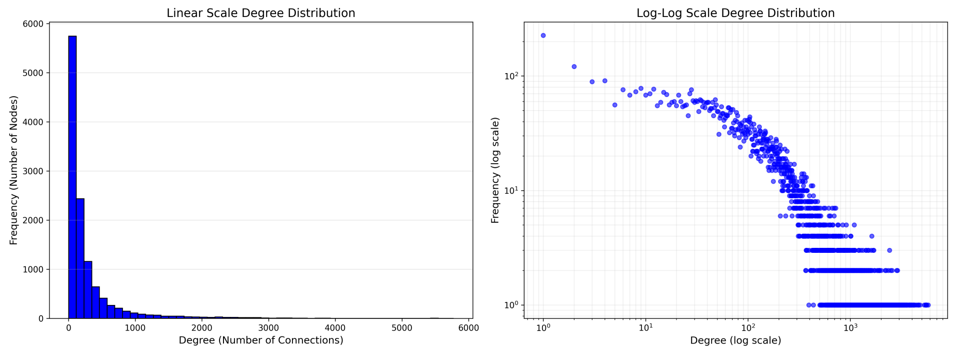

Figure 2: Degree distributions with a linear scale (left) and log-log scale (right) of the bipartite graph G, with investors and companies as nodes.

3.2.2 HITS

Instead of just counting connections, the HITS algorithm uses a recursive logic to rank importance. It consists of the following elements:

- Hub Score (Investors): This measures the "quality" of an investor based on the companies they hold. An investor has a high Hub score if they invest in many high quality (high Authority) companies.
- Authority Score (Companies): This is the primary proxy for success. A company has a high Authority score if it is held by many influential (high Hub) investors.

A high Authority score signifies that a company has the strongest investors from the market's most prominent investors.

3.2.3 PageRank

PageRank is an eigenvector centrality algorithm that measures the importance of a node based on the quantity and quality of links pointing to it. In a bipartite context, it treats the network as a Markov chain where the value of an investor is distributed among the companies they hold. Google's PageRank equation is shown below [7]:

$$r = (1 - \beta) \sum_{i \to j} \frac{r_i}{d_i} + \beta \frac{1}{n}$$

where β is a probability parameter, n is the total number of nodes, and d_i is the out-degree of node i. This represents a model where a random walker follows an edge with probability $1 - \beta$ and "teleports" to a random node with probability β. In other words, the value of a link from node u to v depends on the value of links going into u [6].

3.2.4 Clustering Coefficient

Because standard clustering coefficients are zero for bipartite graphs, I use the bipartite clustering coefficient. This measure assesses the overlap of neighbors between nodes of the same set. For a node $u \in U$, it is defined based on the number of common neighbors. This metric quantifies how much the neighbors of a node (investors) are interconnected through other nodes in the opposite set (companies).

3.2.5 Betweenness Centrality

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes. It is defined from [6] as:

$$C_B(i) = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}}$$

where g_{jk} is the total number of shortest paths from node j to k, and $g_{jk}(i)$ is the number of those paths that pass through i. In a bipartite investment graph, this metric identifies investors or companies that act as critical connectors within the capital flow, controlling the "shortest paths" of information or resource exchange.

3.2.6 Tools

1. Python
2. Pandas: Used for dataset processing and CUSIP standardization.
3. Matplotlib: Used for generating degree distribution plots and network visualizations.
4. NetworkX: Core Python package for complex network analysis.

4 Results

Following the implementation of the bipartite graph G, the results were analyzed through mapping and centrality metrics as discussed in the Methodology section in order to address the research questions defined at the start of this report.

4.1 Q1: Structural Identification of Market Leaders

The first research question asks whether a bipartite network's structural features can accurately identify market leaders without using their financial history. To test this, I performed a blind analysis using the HITS Authority and PageRank algorithms. This methodology allows for an assessment of a company's importance based solely on the "quality" of the investors who hold its shares.

The results in Table 1 show that the HITS algorithm correctly identified Microsoft, NVIDIA, and Apple as the top three authority nodes. In a bipartite investment network, an authority score is not just a count of investors (degree), but a measure of being held by hub investors who themselves have high quality investments. The high authority scores suggest that these companies act as the primary nodes for the most influential institutional capital as of September 2025 (date of this data). This makes sense as top technology and AI companies are expected to have the biggest impact in the more recent AI era. Because large hub investors prioritize companies that are relevant to advances in this sector, these companies naturally emerge as the most central authorities in our graph. Their high scores, thus, indicates that they are the primary nodes through which global capital flows into this economy.

Company (Revealed)	**Authority Score**	**PageRank**
MICROSOFT CORP	0.0105	0.0161
NVIDIA CORPORATION	0.0102	0.0180
APPLE INC	0.0098	0.0127
AMAZON COM INC	0.0094	0.0097
ALPHABET INC	0.0090	0.0086
META PLATFORMS INC	0.0087	0.0075
BROADCOM INC	0.0082	0.0064
TESLA INC	0.0069	0.0058
JP MORGAN CHASE & CO	0.0063	0.0038
VISA INC	0.0060	0.0036

Table 1: Blind Analysis Results: Top 10 Companies by authority scores calculated the HITS algorithm. Associated PageRank scores are shown for these companies.

While NVIDIA possesses the highest PageRank among companies, Microsoft exhibits the highest Authority score. This suggests that while NVIDIA is more broadly connected, Microsoft is held by a more valuable group of hubs, making it mathematically the better authority in the blind analysis. The high overlap of top nodes across authority and PageRank metrics also suggests that market value is generated and invested in other top nodes, making "successful" companies even more successful.

This confirms that the topology of the investment network serves as a reliable, independent way to measure market success, validating my hypothesis of the "blind" predictive power of network science.

4.2 Q2: Investor Centrality and the Certification Effect

The second research question explores the correlation between an investor's centrality and the success of the companies it holds, which in this case is the "authority" score of their holdings. As described previously, this relationship is related to the "Certification Effect" proposed by Bajo et al. [2], which suggests that the presence of a central investor acts as a signal of quality to the broader market.

Company / Investor Name	**PageRank Score**
VANGUARD GROUP INC	0.079998
STATE STREET CORP	0.033531
BlackRock, Inc.	0.024248
NVIDIA CORPORATION	0.018014
MICROSOFT CORP	0.016121
JPMORGAN CHASE & CO	0.014841
GEODE CAPITAL MANAGEMENT, LLC	0.013701
FMR LLC	0.012975
APPLE INC	0.012651
Capital World Investors	0.011541

Table 2: Top 10 Nodes by PageRank Centrality. Results reflect the mixed bipartite nodes (investors and companies) from the PageRank implementation in the code.

The PageRank analysis in Table 2 identifies large institutional hubs, specifically Vanguard Group Inc (0.0799) and State Street (0.0335), as the highest score nodes in the system. The difference between these top hubs and the rest of the network is further illustrated by the heavy-tailed degree distribution (Figure 2). In this 2025 dataset, it is noted that a small number of "large," or successful, investors are connected to nearly every high-authority company as seen partially in Figure 1.

This structure provides empirical evidence for the Certification Effect [2]: the quality of a node like Vanguard flows through the network edges to the companies it holds. Because these central hubs possess such a high PageRank score, their investment link carries significantly more weight than that of another investor. Consequently, the authority of a company like Microsoft is structurally dependent on the incoming value of links of these central connectors. This indicates that a company's structural importance is increasingly determined by its proximity (in terms of the graph) to a concentrated group of investors.

4.3 Q3: Identification of Structural Connections

The final research question identifies companies that may be vulnerable to a "downfall" because they share the same group of owners or investors.

As noted in Table 3, companies like Microsoft, Amazon, Apple, and NVIDIA (all large technology companies), have the highest betweenness scores in the network from 0.00975 to 0.00831. In the context of a bipartite investment network, betweenness measures the extent to which a node lies on the shortest paths between different clusters of investors and companies. Thus, the positions of these top technology companies suggests that these market leaders act as a primary structural bridge for the global economy. In other words, they connect or link thousands of different investors to the rest of the market.

Company Name	**Betweenness Score**
MICROSOFT CORP	0.009748
AMAZON COM INC	0.009511
APPLE INC	0.009163
NVIDIA CORPORATION	0.008306
ALPHABET INC	0.007787
BROADCOM INC	0.006752
Netflix Inc	0.006435
Meta Platforms Inc	0.006058
PHILLIPS 66	0.005662
COUPANG INC	0.005635

Table 3: Top 10 Companies by Betweenness Centrality. Betweenness identifies companies that connect different investor clusters.

On the other hand, high bipartite clustering coefficients for companies as in Table 4 reveal a different type of risk for a company's "downfall." Companies like Ingram Micro and Icahn Enterprises have perfect clustering (1.000), which indicates that a company is held by a nearly identical set of investors. As discussed by Flori et al. [3], we know that this represents a significant downfall risk, where if these specific shared investors face an economic crisis, they may be forced to lose their positions all together. This would cause a rapid drop in the stock price of these highly clustered companies regardless of their individual performances.

Company Name	**Clustering Score**
INGRAM MICRO HLDG CORP COM	1.00000
ICAHN ENTERPRISES LP	1.00000
CVR ENERGY INC COM	1.00000
MCGRAW HILL INC	1.00000
TRIPLE FLAG PRECIOUS METAL	0.60000
SUNCOR ENERGY INC NEW	0.60000
AURORA INNOVATION INC CLASS A COM	0.50000
KEYSIGHT TECHNOLOGIES INC	0.46961
KIMCO RLTY CORP	0.46961
METTLER TOLEDO INTERNATIONAL	0.46961

Table 4: Top Companies by Bipartite Clustering Coefficient. High clustering scores indicates that these companies are held by a nearly identical set of investors.

By identifying high betweenness companies like Microsoft and Amazon as structural bridges, it reveals the companies that are most important for maintaining market value across different investor groups. On the other hand, the high clustering scores for companies like Ingram Micro and Icahn Enterprises highlight a "downfall" risk, where a lack of investor diversity makes these companies vulnerable to fall simultaneously during a market shock.

5 Discussion

The results of this study demonstrate that network science provides a powerful method for understanding the financial landscape. By using HITS and PageRank, we proved that a company's structural position within an investment network is an accurate indicator of its true success, measured in market value. The dominance of technology and AI companies found through these algorithms were expected but further confirms that institutional capital is concentrated around the infrastructure of present trends. This suggests that in modern markets, the topology of capital allocation is not random, but rather forms a deliberate hierarchy where economic success is mirrored by network centrality.

A primary finding of this study is the validation of the "Certification Effect" [2]. By comparing our PageRank and HITS results, it is clear that a company's authority is directly driven by its proximity to a small group of valuable hub investors like Vanguard and BlackRock, which implies that the value of a company is increasingly determined not just by what it produces, but by the value its investors. This finding aligns with the theories proposed by Bajo et al. [2], yet our data shows a much higher degree of centralization than was present in the 2013 datasets. We also saw that in the modern AI-driven economy, capital is significantly more concentrated around a few key technological nodes, reinforcing a system where the "downfall" of a company is often a result of its structural neighbors rather than its own performance.

One limitation of this study is the static nature of the SEC 13F filing data, which provides a quarterly snapshot rather than a real time view of capital flow. Also, while the blind analysis successfully identified market leaders, the dataset lacks specific sector identifiers. In our bipartite network, a connection is treated equally whether it is in the tech, energy, or healthcare domains, which makes it difficult to determine if a node's high betweenness is due to its global importance or merely because it is the only large company in a niche sector. Thus, future research could implement bipartite graphs with different sector layers where one could analyze how capital flows between these industries. Because this study used a static view of the top 5000 nodes, incorporating temporal network analysis of all nodes may uncover hidden patterns as well.

References

[1] Del Giudice, A., Foglia, M., and Gianfrancesco, I. (2025). *Investor networks and social innovation: A stakeholder network analysis of Social Impact Bonds*. International Review of Economics & Finance, 99, 103999. https://doi.org/10.1016/j.iref.2025.103999 (Accessed: 19 Dec 2025).

[2] Bajo, E., Croci, E., and Marinelli, N. (2020). *Institutional investor networks and firm performance*. Journal of Business Finance & Accounting, 47(1-2), 1-35. https://doi.org/10.1111/jbfa.12423 (Accessed: 19 Dec 2025).

[3] Flori, A., Lillo, F., Spelta, A., and Pammolli, F. (2021). *Better to stay apart: asset commonality, bipartite network centrality, and investment strategies*. Annals of Operations Research, 299(1), 177-213. https://arxiv.org/abs/1811.01624 (Accessed: 20 Dec 2025).

[4] Balogh, A. (2023). *Layline Institutional Holding Reports*. Harvard Dataverse. https://doi.org/10.7910/DVN/TZM1QT (Accessed: 23 Dec 2025).

[5] Balogh, A. (2023). *Mapping the common ownership of public firms*. Working paper. (Accessed: 23 Dec 2025).

[6] Zhou, S. (2025). *COMP0123: Complex Networks and Online Social Networks Lecture Slides*. University College London. (Accessed: 30 Dec 2025).

[7] Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems*, 30(1-7), 107-117.

[8] Hagberg, A., Schult, D., and Swart, P. (2008). *Exploring Network Structure, Dynamics, and Function using NetworkX*. Proceedings of the 7th Python in Science Conference (SciPy2008), 11-15.

Code Implemenation: https://github.com/mk1m/Network-Science/blob/main/network_data.ipynb