**Laplacian of a Graph and Application in Analyzing Relationships between AI Stocks and Other Stocks in the Stock Market**

In recent years, the AI sector has witnessed tremendous growth, with NVIDIA (NVDA) emerging as a dominant player. With a market capitalization of $2.9 trillion and a P/E ratio of 69, NVIDIA represents a significant portion of the AI market, reflecting high investor expectations for its future value. This study aims to analyze the relationship between NVIDIA, as a representative AI stock, and other stocks across different sectors, using graph-based methods and advanced modeling techniques.

**Data and Methodology**

We utilized stock data from NVIDIA and 69 other companies, covering sectors such as technology, semiconductors, manufacturing, gaming, energy, biotechnology, and banking. The dataset includes stock prices and financial indicators such as P/E ratio, P/B ratio, and Beta from 2018 to June 2024.

To uncover the relationships between these stocks, we employed a two-phase approach:

1. **Graph Visualization**: We first visualized the connections between NVIDIA and other stocks using a simple graph representation.
2. **Advanced Non-linearity Graph Models**: We applied non-linear graph models to analyze the relationships in more detail.

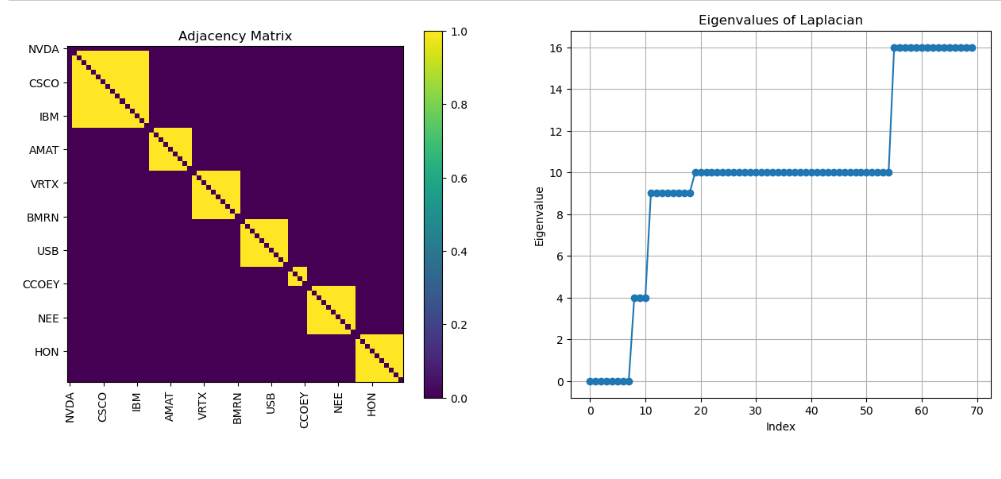
**Part 1 Graph Visualization**

Initially, we created a graph to show the connections between NVIDIA (green node) and other companies (orange nodes). The edges represent potential relationships based on sector classification.

A black and orange lines and dots

Description automatically generated

I then used the GCN model to classify stock groups. The results were quite positive when Test Accuracy score = 0.9286.



Tại hình fig.2 bên trên

#### Adjacency Matrix

The adjacency matrix (left plot) visualizes the connectivity between the nodes (stocks) in the graph. Each cell in the matrix represents the presence (1) or absence (0) of an edge between two nodes. The diagonal elements are always 1 because each node is connected to itself.

* **NVIDIA (NVDA)**: The row corresponding to NVIDIA shows how it is connected to other stocks. If there are yellow cells in this row/column, it indicates that NVIDIA shares a direct connection with those stocks.
* **Sector Clustering**: The blocks of yellow off-diagonal elements indicate strong intra-sector connectivity. For instance, clusters such as IBM, AMAT, VRTX, and others suggest that these stocks are more densely connected within their sectors.
* **Sparse Connections**: There are also sectors with sparse connections, indicating weaker direct relationships between stocks from different sectors.

#### Eigenvalues of the Laplacian

The eigenvalues of the Laplacian matrix (right plot) provide insights into the graph's structural properties. The Laplacian matrix is defined as L=D−AL = D - AL=D−A. The eigenvalues are sorted in ascending order.

* **First Eigenvalue (0)**: The smallest eigenvalue is always 0, corresponding to the constant eigenvector (all ones vector). This indicates that the graph is connected.
* **Eigenvalue Gaps**: The gaps between eigenvalues can indicate the presence of clusters or communities within the graph. Larger gaps suggest that the corresponding eigenvectors can be used to distinguish between different clusters.
* **Multiple Zero Eigenvalues**: If there are multiple eigenvalues equal to zero (not the case here), it would indicate the graph is disconnected with several isolated components.
* **Spectral Clustering**: The number and distribution of eigenvalues can be used for spectral clustering to identify tightly-knit groups within the graph.

#### Embedding of GNN

The final embedding visualization (bottom plot) shows the transformed representation of the nodes (stocks) in a lower-dimensional space using the OptimalMP\_Laplacian\_Phi model. The closer the nodes are in this space, the stronger their relationships.

* **NVIDIA (NVDA) Embedding**: NVIDIA's embedding value is shown in green. It is located near other major technology stocks (e.g., MSFT, GOOGL, AMZN), indicating strong relationships with these stocks.
* **Cluster Formation**: Stocks within the same sector form distinct clusters, showcasing intra-sector relationships.
* **Embedding Values**: The range of embedding values indicates the strength and nature of relationships between stocks.

**Part 2 Using Non Linearity Models**

Use non-linear model (Gaussian), non-linear model (Laplacian), non-linear model with only feature transformation (Laplacian), non-linear model with only non-linear propagation (Laplacian) in turn.**Optimal Non-linear Model (Gaussian)**

The Optimal Non-linear Model (Gaussian) introduces non-linearity by applying a thresholding function on the feature transformations. This model uses a learned threshold to limit the values of the transformed features before adding the result of a linear transformation with the adjacency matrix multiplication.

**Key Components:**

* **Linear Transformation**: Transforms the input features using a linear layer.
* **Thresholding**: Applies a threshold to the transformed features to introduce non-linearity.
* **Adjacency Matrix Multiplication**: Adds the result of the thresholded features with the product of the adjacency matrix and the features.

**Optimal Non-linear Model (Laplacian)**

The Optimal Non-linear Model (Laplacian) enhances the previous model by introducing two separate thresholds for feature transformations. This model applies thresholding before and after the linear transformation, offering a more sophisticated non-linear propagation mechanism.

**Key Components:**

* **Linear Transformation**: Transforms the input features using a linear layer.
* **Double Thresholding**: Applies two thresholds to the transformed features, one before and one after the linear transformation.
* **Adjacency Matrix Multiplication**: Adds the result of the thresholded features with the product of the adjacency matrix and the features.

**Non-linear Model with Only Feature Transformation (Laplacian)**

This model introduces non-linearity solely through feature transformation by applying a threshold before the adjacency matrix multiplication. It focuses on transforming the features non-linearly without altering the propagation mechanism.

**Key Components:**

* **Linear Transformation**: Transforms the input features using a linear layer.
* **Thresholding**: Applies a threshold to the transformed features to introduce non-linearity.
* **Adjacency Matrix Multiplication**: Adds the result of the thresholded features with the product of the adjacency matrix and the features.

**Non-linear Model with Only Non-linear Propagation (Laplacian)**

This model introduces non-linearity solely through the propagation mechanism by applying a threshold after the linear transformation. It focuses on non-linear propagation without altering the initial feature transformation.

**Key Components:**

* **Linear Transformation**: Transforms the input features using a linear layer.
* **Thresholding**: Applies a threshold to the features after the linear transformation to introduce non-linearity.
* **Adjacency Matrix Multiplication**: Adds the result of the thresholded features with the product of the adjacency matrix and the features.

Mô hình non-linear model with only non-linear propagation (Laplacian)( OptimalMP\_Laplacian\_Phi) cho ra kết quả không quá tệ trái ngược với dự đoán ban đầu của tôi. Từ trọng số có được từ mô hình ta thấy

Weights from OptimalMP\_Laplacian\_Phi(Laplacian): [[-0.9260769 1.1825187 0.89947814 -0.15352716]]

**Feature 2 (P/E ratio):** This feature has the highest weight (1.1825187), indicating that the P/E ratio is the most important feature in determining the embedding value of the stocks. **Feature 3 (Stock Price):** This feature has the second highest weight (0.89947814), showing that stock price also plays a significant role. **Feature 1 (P/B ratio):** This feature has a negative weight (-0.9260769), indicating an inverse relationship with the embedding value. **Feature 4 (Beta):** This feature has the lowest weight (-0.15352716), showing that Beta is less important compared to the other features.

**P/E Ratio is the Most Important Feature:**

NVDA and the stocks with embedding values close to it (e.g., MSFT, GOOGL, AMZN) may have similar P/E ratios, suggesting that these companies could have relatively similar profitability compared to their stock prices.

**Stock Price is Also Important:**

Stock price also plays a significant role in determining the embedding value, which may suggest that companies with high stock prices tend to be similar in your model.

**Inverse Relationship with P/B Ratio:**

The negative weight of the P/B ratio indicates that companies with higher P/B ratios might have lower embedding values, and vice versa.

**Beta is Less Important:**

Beta has the lowest weight, indicating that stock volatility is not a major factor in determining the relationship between stocks in your model.

And after embedding we get the result as shown in fig.3 below

A chart with numbers and dots

Description automatically generated with medium confidence

The provided embedding visualization represents the 1D embeddings of various stocks, including NVIDIA, obtained from the OptimalMP\_Laplacian\_Phi model. The embedding values reflect how closely related each stock is to the other stocks in terms of their graph-based relationships.

#### Key Observations:

1. **NVIDIA (NVDA)**:
   * NVIDIA is clustered with other major technology stocks like Microsoft (MSFT), Google (GOOGL), Amazon (AMZN), and Apple (AAPL), which is indicated by their similar embedding values close to 1.0.
   * This suggests that NVIDIA shares a strong relationship with these tech giants, likely due to similarities in market behavior, investor sentiment, and sector-specific factors.
2. **Cluster Formation**:
   * The plot shows distinct clusters of stocks with similar embedding values, indicating strong intra-cluster relationships.
   * For example, one cluster includes stocks like ADBE, CRM, and CSCO, while another includes INTC, HPQ, and NFLX.
3. **Outliers**:
   * A few stocks are positioned away from the main clusters, indicating weaker or more distinct relationships. For instance, there is a stock with an embedding value around 0.4, which suggests a unique relationship with other stocks.
4. **Sector-Based Grouping**:
   * The colors representing different stocks also highlight sector-based groupings. Stocks from similar sectors tend to have closer embedding values, emphasizing the model’s ability to capture sector-specific relationships.

Reference:

*Fiedler Regularization: Learning Neural Networks with Graph Sparsity*-Edric Tam, David Dunson

*Understanding Non-linearity in Graph Neural Networks from the Perspective of Bayesian Inference-*Rongzhe Wei , Haoteng Yin , Junteng Jia , Austin R. Benson , Pan Li