Graph Neural Network (GNN)-Based Deep Learning for Optimised Corporate Reorganisation

Submission for

NUS Datathon 2026

Team RedBullz Airways

Ilyas Bin Mohd Izhar Khan Surattee
Lai Woh Jon
Lee Weng Heng, Jon
Lim Hong Jing, Luke

1 Introduction

Our project aims to develop a deep learning model prototype which conglomerates can consult for strategic planning with respect to corporate reorganisation, specifically in the aspect of internal entity mergers and acquisitions (M&A). This is done by leveraging data analytics, network modelling, deep learning techniques and large language models (LLMs) to analyse entities' different attributes and corporate relationships with one another. We aim for the model to help conglomerate-level decision-makers of Champions' Group overcome the complexity of their entities' corporate networks by prescribing optimal solutions on internal M&As for their consideration.

2 Champions Group Dataset Overview

The Champions Group dataset has 8559 rows and 72 columns, corresponding to each company's records and the features of each company respectively. The dataset contains **firmographic attributes** (e.g. ownership type and active status), **geographic fields** (e.g. addresses and regions of operation), **industry classifications** (e.g. SIC, NAICS and related systems), **business scale indicators** (e.g. total number of employees, revenue and market value in USD), **IT capacity indicators** (e.g. IT budget, IT spend and approximate counts of the various IT devices), and **corporate structure** (e.g. parent, global ultimate, domestic ultimate company relationships and corporate family size).

In the report, we will explore the relationships between these attributes and information regarding corporate structure will be used for network analysis and the creation of a graph neural network.

3 Methodology and Results

We started by performing exploratory data analysis (EDA) and generating a network graph on Gephi[1] to visualise the relationships between companies.

3.1 Exploratory Data Analysis (EDA)

3.1.1 Relevant attributes for our analysis

In our EDA, we want to determine which attributes are relevant for our analysis.

Figure 3.1 shows the percentage of missing values for each variable. We found that many variables have a high percentage of missing values. For example, the variables 'Ticker' and 'Registration Number Type' have close to 100% missing values.

[1] Gephi is a network graph visualisation platform.

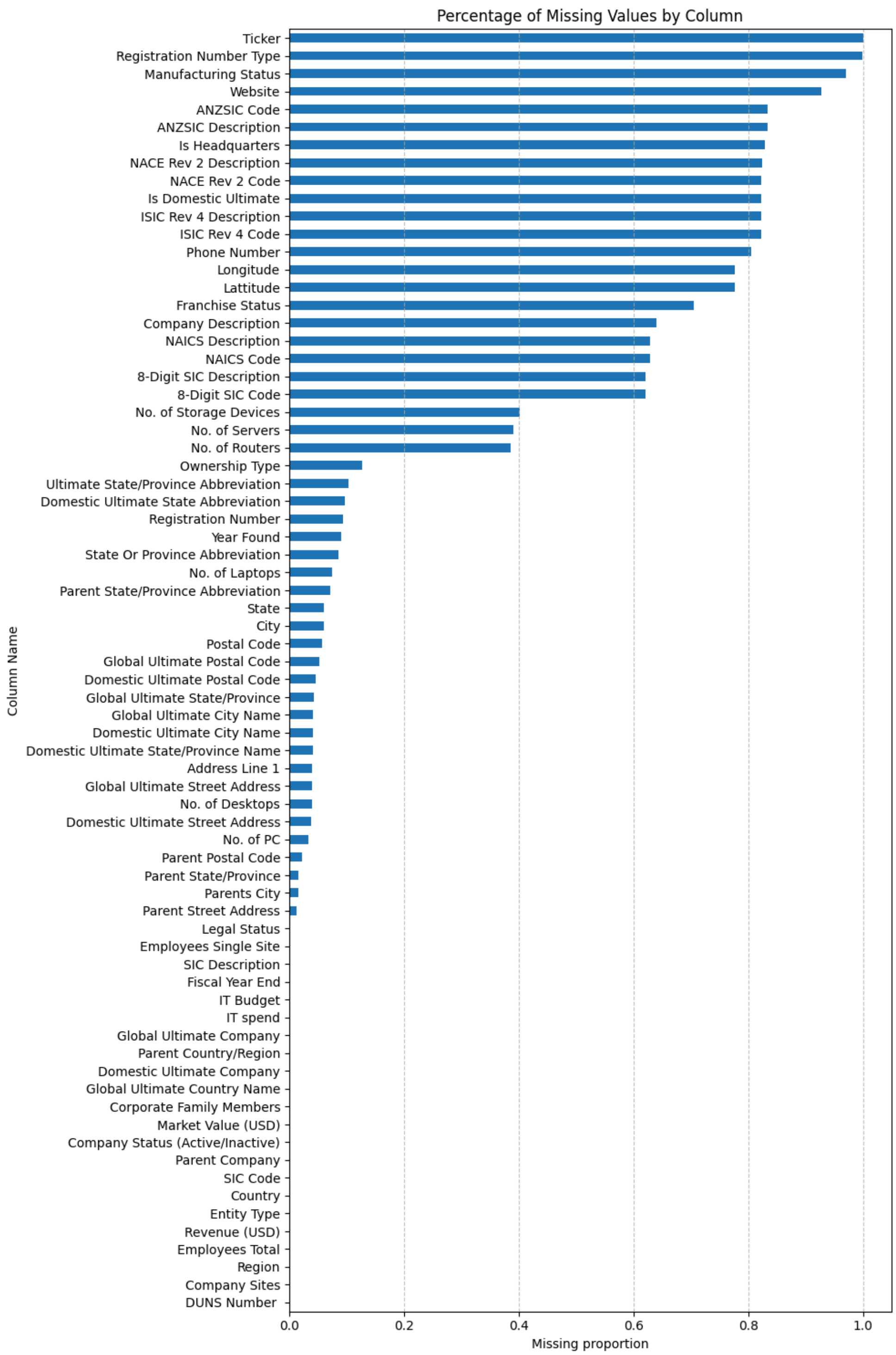

Figure 3.1: Percentage of Missing Values by Column

Additionally, while some variables have a low percentage of missing values, we drop some of these columns, which may not be relevant to our analysis. A full list of columns dropped, and our rationale for dropping each column can be found in **Annex A**.

We also used EDA to determine the relationships between key variables.

3.1.2 Relationship between Age (of a company) and Company Type

From our analysis, we removed an anomaly where Age = 2026 (this company has seen the rise and fall of empires). We then obtained the following plot:

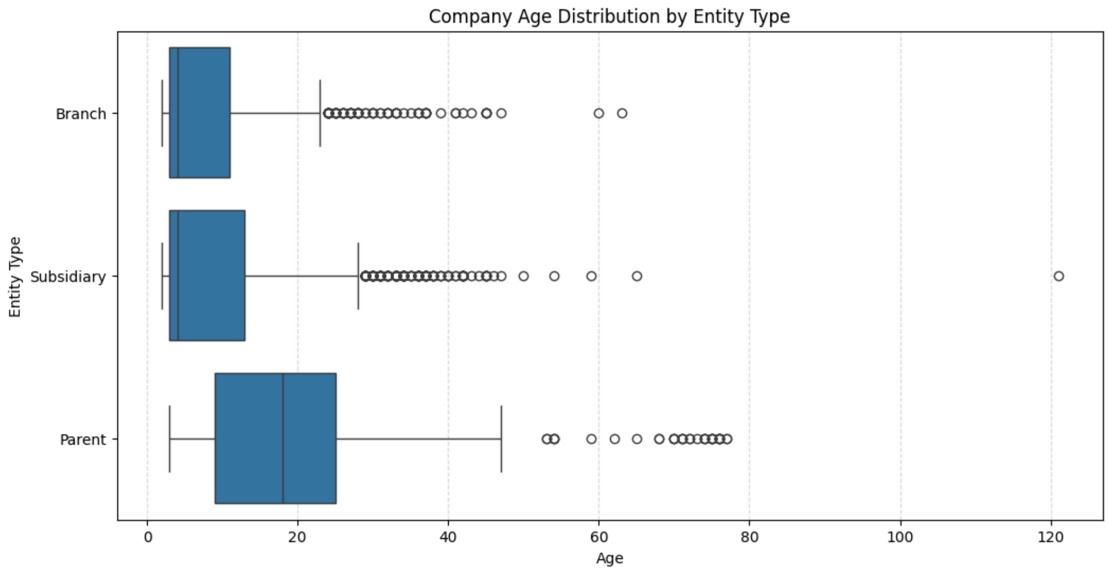

Figure 3.2: Company Age Distribution

From the above plot, **Parent** companies are generally much older than **Branch** and **Subsidiary** companies, which is in line with expectations.

3.1.3 Relationship between Market Value and Revenue

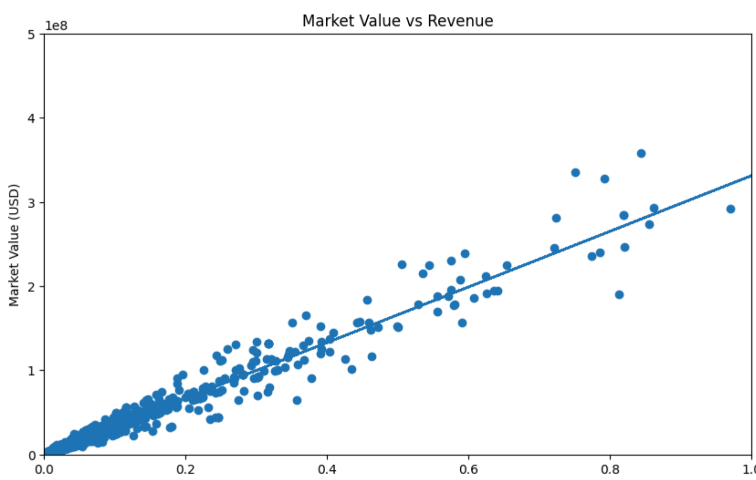

Figure 3.3: Plot of Market Value vs Revenue

Plotting market value (USD) against revenue (USD), we observed a strong positive relationship (R^2 = 0.9725), which is in line with expectations. This strong positive relationship is also reflected in the business attributes PCA biplot (Figure 3.13).

Profit = Revenue - cost. The relatively variance of the graph shows that revenue generally plays a more important role in determining the market value of a company compared to business cost.

3.1.4 Relationship between Revenue and IT Spending

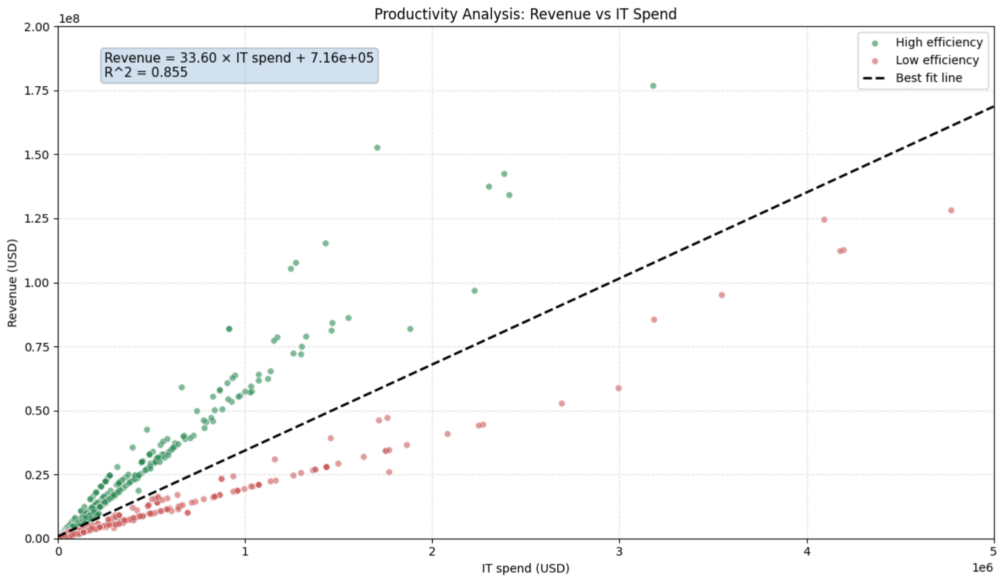

Figure 3.4: Plot of Revenue against IT Spending

Plotting revenue (USD) against IT spending (USD), we observed a clear positive relationship. However, the R^2 value is not very good. There seems to be at least two categories of companies, each with different revenue responses to IT spending.

By grouping points above the best-fit line into high responders (we call it high “efficiency”) and points below the best-fit line into low responders, we can verify this clustering behaviour, and this is visually reflected in the plot above.

3.1.5 Market Value Distribution by Company Age Group

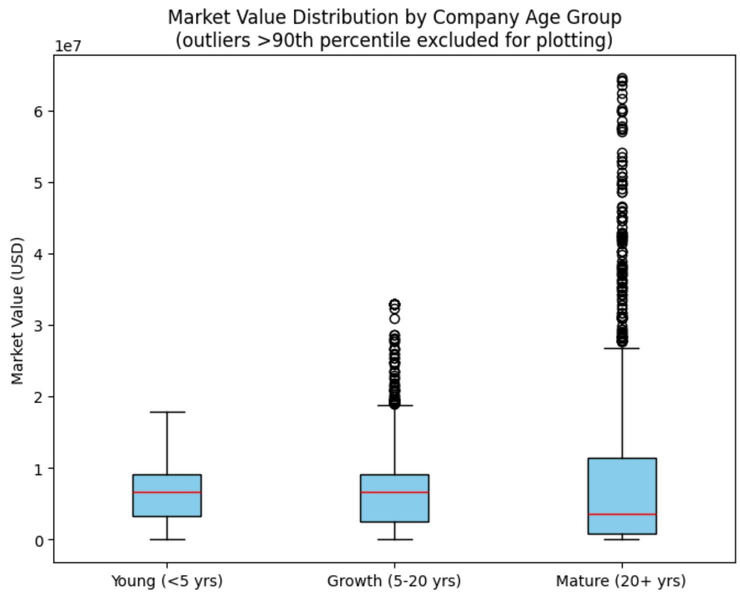

Figure 3.5: Market Value Distribution by Company Age Group

Age Group	Company Count	Median	P25	P75	Variance
Young (<5 yrs)	1716	6,675,812	3,828,020	10,094,246	2.17×10^{14}
Growth (5-20 yrs)	1336	6,675,812	2,883,551	10,873,378	2.97×10^{15}
Mature (20+ yrs)	1169	4,663,053	980,982	18,846,145	1.49×10^{16}

There is no difference between the median market value (MV) of young and growth companies, while mature companies exhibit a lower median MV than both. At the same time, the variance of MV increases substantially with firm age, with mature firms displaying far greater dispersion than younger firms.

The fact that the median mature firm is smaller than the median young or growth firm suggests that many firms survive without ever scaling meaningfully and may gradually stagnate over time. Survival, therefore, does not imply growth.

However, firm outcomes diverge dramatically as firms age. Over longer horizons, small initial advantages can compound, leading to highly unequal outcomes. This creates room for winner-takes-most dynamics, where a small subset of mature firms grows disproportionately large while the majority remain small.

When we examine the mean market value, this divergence becomes even clearer. The mean MV of mature companies (31.3 million USD) is substantially higher than that of young (10.3 million USD) and growth firms (1.65 million USD). This gap between the mean and the median indicates a highly right-skewed distribution for mature firms, where extreme outliers dominate aggregate value creation.

For a venture capitalist seeking to diversify across firms of different age groups, investing in mature companies may still be attractive despite their lower median market values. The substantially higher variance and mean MV suggest a greater probability of exposure to extreme positive outliers, which can significantly increase the expected value of a diversified portfolio.

In contrast, younger firms appear to offer more homogeneous outcomes with less upside dispersion, implying lower tail risk but also fewer opportunities for transformational returns.

3.1.6 IT Budget against IT Spending

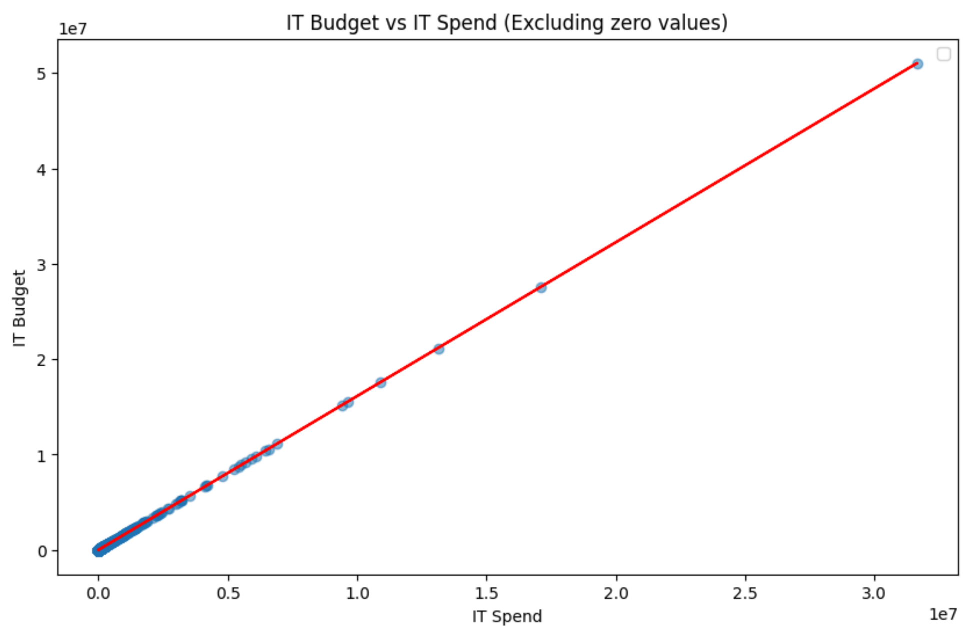

Figure 3.6: Plot of IT Budget against IT Spending

Plotting IT Budget (USD) against IT spending (USD), we observe a perfect correlation ($R^2 = 1$). Hence, it makes sense to drop one column for our analysis. In this case, we drop the IT Budget variable.

3.1.7 Revenue per Employee

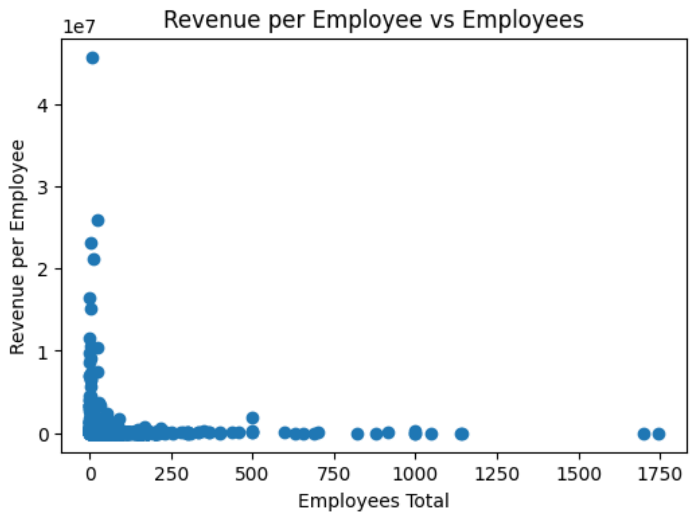

Figure 3.7: Plot of Revenue per Employee against Number of Employees

Generally, the higher the total number of employees in a company, the lower the revenue per employee. This suggests that there is “friction” associated with scaling, where organisational complexity and overhead often dilute individual output.

The left side of the graph (low number of employees) has a “spike”, indicating that some companies have very few employees and high revenue per employee. These high-ratio outliers suggest that companies generally achieve high monetisation with smaller teams, and these companies could be attractive acquisition targets.

3.2 Network Analysis

We used Gephi to visualise relationships between entities and calculate different network parameters (e.g. indegree, outdegree). We assume that the nodes represent individual companies and the directed edges represent ownership/control relationships. The directional edges point from the parent entity → child entity.

We then merged our network graph with the original dataset (champions_group_data.xlsx) after dropping irrelevant columns to create a new dataset (companies_merged.xlsx).

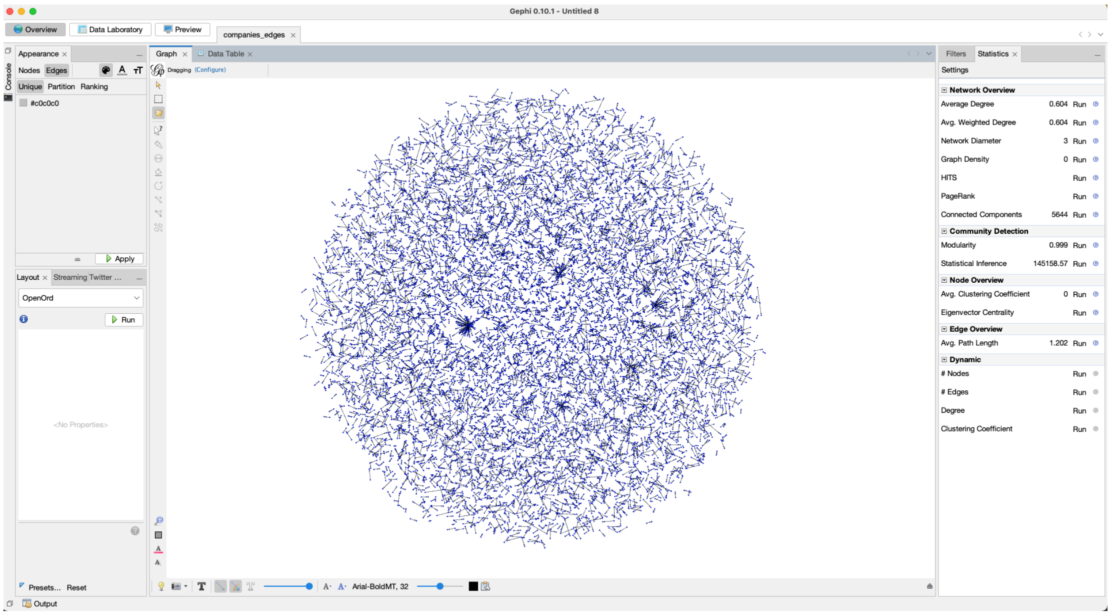

Figure 3.8: Network visualisation of relationships between Champions Group entities

The plot looks really complex from here, but notice the presence of nexuses and "hub and spoke" features in the network. These features correspond to components of companies linked together through ownership and control relationships. Exploring these connections provides us a visual way to chart out the connection between companies.

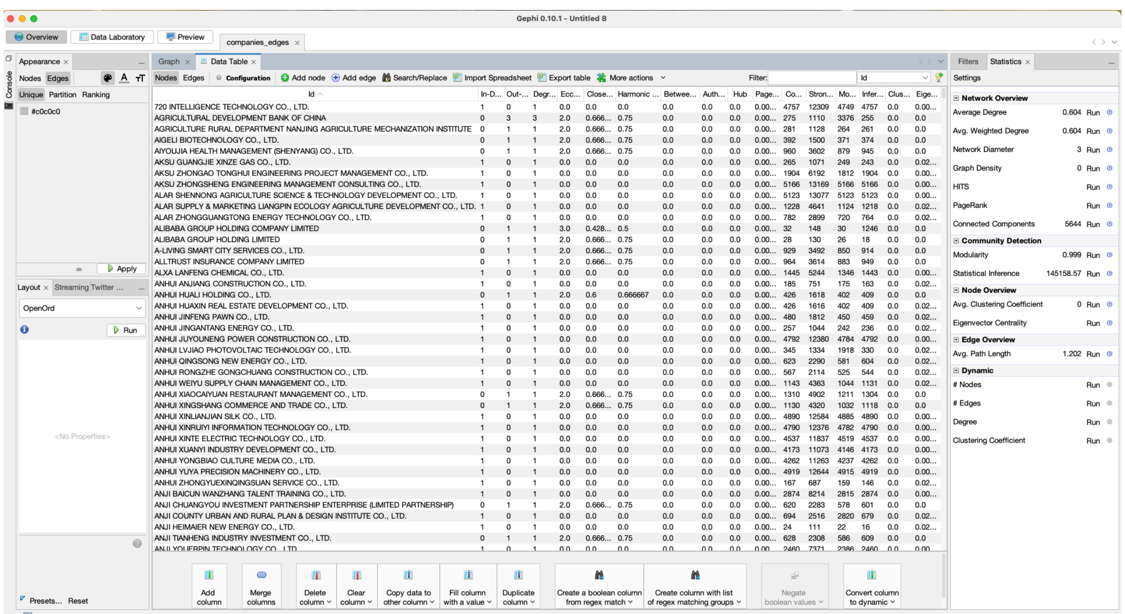

Figure 3.9: Network data on Champions Group entities as computed by Gephi

A full list of the quantities of interest which we have defined is in **Annex B**.

3.2.1 Detection of circular ownership

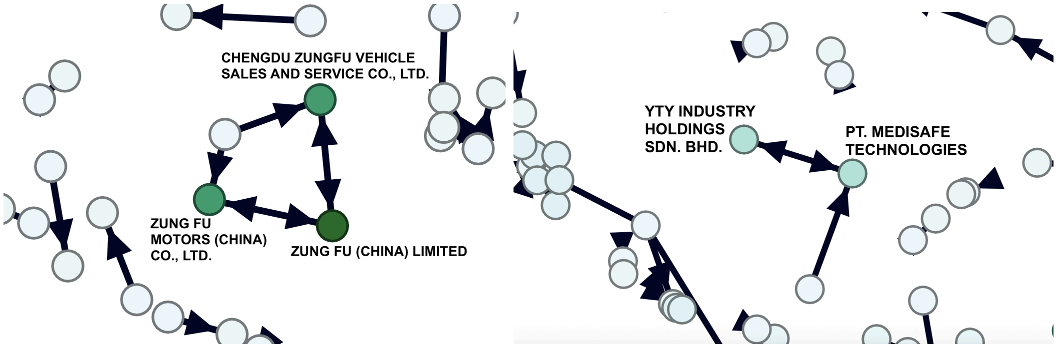

Figure 3.10: Evidence of circular ownership among sub-conglomerates in our dataset. Arrows are directed from parent to subsidiary.

In addition, the use of network analysis on our data has allowed us to detect anomalies such as the circular ownership of entities within some sub-conglomerates. By defining circular ownership as sub-conglomerates with an strongly connected component (SCC) size of at least 2 (i.e. 2 or more companies share the same SCC ID), we can rely on processed data such as the eigenvector centrality and SCC ID of the vertices to detect the presence of cycles. These cycles in our ownership network of corporate entities thus implies the presence of additional subsidiary ownership in parent companies.

Identifying circular ownership might be of interest to regulators due to the potential ambiguities regarding ownership, tax regulations etc.

3.3 Principal Component Analysis (PCA)

We performed PCA as there are many potentially overlapping numeric variables (e.g. variables highly correlated with one another). We want to reduce the dataset's high dimensionality. With the presence of many variables, it is difficult to discern which variable(s) are actually driving variance in the dataset (e.g. differences between various companies), especially if there exist attributes irrelevant for data analysis.

3.3.1 Scree Plot

We begin by creating a scree plot. A scree plot is used to determine how many components to keep, plotting eigenvalues (variance explained) against component number. Plotting a scree plot reveals an "elbow" (i.e. the point where the curve flattens), indicating diminishing returns. This suggests that components before the elbow are most significant in contributing to the variance observed in the dataset.

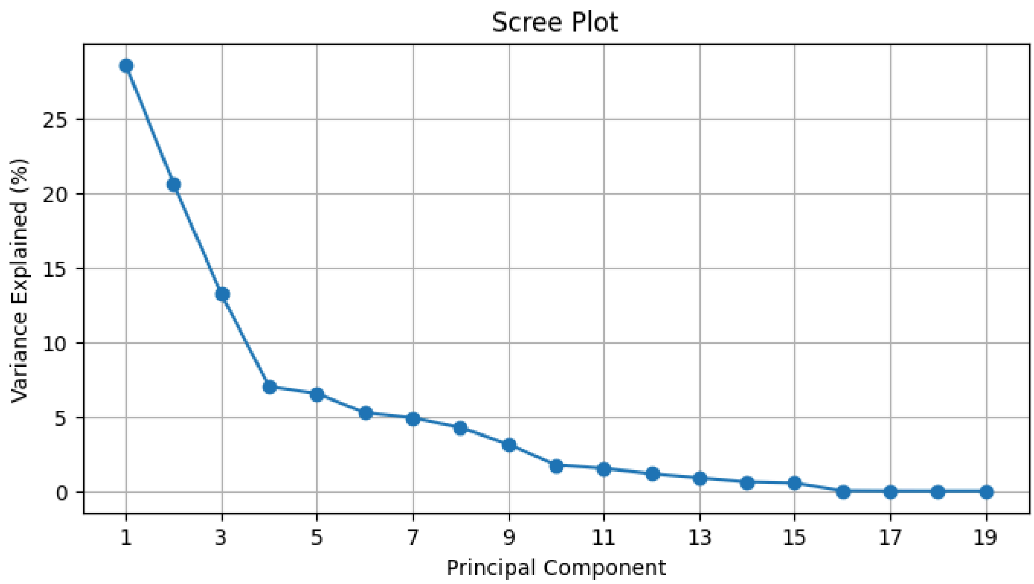

Figure 3.11: Scree Plot Analysis

From **Figure 3.10**, we obtain the following values for the percentage of variance explained by each principal component:

- PC1 = 28.57%
- PC2 = 20.59% and cumulatively, PC1+PC2 = 49.16%
- PC3 = 13.20% and cumulatively, PC1+PC2+PC3 = 62.36%
- PC4 = 7.00% and cumulatively, PC1+PC2+PC3+PC4 = 69.36%

This implies that in our dataset, the variance is not dominated by just one or two dimensions, since PC1 and PC2 cumulatively account for only ~50% of the dataset's total variance. Hence, a 2D PCA plot shows only about half of the structure in the data.

Instead, the dataset has multiple underlying factors driving differences across companies. This means that companies do not differ in only one main way. Differences are spread across several factors, suggesting that there are multiple independent reasons companies look different, not just based on one single ranking.

We can observe an “elbow” at PC4, showing a clear drop-off past PC4. To elaborate, after the point of PC4, each extra component adds relatively little to the variance in the dataset, where the majority of the variance (~69% in the dataset) is attributed to PC1-PC4. Hence, based on our analysis, it is reasonable to keep 3-4 components for our analysis.

Overall, the scree plot suggests multidimensional differences between companies, not dominated by a single factor (since PC1 accounts for only around 28.57% of the dataset's variance).

3.3.2 Biplot for PC1 and PC2

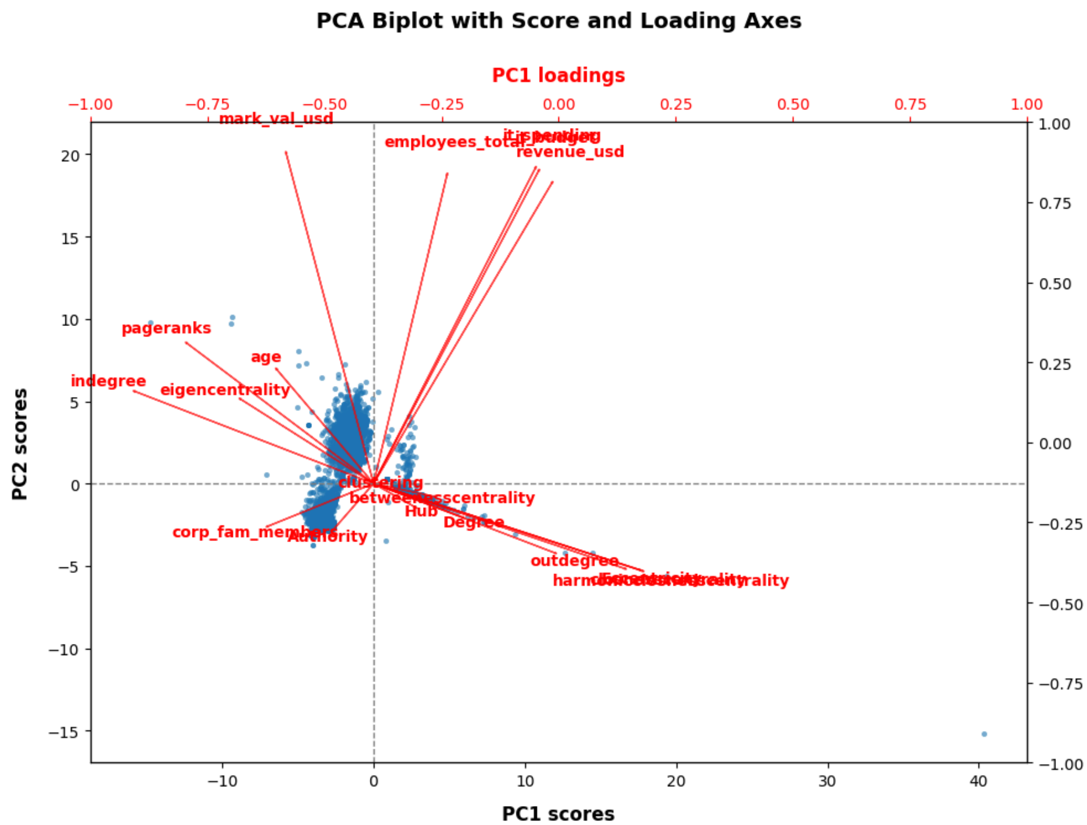

Figure 3.12: Biplot for PC1 and PC2

From **Figure 3.11**, we obtain the loading values for PC1 and PC2, indicating which variables matter the most for PC1 and PC2.

PC1 is primarily driven by distance-related network measures, with the largest positive loadings coming from the following 3 variables:

- harmonic closeness centrality (≈ +0.386)
- closeness centrality (≈ +0.382)
- eccentricity (≈ +0.361)

and also positive contributions from outdegree and several firm size/spend variables.

PC1 can be interpreted as separating companies that appear more “reachable” in the network (i.e. high (harmonic) closeness and related distance structure) from companies characterised more by incoming importance (corresponding to the variables of high indegree, PageRank and eigencentrality). Companies with high PC1 scores tend to align with the former, while low PC1 scores align with the latter.

PC2 is dominated by business scale variables, with large positive loadings on:

- IT spending (≈ +0.416)

- IT budget (≈ +0.413)
- employees_total (≈ +0.408)
- revenue_usd (≈ +0.397)
- mark_val_usd (≈ +0.436)

PC2 can be interpreted as a clear company scale or financial capacity dimension. Companies with higher **PC2** scores are larger organisations with higher revenue, market value, headcount, and IT spend/budget.

We also want to determine the heavyweight variables in the **PC1-PC2 plane**. To help us understand which variables most strongly influence the 2D PCA visualisation, we computed each variable's PC1-PC2 strength. The greatest contributors to the PC1-PC2 scores were the following:

- it_spending (≈ 0.477)
- it_budget (≈ 0.476)
- revenue_usd (≈ 0.472)
- mark_val_usd (≈ 0.454)
- employees_total (≈ 0.421)
- harmonic closeness centrality (≈ 0.403)
- closeness centrality (≈ 0.399)
- eccentricity (≈ 0.378)
- indegree (≈ 0.364)
- pageranks (≈ 0.327)

The first two components (**PC1** and **PC2**) provide a meaningful low-dimensional view of the data (**PC1** and **PC2** cumulatively capture a substantial ~50% of the overall variation). Interpreting the first two principal components, we examined the component loadings (weights) and the **PC1-PC2** plane strength, represented by the arrow length in the biplot.

This suggests that separation among companies in the **PC1-PC2** plane is driven primarily by firm size and spending capacity variables, and secondarily by a set of network position variables.

PC1 captures a "network position" contrast (reachability/distance-based centrality vs incoming importance/prestige), and **PC2** acts as a strong "size/spend" axis (big firms vs small firms).

This provides a compact way to compare companies along two interpretable dimensions - scale/spending capacity and network position type, useful for segmentation and identifying unusual firms (outliers) or distinct groups.

3.3.3 Biplot for PC3 and PC4

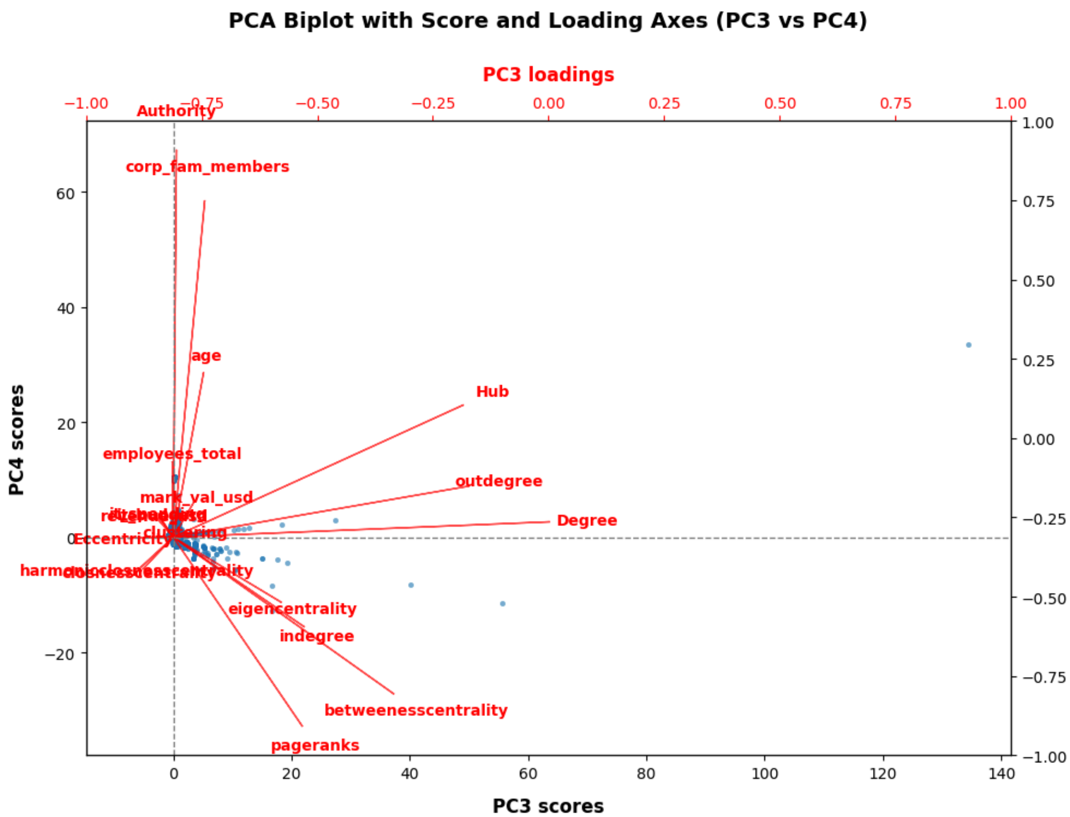

Figure 3.13: Biplot for PC3 and PC4

For **PC3**, the highest absolute loadings include the following:

- Degree = +0.584
- outdegree = +0.459
- Hub = +0.450
- betweenness = +0.343
- (smaller) indegree/pagerank around +0.20

The variables of closeness, harmonic and eccentricity have small absolute loadings on **PC3**. From this, we can derive that **PC3** is a connectivity dimension. This means that a high **PC3** score can be used to imply companies with many links (degree/outdegree) behave like hubs (Hub), and often sit on paths between others (betweenness), whereas low **PC3** loadings would imply less connected nodes.

For **PC4**, the highest absolute loadings include the following:

- Authority = +0.617
- corp_fam_members = +0.536
- age = +0.263
- Hub = +0.211

The strongest negatives for PC4 include the following:

- pageranks = −0.302
- betweenness = −0.250
- indegree = −0.143

PC4 can be interpreted as separating authority, corporate-family and age characteristics from PageRank, betweenness-type and network influence characteristics. A high **PC4** loading score implies that companies score high on the Authority dimension, while a low **PC4** loading score indicates companies characterised more by PageRank and betweenness (global influence/brokerage signals), rather than authority.

The **PC3** and **PC4** components are driven primarily by network structure metrics, rather than firm size or spending variables. For **PC3**, the largest positive loadings occur for Degree (0.584), outdegree (0.459), Hub (0.450), and betweenness centrality (0.343), indicating that **PC3** captures a connectivity dimension. Companies with high **PC3** scores tend to be highly connected nodes that act as hubs and frequently occupy bridging positions in the network.

For **PC4**, the strongest loading is Authority (0.617), with additional positive contributions from corporate family members (0.536) and age (0.263). Conversely, PageRank (−0.302) and betweenness centrality (−0.250) load negatively on **PC4**. This suggests that **PC4** captures a contrast between authority-style influence and related firm attributes (e.g. corporate family size, age) vs PageRank and betweenness-type network influence, reflecting different notions of prominence within the network.

Overall, the variables with the greatest influence in the PC3-PC4 plane are Authority, Degree, corporate family members, Hub, outdegree, and betweenness centrality, while business-scale variables (revenue, IT budget/spend, market value) contribute minimally to these components.

3.3.4 Biplot for Business Attributes (PC1 and PC2)

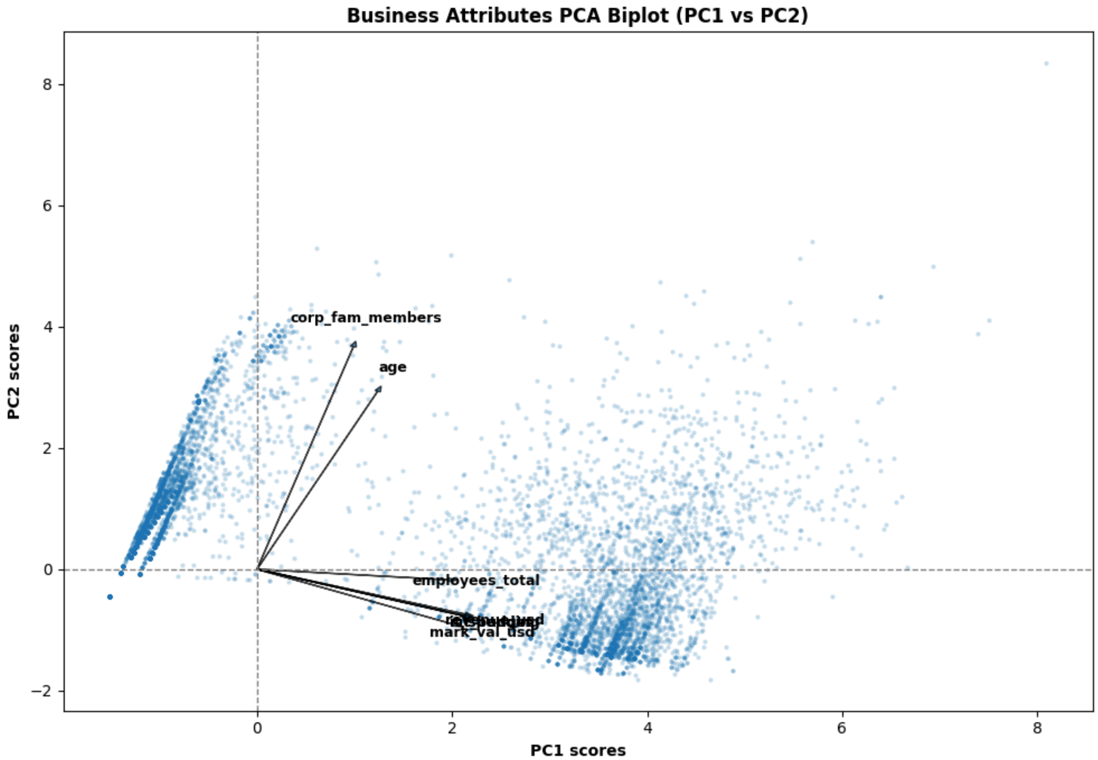

Figure 3.14: Business Attributes PCA Biplot (PC1 vs PC2)
Arrows that point to the bottom right are revenue_usd, mark_val_usd and it_budget

We can also perform PCA only using business attributes. We observe a very high correlation between IT spending and revenue (USD). This corroborates well with the revenue against IT spending plot we created, showing a strong positive relationship between these 2 features.

We also observe that certain features are correlated with PC1 (e.g. employees_total, it_spending, revenue_usd) and PC2 (e.g. age of the company and number of corporate family members), respectively.

Our interpretation of the PC1 and PC2 loadings is as follows:

1. **PC1 seems to be a “firm scale/economic size” axis.** On the right side, we have firms with high firm scales, while on the left, we have firms with much smaller firm scales. Bigger companies spend more on IT, employ more people and are worth more.
2. **PC2 seems to capture organisational maturity and ownership structure.** The high correlation between age and corporate family size is expected, as older firms would have had more time to acquire family/corporate members.

The two apparent clusters in the PCA biplot (one on the left and the other on the right) could be driven by differences in firm scale (IT spend, revenue etc). We decided to investigate the reason by only plotting the points that have efficiency values, which means these companies have both revenue and IT spend values (and these values are not empty).

The results obtained are represented below:

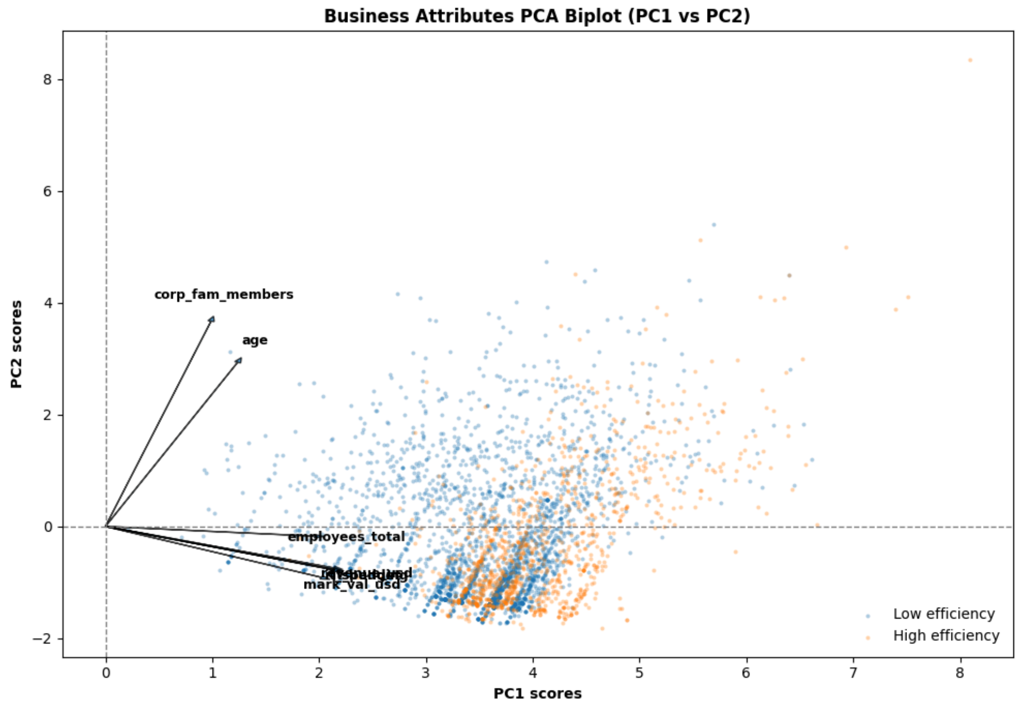

Figure 3.15: Business Attributes PCA Biplot (PC1 and PC2)

The PCA biplot suggests that firms with higher revenue-to-IT-spending ratios tend to project more strongly along PC1. Since PC1 loads heavily on variables associated with firm scale and economic size (such as revenue and employment), this suggests that more IT-efficient firms are generally positioned toward the larger-scale end of the firm spectrum. This pattern is consistent with the notion that firms which convert IT expenditures into revenue more effectively tend to achieve greater economic scale.

A natural extension of this analysis is to investigate which firm-level characteristics are associated with higher or lower IT efficiency. Identifying the features that systematically contribute to effective conversion of IT expenditure into economic output would provide actionable insights for firms seeking to scale sustainably. Such analysis could inform strategic decisions on organisational structure, investment priorities, and digital capability development, enabling firms to better align IT spending with growth objectives.

3.4 Classification Models

We initially implemented two classification models, **XGBoost** and **Random Forest**, to predict acquisition relationships between companies using both firm attributes and network-derived features.

Both models achieved a **perfect ROC-AUC and accuracy of 1**, indicating that the classes are almost perfectly separable. While this may suggest strong predictive performance, it more likely signals **data leakage** rather than genuine forecasting ability.

Many of the most important features (e.g. Modularity Class, Component Number, Strong Component Number, and Eccentricity) are graph-structural measures computed directly from the acquisition network itself.

Since these metrics encode community membership and connectivity patterns that already reflect the observed edges, the models effectively learn the network structure rather than the underlying economic or strategic drivers of acquisitions.

Feature importance rankings from both models show strong agreement, consistently prioritising differences or ratios of these structural variables over financial characteristics (e.g. revenue, number of employees). This suggests that companies within the same community cluster or supply-chain ecosystem are mathematically inseparable from their acquisition partners, turning the task into descriptive mapping rather than prediction.

Hence, we require alternative models. We consider reinforcement learning and deep learning models.

3.5 Reinforcement Learning

In assessing alternatives to classification models, we trialled a proximal policy optimisation (PPO) reinforcement learning model. To compensate for the imbalanced data, a significantly higher reward score of 10 was awarded to correctly identified acquisitions as compared to the score of -1 for wrongly identified ones. However, as the PPO model could not process the data with respect to the network topology of the entities, this resulted in an over-reporting of false negatives as the model was trained to recognise the entities without any respect to the networks of sub-conglomerates that they were members of.

3.6 Graph Neural Network (GNN)

To more comprehensively account for the topology of the corporate network, we used a GNN deep learning model. This involved us modelling acquisition prediction as a graph link prediction task where companies are nodes and known acquisitions are edges. Node features include financial metrics, structural network properties and categorical industry indicators. A GNN learns node embeddings by propagating information across the company network. Acquisition likelihood is computed via embedding similarity and trained using positive and negative edge sampling. Performance is evaluated using a deterministic train/validation/test split for reproducibility.

The model achieved the following:

- Validation AUC: 0.958
- Test AUC: 0.969
- Validation Average Precision (AP): 0.940
- Test AP: 0.956

These scores indicate excellent predictive performance, with the model correctly ranking true acquisitions above non-acquisitions approximately 96-97% of the time and maintaining high precision. Similar validation and test results demonstrate strong generalisation and minimal overfitting.

3.7 Textual analysis on sic_desc

In the original dataset, the SIC description was quite specific. Using textual analysis through Huggingface LLM, we categorised the descriptions into 8 categories (may vary slightly every time you run the file), allowing us to group the companies according to their industry more easily.

These industries are as follows:

1) Manufacturing
2) Industrial and Trade Services
3) Utilities and Energy
4) Agriculture and Livestock
5) Education, Research and Social Services
6) Retail and Wholesale Trade
7) Information and Communication
8) Transportation and Logistics

Companies_working.xlsx has all the generic labels.

4 Evaluation and Insights

4.1 Evaluation of Model Performance

We evaluated the performance of each model to determine the most reliable framework for predicting acquisition relationships. The results highlight the importance of modelling relationships between companies.

The **tree-based classification models** (XGBoost and Random Forest) achieved perfect ROC-AUC and accuracy scores of 1.0. However, this performance is misleading due to data leakage. Many important predictors (e.g. modularity class, component identifiers) were computed directly from the observed acquisition network.

Hence, we attempted to address this limitation by implementing **reinforcement learning** (PPO). However, the model performed poorly with a high percentage of false negatives. Since the model treated companies as independent entities and did not incorporate relational or topological information, it did not capture the structural dependencies between companies.

The **Graph Neural Network (GNN)** aimed to incorporate both firm attributes and network structure. This approach produced strong results with a test AUC and AP of 0.969 and 0.956, respectively. This indicates a high discriminative power and good generalisation to unseen edges. Overall, the GNN has the most robust performance, suggesting that acquisition behaviour is best modelled using a network graph, since it improves predictive reliability.

4.2 Key Insights

These are some key insights we have drawn from our analysis and evaluation:

1. **Network topology is important in determining acquisition behaviour.** Approaches based solely on firm-level attributes do not capture the inherently relational nature of acquisitions. The strong performance of the GNN demonstrates that a company's position within the ownership network is a better indicator of acquisition likelihood than financial metrics.

2. **A company's high operational efficiency can make it attractive for acquisition.** Our analysis of revenue per employee (in **Section 3.1.7**) shows a distinct group of companies that generate a disproportionately high revenue with a relatively small workforce. These companies are operationally efficient and scalable, suggesting they can deliver strong returns when integrated into a larger corporate structure. Hence, these companies are more likely to be acquired by larger companies.

3. **Mature companies tend to have different dynamics.** Our analysis of market value (in **Section 3.1.5**) shows that mature companies have lower median market values, but substantially higher variance. This could suggest that many mature companies tend to stagnate, but companies that can continue to innovate and stay relevant will continue

growing their market value in the long term. Targeting mature companies may involve a higher risk but also offers the potential for larger strategic gains.

5 Conclusion

Our project has developed a GNN-based deep learning prototype to support acquisition decisions within companies in the Champions Group dataset. Through EDA and network modelling, we found that acquisition activity is influenced not just by company-level characteristics but also by the relationships between the companies.

Overall, our findings show that an effective reorganisation strategy requires an awareness of the relationships between companies. If we prioritise structurally central and operationally efficient companies, and mature companies with high potential, there will be many opportunities to identify acquisition opportunities, reduce organisational complexity and maximise long-term value creation in the corporate ecosystem.

Annex A: Columns Dropped from our Analysis

From **Figure 3.1**, and based on our analysis, we drop the following columns:

Column to drop	**Rationale**
Website, Phone Number and Address Line 1	Not relevant to our analysis
8-Digit SIC Code and 8-Digit SIC Description, NAICS Code and NAICS Description, NACE Rev 2 Code and NACE Rev 2 Description	Not relevant to our analysis
Ticker	High percentage of missing values
Parent Street Address	Not relevant to our analysis
Registration Number Type	High percentage of missing values
Global Ultimate Street Address and Global Ultimate Postal Code	Not relevant to our analysis
Domestic Ultimate Street Address and Domestic Ultimate Postal Code	Not relevant to our analysis
Registration Number and Registration Number Type	Not relevant to our analysis
Company Description	Not relevant to our analysis
Fiscal Year End	All values are the same at 2031
ANZSIC Code and ANZSIC Description	Not relevant to our analysis
ISIC Rev 4 Code and ISIC Rev 4 Description	Not relevant to our analysis
No. of PC, No. of Desktops, No. of Laptops, No. of Routers, No. of Servers and No. of Storage Devices	Information is captured by the IT Spend variable
IT Budget	IT Budget and IT spend are perfectly correlated

Annex B: Definitions of Network Graph Variables

The following are quantities of interest which we have defined:

Quantity	**Definition**
Indegree	**The number of direct owners / controlling entities** this company has. A high indegree indicates that a company is jointly owned or influenced by multiple parents.
Outdegree	**The number of subsidiaries or entities this company directly controls**. A high outdegree indicates that a company is a holding company, conglomerate, or a group-level entity.
Degree	**The total number of connections (Indegree + Outdegree).** It measures the overall **connectedness** of a company in the ownership network. A high degree indicates that a company is structurally important.
Eccentricity	The **maximum shortest-path distance** from this company to any other reachable company. It measures how far the company is from the most distant part of the network. A low eccentricity indicates that a company is centrally located in the corporate structure, while a high eccentricity indicates that a company is a peripheral or deeply nested subsidiary.
Closeness centrality	The **inverse of the average shortest-path distance** to all other reachable companies. It measures how **quickly influence or information** could spread from this company. A high closeness indicates that a company is strategically central in ownership chains.
Harmonic closeness centrality	A variant of closeness that handles **disconnected networks** better. It is more appropriate for corporate networks with isolated groups. A high value of harmonic closeness indicates that a company is close to many others, even across components.
Betweenness centrality	The fraction of the shortest paths between other companies that pass through this company. It measures **brokerage or gatekeeping power**. A high betweenness value indicates that a company is an intermediary holding company, tax vehicle or regional HQ. These firms often sit between global ultimate owners and operating subsidiaries.
Authority	(from HITS algorithm) It measures how much a company is **owned by influential parents**.

	A high authority score indicates that the key operating company is controlled by major players.
Hub	(from HITS algorithm) It measures how much a company **owns important companies**. A high hub score indicates that the company is a strong parent, holding or investment company.
Pageranks	The probability that a random walk lands on the company. It measures **overall structural importance**, considering both quantity and quality of connections. A high PageRank indicates that the company is an influential firm in ownership/control flow. It often identifies global ultimate owners or central intermediaries.
Componentnumber	The ID of the **connected component** the company belongs to. Each component represents a **separate corporate group** or ownership cluster.
Strongcompnum	The ID of the **strongly connected component** (for directed networks). It indicates groups of companies with **mutual ownership or circular control**. It is useful for detecting cross-shareholding structures.
Modularity class	Community detected via modularity optimisation. It represents **corporate subgroups**, often aligned with business units, regions, legal entities within a multinational group, etc.
Stat_inf_class	Statistical inference-based community assignment (e.g. stochastic block model). It groups companies with **similar structural roles**, not just dense connections (e.g. all regional holding companies across different countries).
Clustering	Measures how interconnected a company's neighbours are. It indicates **triangular ownership structures** or cross-holdings among affiliates. A high clustering indicates complex corporate structuring, while a low clustering indicates clean hierarchical ownership.
Eigencentrality	Measures importance based on connections to other important companies. It captures **embedded influence** in elite corporate structures. It can be used to uncover hidden control (e.g. revealing that certain shareholders exercise disproportionate influence beyond their direct equity shares). A high eigenvector centrality indicates that a company is connected to powerful parents or subsidiaries.