



# **GNN-Based Deep Learning for Optimised Corporate Reorganisation**

Team RedBullz Airways

# Table of contents



**01**

## **Problem Definition**

Business Objective, Operational Constraints and Aims

**02**

## **Key Findings**

Network Analysis, EDA, PCA and Textual Analysis

**03**


## **Evaluation**

Comparison of Possible Models and Model Performance

**04**

## **Key Insights**

What does this mean for companies looking to acquire?





01

# Problem Definition

Business Objective, Operational Constraints and Aims



# Business Objective and Constraints

**Recommend** potential mergers and acquisitions (M&A) from the Champions Group dataset based on compatibility of potential parent and child entities

## Constraints:



Existing ownerships and conglomerate **structures**



Current financial, technological and manpower **capabilities** of entities

# Overarching aims of this project



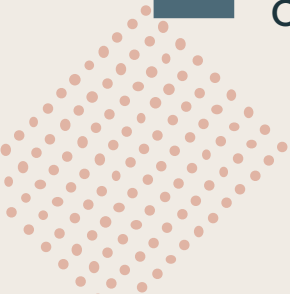
Discover interesting trends in the data such as how various features are related so that a viewer gain a **broad understanding of the dataset.**



Develop a machine learning model to **prescribe the most compatible M&A.**



**Identify outliers** using EDA and network analysis (e.g circular ownership).





02

# Key Findings

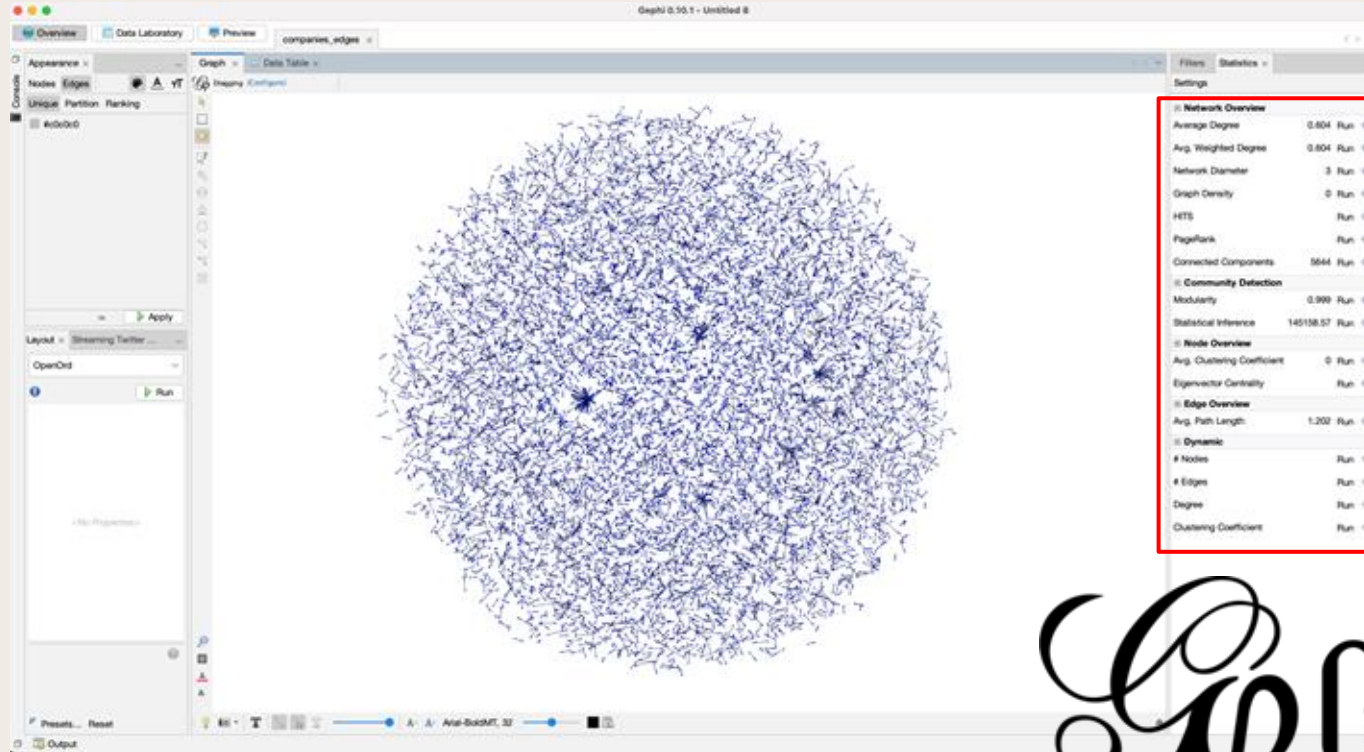
Network Analysis, EDA, PCA and Textual Analysis



# Network Analysis

Parent

Subsidiary



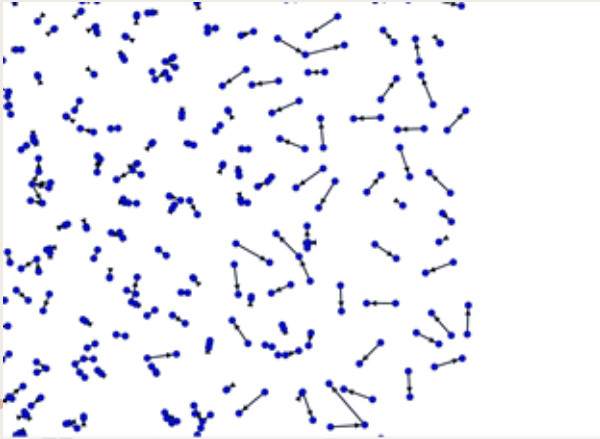
Network parameters

Gephi

# Much potential for future M&A

Largely **disconnected** graph

Majority of sub-conglomerates exist as **standalone parent-child pairs** (degree = 1)



Network graph created using Gephi.

Network Overview			
Average Degree	0.604	Run	?
Avg. Weighted Degree	0.604	Run	?
Network Diameter	3	Run	?
Graph Density	0	Run	?
HITS	Run		?
PageRank	Run		?
Connected Components	5644	Run	?
Community Detection			
Modularity	0.999	Run	?
Edge Overview			
Avg. Path Length	1.202	Run	?

# Presence of entities with circular ownership

Significantly higher eigenvector centralities (prestige scores) for entities with **low out-degree**

Such entities belonging to the **same strongly connected component** (SCC)



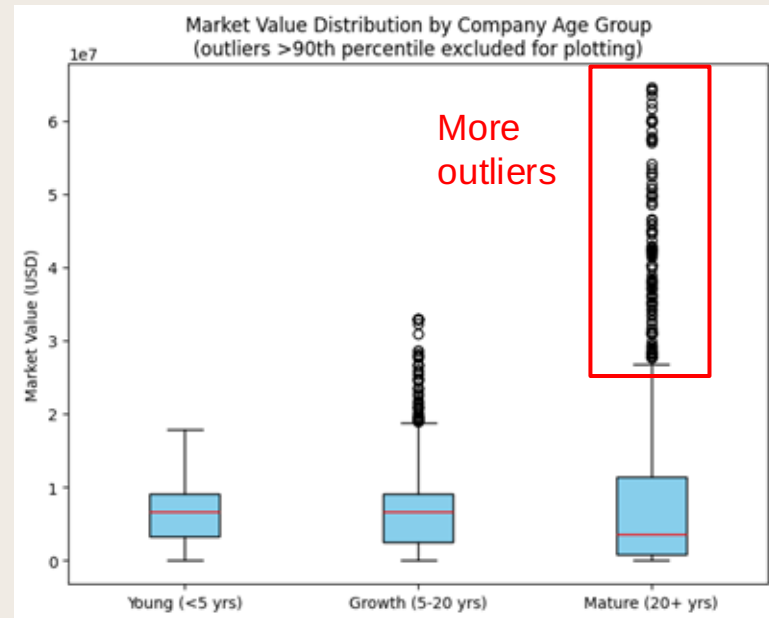
Network graph created using Gephi.

Id	Out-Degree	Strongly-Connected ID	Eigenvector Centrality ▾
ZUNG FU (CHINA) LIMITED	2	0	1.0
ZUNG FU MOTORS (CHINA) CO., LTD.	1	0	0.70713
CHENGDU ZUNGFU VEHICLE SALES AND ...	1	0	0.70713
SHENZHEN BAIROU INDUSTRIAL CO., LTD.	1	48	0.200049
PT. MEDISAFE TECHNOLOGIES	1	204	0.200049

# Early survival of a company does not always lead to long term growth

Mature companies generally have a **lower market value** than young and growth companies

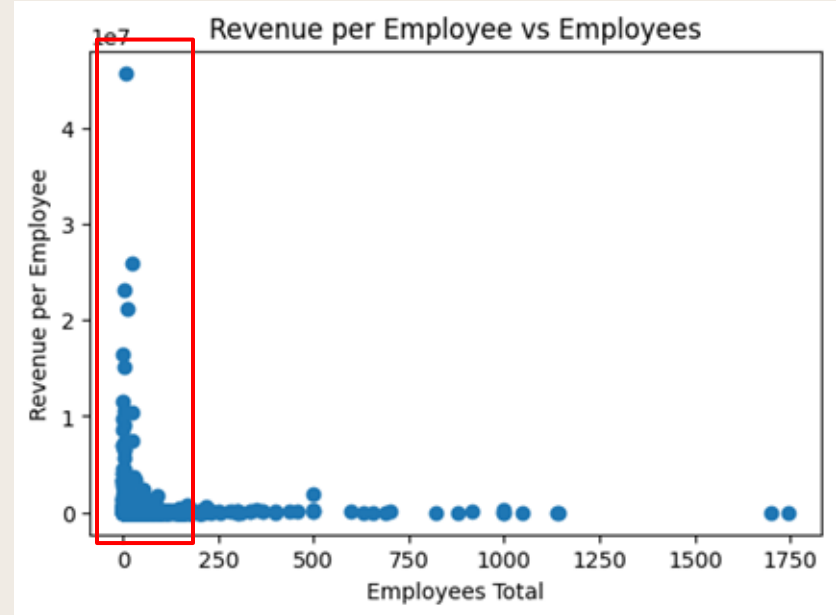
**“Winner-takes-most”** dynamic for the long term



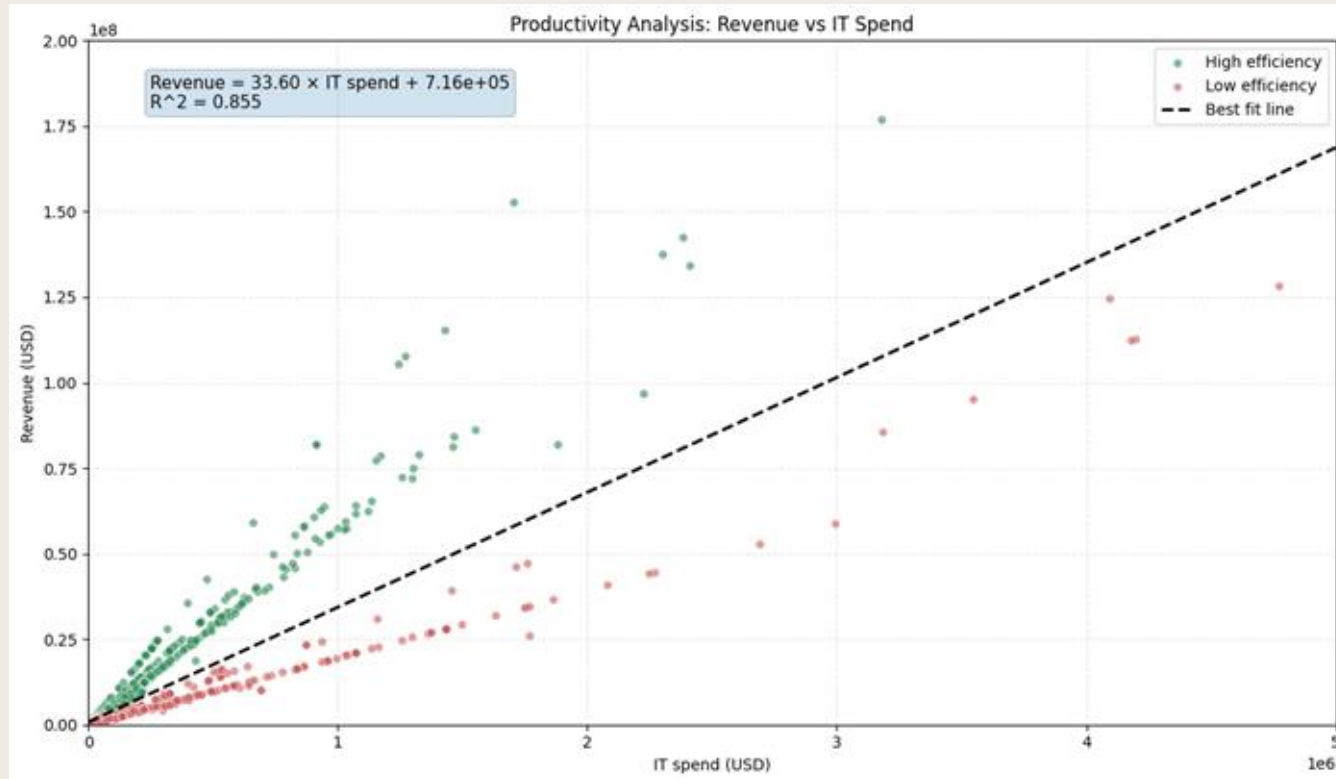
# Companies with high revenue per employee could be attractive acquisition targets

Some companies have very few employees but **high revenue per employee**.

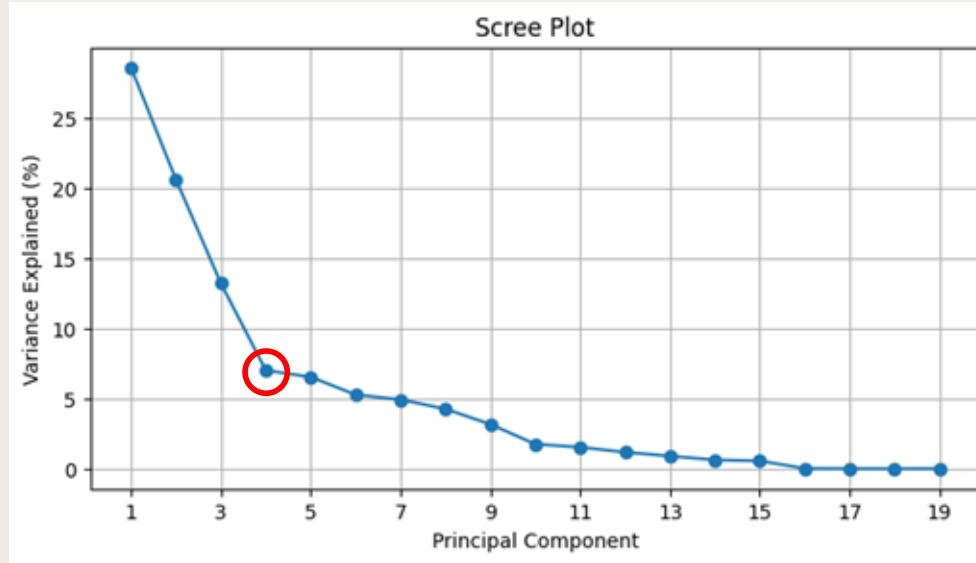
Organisational complexity and overhead in larger firms could **overshadow individual output**



# Firms fall into 2 groups based on the sensitivity of their revenue to IT spending

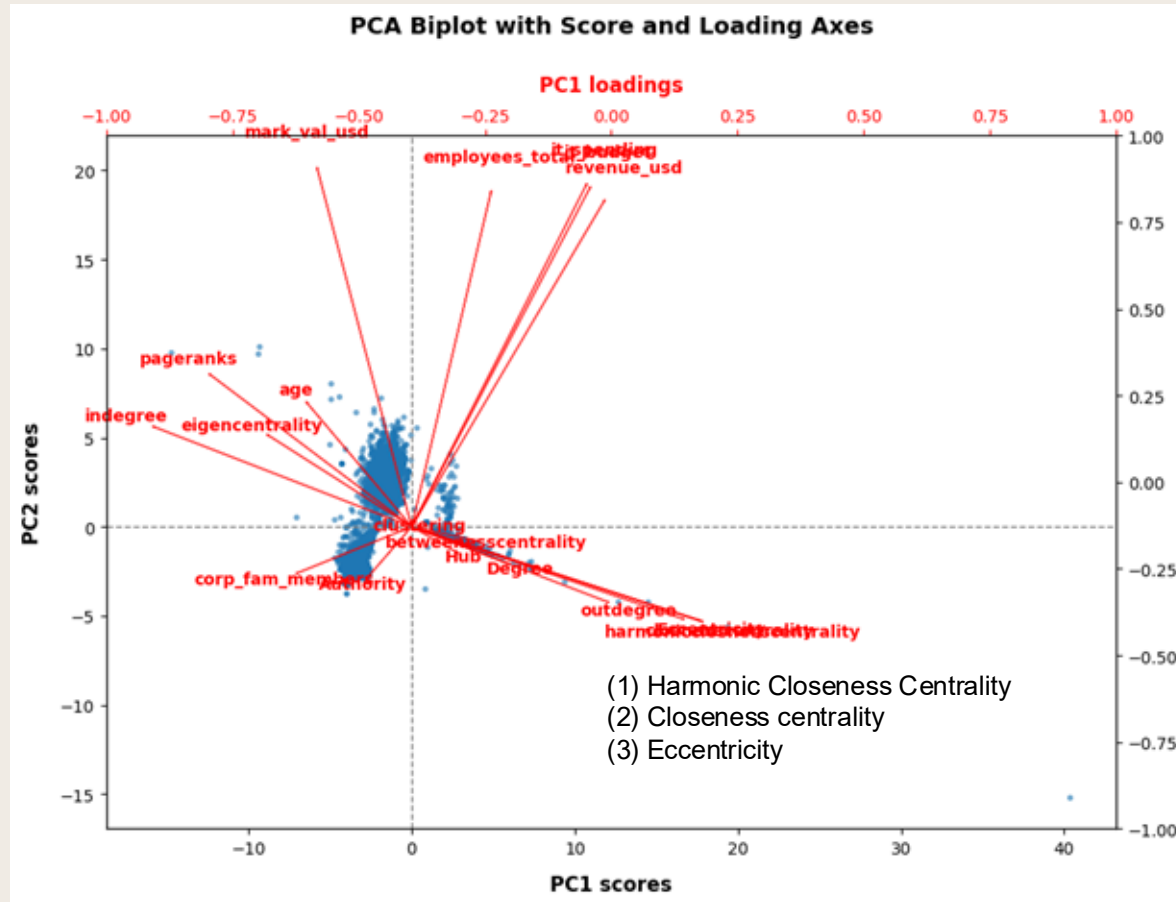


**There are 4 optimal principal components, accounting for ~69% of variance in the dataset**



PC1	PC2	PC3	PC4	Total
28.57%	20.59%	13.20%	7.00%	69.36%

# PC1 and PC2 account for ~50% of the variance



## PC1 and PC2 account for ~50% of the variance

PC1

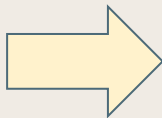
**How accessible is this company within our ecosystem?**

driven by Harmonic Closeness, Eccentricity and Outdegree

PC2

**What is the firm's resource capacity?**

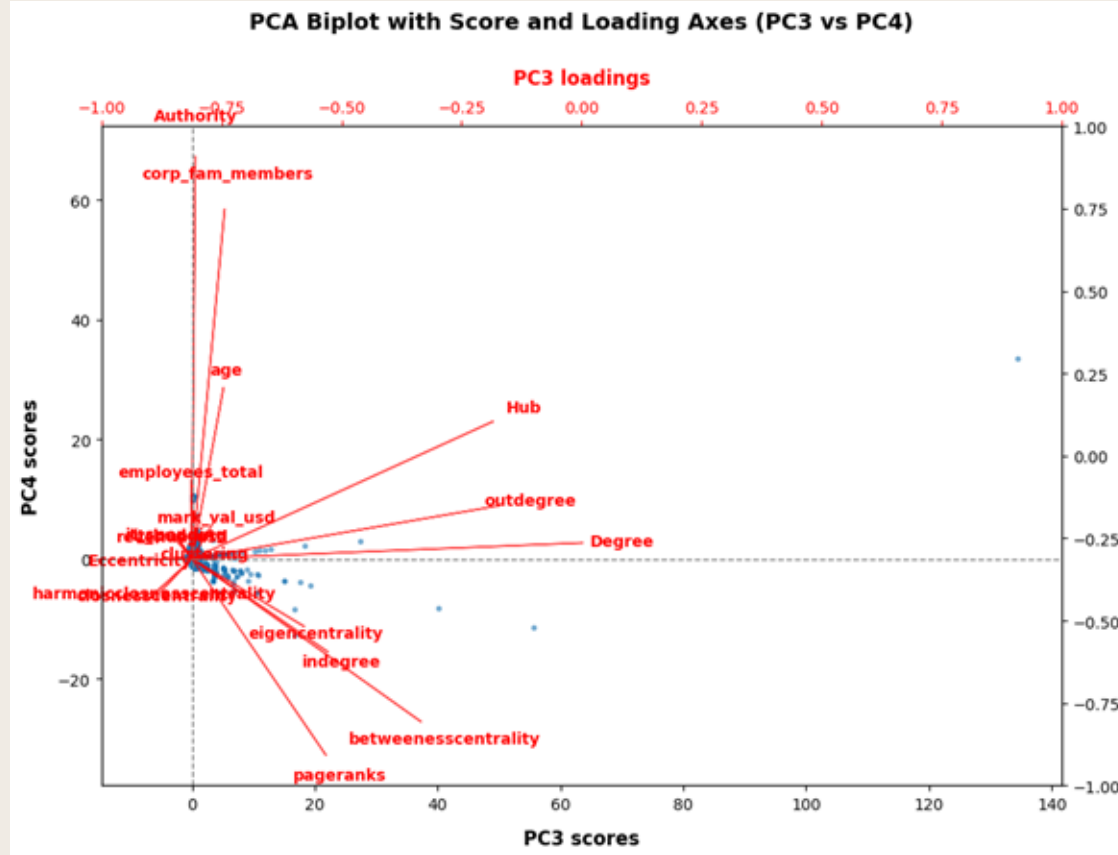
driven by Market Value, IT Spend, Revenue and Manpower



**Is the firm a “high efficiency” responder?**

Firms maximising **Revenue** relative to **IT spending** correlate strongly with these axes, and tend to achieve the greatest economic scale.

# PC3 and PC4 account for ~20% of the variance



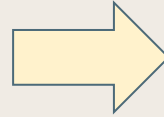
## **PC3 and PC4 account for ~20% of the variance**

**PC3**

**Is this firm highly  
connected with different  
parts of the network?**

**PC4**

**How influential is the firm  
relative to other firms?**



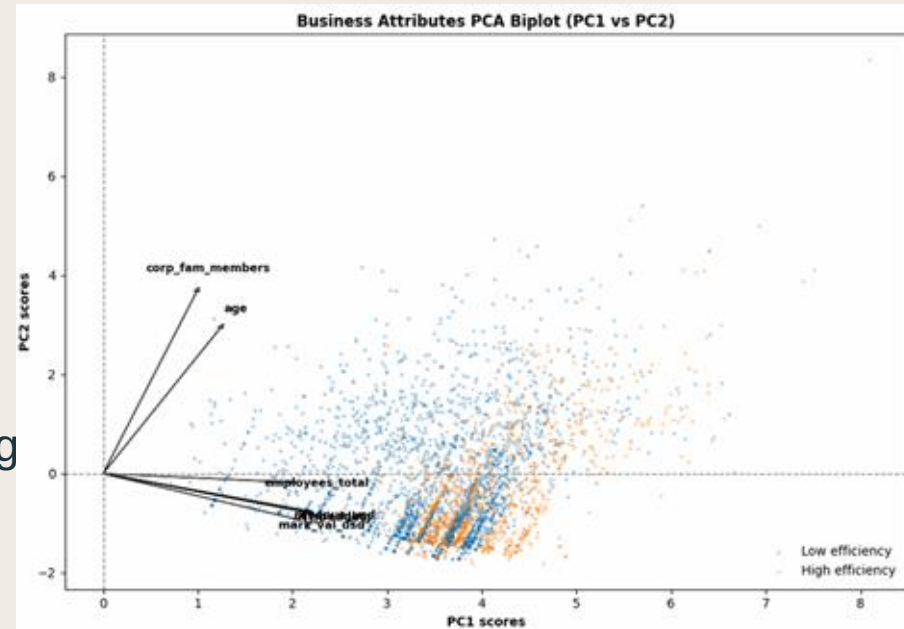
**Is the firm a “strategic  
anchor” target?**

# Revenue-IT efficiency correlates strongly with PC1 in the business attributes specific PCA

Business attribute PC1 seems to be a “firm scale/economic size” axis.

High efficiency firms generally do better on this axis, which correlate to market value as well.

**High efficiency**= high revenue/IT spending






03

# Evaluation

Comparison of Possible Models and Model Performance



# Textual analysis with Hugging Face LLM 🧠

**Broad industrial  
categorisation**

**Granular SIC (Standard Industrial Classification) →  
8 broad categories** (Manufacturing, Transportation  
and Logistics etc)

# Textual analysis with Hugging Face LLM 🧠

Broad industrial  
categorisation

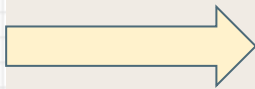


Numerical data  
imputation

**Granular SIC (Standard Industrial Classification) → 8 broad categories** (Manufacturing, Transportation and Logistics etc)

**Provides a more complete dataset for ML applications/inference.**

revenue_usd	sic_desc
517593000	Whol autos/motor vehicles
3330398	Automotive repair
0	Nonclassified establishment
715605	Real estate agent/manager
365224	Commercial physical research



revenue_usd	sic_desc
5067400	Utilities and Energy
441376	Industrial and Trade Services
8336460	Industrial and Trade Services
46441	Industrial and Trade Services
2710250	Information and Communicati
1495050	Transportation and Logistics
381784	Industrial and Trade Services
7212268	Utilities and Energy
40082400	Industrial and Trade Services

# Textual analysis with Hugging Face LLM 🧠

Broad industrial  
categorisation



Numerical data  
imputation



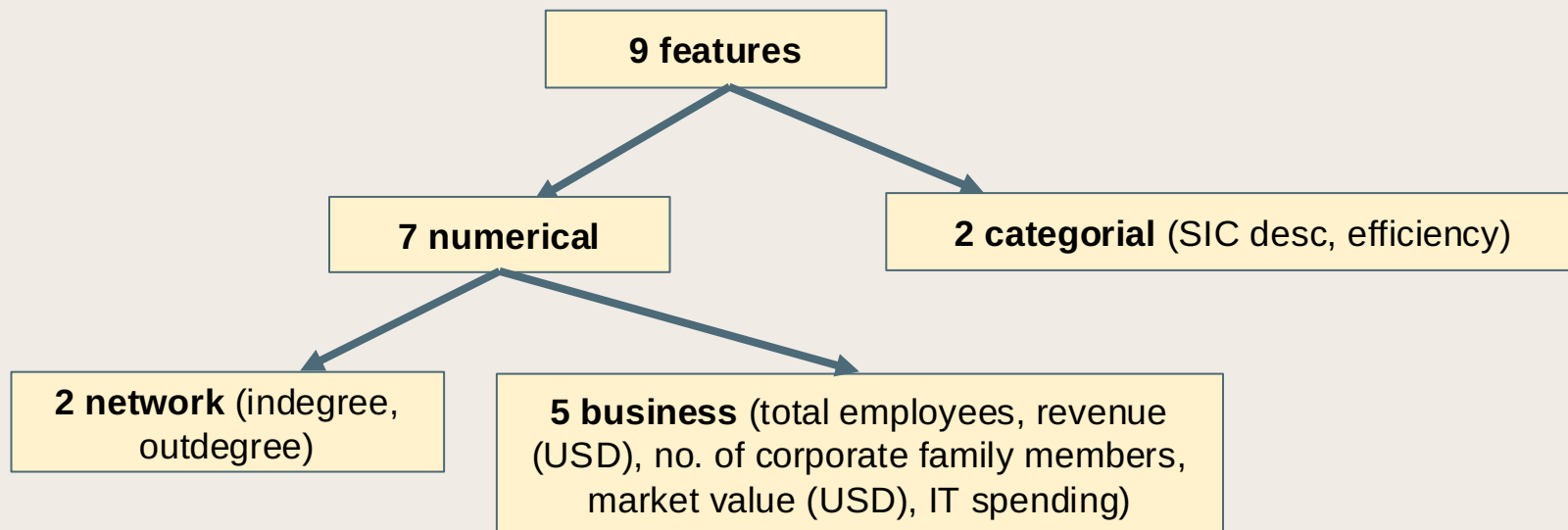
Creation of clean  
dataset for further  
analysis

**Granular SIC (Standard Industrial Classification)** →  
**8 broad categories** (Manufacturing, Transportation  
and Logistics etc)

**Provides a more complete dataset for ML  
applications/inference.**

New dataset has **broad industrial categorisation**  
(more companies per category) and **more  
complete numerical data.**

# The features we used to prescribe the acquisition relationship



# Graph Neural Network is most appropriate to predict acquisition relationships between companies

XGBoost +  
Random Forest

**Perfect ROC-AUC and accuracy of 1**  
Overfitted due to data leakage



Reinforcement  
Learning

**Over-reporting of false negatives**  
Ignored relational context

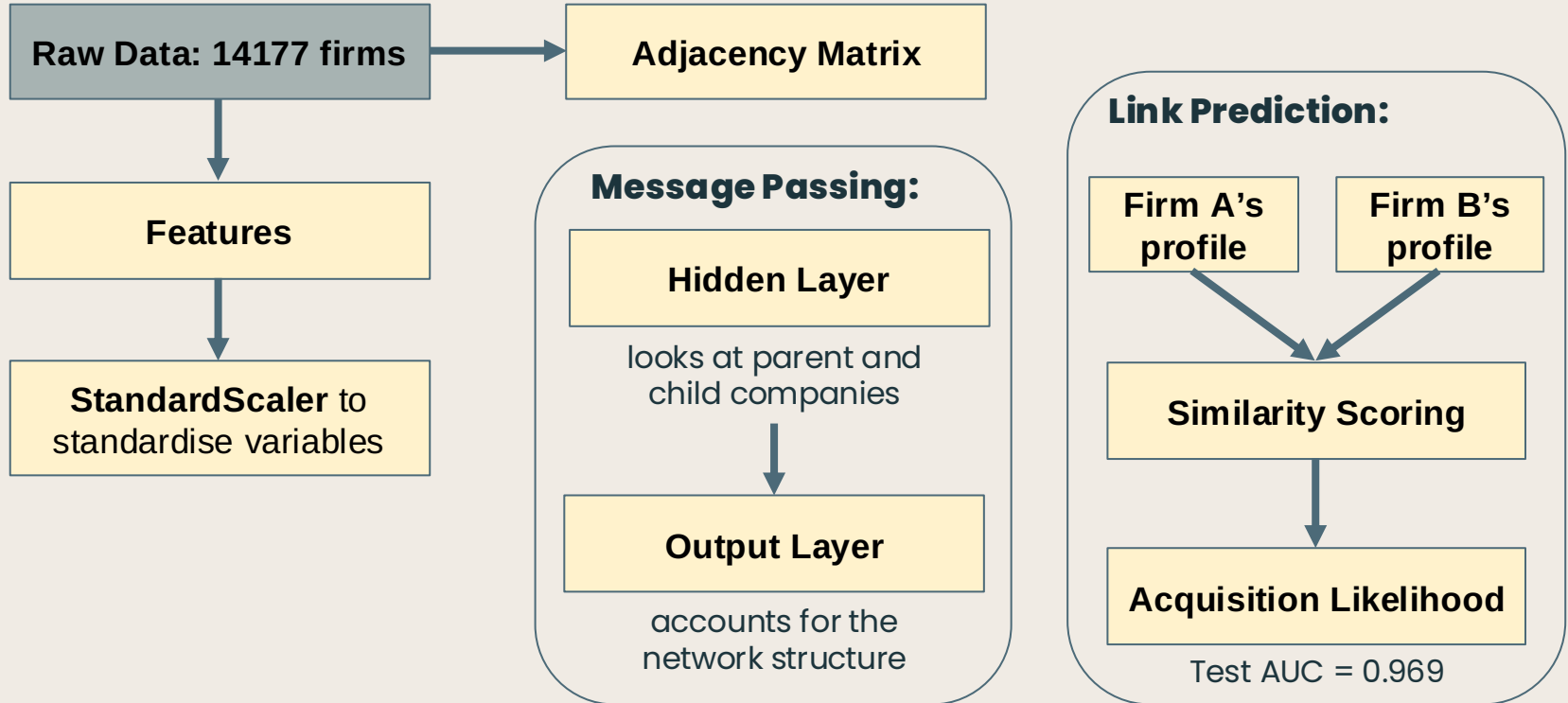



Graph Neural  
Network

**Strong generalisation and minimal overfitting**  
Test AUC = 96.9%; Test Average Precision = 95.6%



# How the GNN model works







04

# Key Insights

What does this mean for companies looking to acquire?



# **Relationships between companies drive acquisition behaviour, not just financial metrics**



**20% of variance is explained by structural connectivity and authority in PC3**




**GNN vastly outperformed attribute-only models**




**Companies should use network-based scoring to prioritise their acquisition targets**

# Operationally productive firms are high-value acquisition targets

Higher revenue per employee = higher acquisition attractiveness



Who should we acquire?



We should target the most productive firms first!

Productivity can bring about economies of scale!

Boss

Me

# Mature firms are not stagnant; they are high-variance strategic anchors



My boss

Most mature firms look stagnant. Should we avoid them?

We should diversify our investment in mature firms!



Me

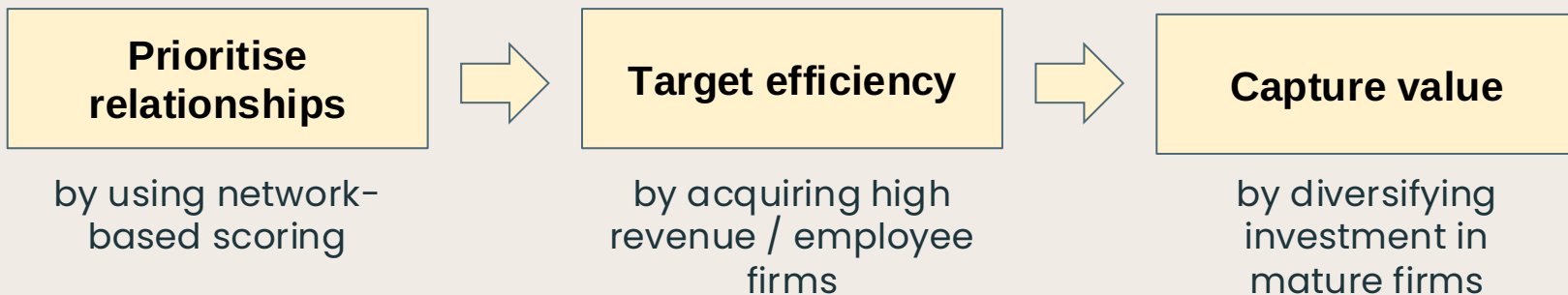
A few firms dominate total value. We should target them!



My workpal

# RedBullz Airways gives your strategy Wings

GNN moves corporate strategy from intuition-based guesses to data-driven precision.





# Thank you

**CREDITS:** This presentation template was created by **Slidesgo**, and includes icons, infographics & images by **Freepik**

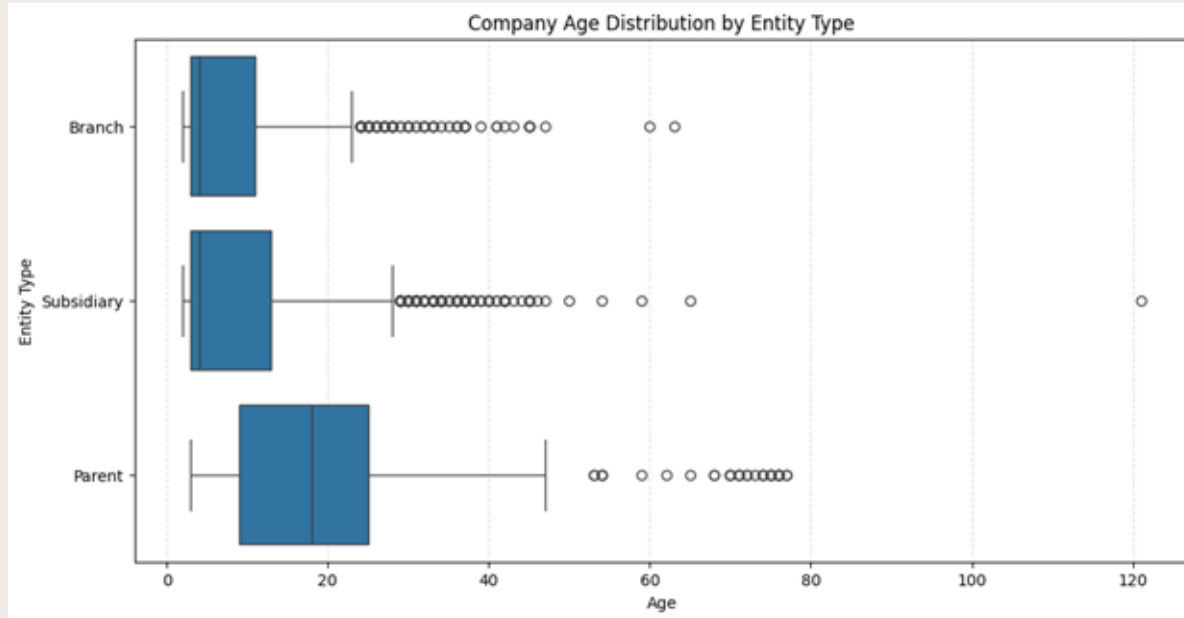


**05**

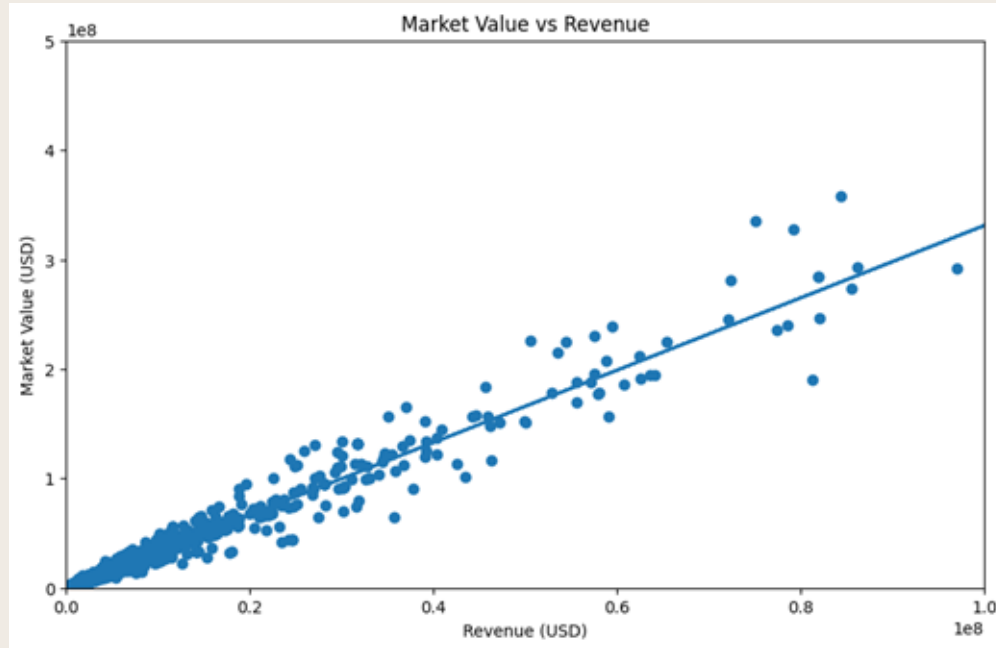
# **Annex**



# Parent companies are generally much older than Branch and Subsidiary companies



**There is a strong positive relationship between market value and revenue**



# PC1 Analysis

- PC1 is primarily driven by distance-related network measures
- PC1 can be interpreted as separating companies that appear more “reachable” in the network

PC1 Variable(s)	Loading value
Harmonic Closeness Centrality	0.386
Closeness Centrality	0.382
Eccentricity	0.361

# PC2 Analysis

- PC2 can be interpreted as a clear company scale or financial capacity dimension.

PC2 Variable(s)	Loading value
IT Spending	0.416
IT Budget	0.413
Employees_total	0.397

# PC1 and PC2 Analysis

- PC1-PC2 plane is driven primarily by firm size and spending capacity variables, and secondarily by a set of network position variables.

Variables	Loading value (PC1 + PC2)
IT Spending	0.477
IT budget	0.476
Revenue (USD)	0.472

# PC3 Analysis

- 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 and often sit on paths between others (betweenness)

PC3 Variable(s)	Loading value
Degree	0.584
Outdegree	0.459
Hub	0.450

# PC4 Analysis

- PC4 can be interpreted as separating authority, corporate-family and age characteristics from PageRank, betweenness-type and network influence characteristics.

PC4 Variable(s)	Loading value
Authority	0.584
Corp_fam_members	0.536
Hub	0.211

# PC3 and PC4 Analysis

The PC3 and PC4 components are driven primarily by network structure metrics, rather than firm size or spending variables.

Variables	Loading value (PC1 + PC2)
Authority	0.617
Corp_fam_members	0.536
Age	0.263

# Risks of acquiring high-efficiency firms

Killing the firm's high efficiency by overwhelming it with parent company's administrative overhead.

Suggestions:

- Allow high-efficiency targets to maintain their high efficiency structures and specialised teams.
- Inject parents' IT budget and spending capacity to scale efficient processes.