Project BDO Austria GmbH

Network Analysis for Detecting Shared Risk Factors Between Companies

Supervisor: PD Dr. Ronald Hochreiter

Simon Böck – 12115066
Manuel Petschinger – 12026023

Jakob T. Koller – 11741732
Markus C. Weiss – 12030190

Contents

List of Figures

List of Tables

Abstract

The project conducts a detailed analysis of a complex network, focusing on four main areas: natural language processing (NLP), building the network, analysing it and predicting new connections. An important part of the work is the detailed categorisation of company departments and the accurate identification of individuals in the data. This helps to better understand the organisations' structure and operations. A key part of the project is monitoring how the network changes over time to identify the most influential members and their roles.

The project uses visual tools to point out important entities and to study the network's complex connections. These visualisations are useful for spotting specific connections and for a clearer view of how different nodes in the network interact. The project also examines the financial performance of the companies, looking at indicators like "Return on Assets", "Gross Margin" and "Market Capitalisation". This analysis provides insight into how companies within the network are related financially. The final part of the document provides an in-depth look at smaller, label-specific networks within the larger network. This is essential for understanding the network data more clearly, allowing a detailed study of interactions and connections based on specific categories such as "person", "organisation", "country" and "city".

1 Introduction to BDO-Project

This project, conducted as part of the Data Science Lab course at WU (Vienna University of Economics and Business), is designed to provide hands-on experience in applying data science techniques to real-world business scenarios. Partnering with BDO, a global leader in audit, tax and advisory services, the project aims to leverage academic knowledge in a practical setting, focusing on network analysis, corporate finance and risk management. This collaboration not only enhances our learning experience, but also contributes valuable insights to BDO, demonstrating the practical applications of data science in the business world.

1.1 Overview

The Company: BDO is a global network of public accounting firms, the fifth largest in the world. Operating in 167 countries, with over 88,000 employees, BDO provides audit, tax and advisory services to a wide range of clients. Founded in 1963, the firm's name originally stood for Binder Dijker Otte & Co. Over the years, BDO has expanded through a series of significant mergers and acquisitions, enhancing its capabilities in various sectors [1].

The Related Field: Data Science, Corporate Finance and Risk Management:
The project associated with BDO delves into the realms of data science and analytics, corporate finance and risk management. These fields are increasingly intertwined in the modern business landscape [2].

Data Science & Analytics: This field involves using statistical techniques, machine learning and big data analytics to extract insights from data. In the context of corporate finance and risk management, data science helps in predicting market trends, assessing risks and making informed financial decisions [3].

Corporate Finance: This area focusses on the financial activities of corporations, including capital structure, funding strategies and investment decisions. It plays a crucial role in the maximisation of shareholder value through long- and short-term financial planning and the implementation of various strategies [4].

Risk Management: Involves identifying, assessing and prioritising risks followed by coordinated application of resources to minimise, control, or mitigate the impact of unfortunate events. In the financial sector, this includes market risk, credit risk and operational risk [5].

1.2 Project Goals

For network visualisation and understanding, the goal is to create a detailed depiction of the interconnections between companies and other entities, ensuring that the key players, clusters and potential future links are prominently featured. More specifically, the focus lies on the exploratory analysis of networks, identification of patterns and connections between the nodes. In identifying risk factors, the focus is on analysing the network to pinpoint shared risks among companies, using attributes between companies to uncover possible threats and weaknesses. Here, we will try to additionally use the financial data to uncover any insight. Predictive analysis aims to utilise the trends visible in existing data to forecast future potential relationships between companies and anticipate the emergence of new risks.

Success criteria are defined by achieving a lucid network visualisation that emphasises the main actors and links, accurately identifying and classifying risk factors, providing precise forecasts for future connections and risks and receiving affirmative feedback from BDO Austria GmbH, the data coaches and PD Dr. Ronald Hochreiter regarding the pertinence and precision of the project's results. The ultimate goal is to deliver a project that offers significant insight into the network of companies and their collective risk factors. The final presentation and the report will encapsulate all findings, analyses and visualisations into a comprehensive report, which will then be presented to BDO Austria GmbH and other interested parties.

2 Literature Review

Introduction: This literature review critically examines recent progress in network analysis, with a focus on its application in corporate finance and risk management. It synthesises findings from key studies to provide a nuanced understanding of current methodologies and challenges in this area. The reviewed studies include "An Empirical Analysis of Corporate Financial Management Risk", "Business Analytics for Corporate Risk Management and Performance" and "Systemic Risk Management and Investment Analysis with Financial Network Analytics", as well as a review of network models in financial systemic risk.

2.1 Current Research

Associative Memory Neural Networks in Financial Management: The study "An Empirical Analysis of Corporate Financial Management Risk" introduces associative memory neural networks for managing financial risks. These networks, modelled after brain functions, offer a comprehensive framework for tackling financial management issues, particularly in risk synchronisation and stability analysis. Research emphasises improving enterprise risk management systems and financial risk awareness, presenting a stochastic amnestic neural network model to predict and manage financial risks in corporations [6].

Utilising Accounting Narratives for Corporate Risk Assessment: "Business Analytics for Corporate Risk Management and Performance" highlights the importance of accounting narratives in risk assessment. It uses text mining to extract risk information from narratives, combining this with readability metrics to gauge corporate risk attitudes. Using a two-stage network data envelope analysis and fuzzy rough set theory, the study offers a comprehensive view of a firm's risk profile by incorporating financial and non-financial data [7].

Financial Network Analytics: "Systemic Risk Management and Investment Analysis with Financial Network Analytics" discusses the interconnectedness of the global financial system and the need for a network-based approach to manage systemic risks. The paper underscores the role of big data in understanding global financial dynamics and developing intelligent business applications, advocating financial network analytics to aid in the formulation of financial policies [8].

Network Models of Financial Systemic Risk: The study "Network Models of Financial Systemic Risk" reviews the use of network approaches in modelling financial systemic risk. It examines the empirical structure of interbank networks, the mechanics of algorithms like Eisenberg–Noe and models such as DebtRank. The study emphasises the understanding of hidden interbank links through cross-asset holdings, crucial for modelling systemic risk [9].

Conclusion: The literature review reveals significant advances in the application of network analysis to corporate finance and risk management. The integration of associative memory neural networks, accounting narratives and financial network analytics reflects a diverse approach to understanding financial systems. These developments contribute both practical tools for risk management and theoretical insights into financial network analysis.

2.2 Project Relevance

Relevance to the Project: This literature review's insights into network analysis in corporate finance and risk management are highly relevant to our project. The reviewed studies provide a multifaceted understanding of network dynamics and analytical methodologies, crucial for our project objectives.

Associative Memory Neural Networks: Aligning with our data-driven focus, the associative memory neural networks discussed in "An Empirical Analysis of Corporate Financial Management Risk" can enhance our project's predictive accuracy in financial risk management. These networks could analyse complex financial data innovatively, offering insights into market trends and stability [6].

Accounting Narratives for Risk Assessment: Adopting methodologies from "Business Analytics for Corporate Risk Management and Performance", such as text mining for risk information in accounting narratives, can enrich our understanding of corporate risk profiles for a more comprehensive risk assessment [7].

Financial Network Analytics: "Systemic Risk Management and Investment Analysis with Financial Network Analytics" provides relevant insights for understanding financial networks' interconnections. Applying network analytics will enable us to identify critical nodes and connections, aiding in systemic risk identification [9].

Network Models for Systemic Risk Analysis: Our project can benefit from the frameworks in "Network Models of Financial Systemic Risk", particularly in understanding financial distress propagation and systemic risk modelling [8].

Conclusion: Overall, the reviewed literature significantly informs and enhances our project's approach to network analysis in corporate finance and risk management. By integrating these advanced methodologies, we can achieve a deeper understanding of financial systems, improve predictive accuracy and make informed decisions to manage financial risks effectively.

3 Methodology

3.1 SpaCy

In this project, we encountered a significant challenge with the data: the attributes associated with each company were unlabelled, providing strings of text without clear indications of their categories (e.g., country, city, company, person). To address this, we integrated *Spacy*, a cutting-edge natural language processing (NLP) library, into our methodology. Specifically, we utilised Spacy's English Transformer-based (TRF) model to efficiently categorise and label these ambiguous attributes, enhancing the accuracy and interpretability of our network analysis. Spacy is renowned for its performance in various NLP tasks, offering pre-trained models that are both highly efficient and accurate. The library is particularly well-suited for tasks like tokenization, named entity recognition (NER), part-of-speech tagging and dependency parsing. For our purposes, the English TRF model provided by Spacy was pivotal.

The English TRF model in Spacy is a transformer-based model, which leverages the capabilities of transformer architectures to understand the context and relationships within text. Transformer models have revolutionised the field of NLP by enabling models to consider the entire context of a word by focusing on the relevant parts of the text. This is particularly beneficial for disambiguating words that might have multiple meanings in different contexts. In our project, we employed the English TRF model primarily for its excellence in Named Entity Recognition (NER). NER is a crucial feature of the model, allowing us to identify and categorise entities within text into a broad spectrum of predefined entity types. Below is a detailed description of each named entity label used in our analysis:

- **CARDINAL**: Numerals that do not fall under another type. These are numbers that do not have a numeric value (like a quantifier or a part of a measurement) but are rather part of a named entity, like "Three" in "The Three Musketeers".
- **DATE**: Absolute or relative dates or periods. This label identifies entities that represent specific dates, years, or ranges of time, such as "1999", "20th century", or "the 1980s".
- **EVENT**: Named events like battles, sports events, or hurricanes. Entities under this label represent significant occurrences that are commonly recognised by a proper name, for instance, "World War II", "Olympics 2020", or "Hurricane Katrina".
- **FAC** (Facility): Buildings, airports, highways, bridges, etc. This label is applied to entities that are man-made structures designed for a specific purpose, like "JFK Airport", "Golden Gate Bridge", or "Empire State Building".

- **GPE** (Geopolitical Entity): Countries, cities and states. The GPE label encompasses names of political or geographic areas, including nations ("Germany"), cities ("New York City"), or provinces/states ("Bavaria").
- **LANGUAGE**: Any named language. Entities that are recognised as languages are categorised under this label, such as "English", "Spanish", or "Mandarin".
- **LAW**: Named documents made into laws, or legal document titles. This label is for entities that are legal documents or laws, like "Constitution", "Civil Rights Act", or "GDPR".
- **LOC** (Location): Non-GPE locations, mountain ranges and bodies of water. Locations that are not categorised as geopolitical entities but are still recognised geographical places fall under this label, including "Himalayas", "Sahara Desert", or "Mississippi River".
- **MONEY**: Monetary values, including unit. This label is applied to entities representing a specific amount of money, such as "$100", "€50 billion", or "7 million pounds".
- **NORP** (Nationalities or Religious or Political Groups): This label is for entities that represent nationalities, religions, or political affiliations, like "Americans", "Hindu", or "Democrats".
- **ORDINAL**: "First", "second", etc. ORDINAL entities represent position in an ordered sequence, such as "first", "23rd", or "last".
- **ORG** (organisation): Companies, agencies, institutions, etc. Entities that are recognised as organisations, including businesses, government entities and other institutions, are categorised with this label, like "United Nations", "Google", or "Harvard University".
- **PERCENT**: Percentage (including "%"). This label identifies entities that represent percentages, such as "50%", "twenty percent", or "5.2%".
- **PERSON**: People, including fictional. The PERSON label is for names of individuals, whether real or fictional, like "Elon Musk", "Harry Potter", or "Cleopatra".
- **PRODUCT**: Objects, vehicles, foods, etc. (not services). This label applies to entities that are tangible products, like "iPhone", "Boeing 747", or "Coca-Cola".
- **QUANTITY**: Measurements, as of weight or distance. QUANTITY entities represent a measurement or an amount of something, such as "15 kilograms", "100 miles", or "six liters".
- **TIME**: Times smaller than a day. This label is for entities that represent times of the day or durations of time shorter than a day, like "6:30 am", "two hours", or "midnight".
- **WORK_OF_ART**: Titles of books, songs, etc. The WORK_OF_ART label is used for the names of creative works, including books, songs, movies and other artistic creations, such as "Mona Lisa", "Bohemian Rhapsody", or "Inception".

The process involved feeding the text data through the model, which then utilised its pre-trained understanding of the English language to identify and categorise entities. The model's ability to understand context was particularly useful in our dataset, which contained numerous instances where the same string could represent a company name or a city name, depending on the context.

The utilisation of Spacy's English TRF model significantly enhanced the efficiency and accuracy of our data categorisation process. By automating attribute labelling, we not only expedited the data preparation phase, but also ensured a good level of consistency and reliability in the categorisation of company-related attributes. The application of Spacy and its English TRF model in our methodology underscores our commitment to employing advanced, reliable tools in our data science endeavors. This approach not only improved the quality of our network analysis but also set a robust foundation for any future analytical tasks involving the Eurostoxx 600 companies datasets.

For further technical details and documentation on Spacy and the English TRF model, refer to the official Spacy documentation and user guides [10].

3.2 Cartopy

Visual representation of data was crucial for understanding the intricate relationships and geographical distributions. To this end, we incorporated *Cartopy*, a Python library designed for cartography, to plot network graphs on world and Europe maps. Cartopy is particularly adept at handling geographic data and creating maps, making it an invaluable tool for our project. Cartopy is built on top of Matplotlib and provides a powerful interface for map creation, allowing for the plotting of data on a 2D map. It is well suited for geospatial data visualisation due to its ability to handle projections, coastlinesand other geographical features with high precision and ease.

1. **Mapping and Projection**: Cartopy excels in handling different map projections. It enables the transformation of geographical coordinates into 2D map projections, ensuring that the spatial relationships between points are accurately represented. In our project, we used Cartopy to project our network graphs onto both a global scale and a more focused European scale, depending on the level of detail required for specific analyses.

2. **Geographical Features**: Cartopy provides extensive support for adding geographical features to maps, such as coastlines, rivers and political boundaries. This feature was particularly useful in our project to contextualize the network relationships of companies. By overlaying network graphs onto maps with these geographical features, we were able to visualize the connections between companies in relation to their physical locations.

3. **Integration with Matplotlib**: Cartopy's integration with Matplotlib, a widely-used Python plotting library, allowed us to leverage Matplotlib's extensive range of plotting tools and customization options. This made it possible to create highly detailed and customized visual representations of our data, enhancing the interpretability and aesthetic appeal of our network graphs.

4. **Customization and Extensibility**: Cartopy provides a wide range of options for customizing the appearance of maps. This includes the ability to change map colours, add custom labels and incorporate additional layers of data. In our project, this level of customization was crucial for creating clear, informative visualisations that effectively communicated the results of our network analysis.

By utilising Cartopy in our project, we were able to create detailed and informative visual representations of the network relationships between Eurostoxx 600 companies. The library's powerful mapping capabilities, combined with its integration with Matplotlib and extensive customization options, made it an ideal tool for our geographical data visualisation needs.

For further technical details and documentation on Cartopy, refer to the official Cartopy documentation and user guides [11].

3.3 Country Converter

Ensuring the accuracy and consistency of country-related data in our project was paramount. To achieve this, we utilised the *Country Converter* (coco), a comprehensive Python package designed to convert and match country names between different classification schemes and standards. This tool was particularly effective in standardising country names, improving the accuracy of our attribute labelling and ensuring consistency across our dataset.

The Country Converter facilitates the harmonisation of country names by providing a wide range of conversion options, including standard country names, ISO codes and various classification systems. For our project, we focused on converting all country names to coco's "name_short" format, which provides a standardised, concise representation of country names. This standardisation was crucial to eliminate discrepancies and ambiguities in country-related data.

The key features and applications of the Country Converter in our project included:

1. **standardisation of Country Names**: Coco's "name_short" format provided a consistent naming scheme for countries, which was essential for accurate data analysis and comparison. This standardisation ensured that all country-related attributes were uniformly labelled, preventing any inconsistencies that might arise from varied naming conventions or typos.

2. **Wide Range of Conversion Options**: The Country Converter supports a broad array of naming conventions and standards, making it a versatile tool for dealing with country names. Its ability to convert between different standards allowed us to easily align our data set with other data sources or requirements.

3. **Handling of Ambiguous Names**: Coco is equipped to handle ambiguous or unclear country names, including historical names or alternative spellings. This feature was particularly valuable to ensure the integrity of our data, as it minimised the risk of mislabeling or overlooking country-related attributes.

4. **Integration with Data Processing Pipelines**: The ease of integrating coco with our data processing pipelines streamlined the data cleaning and preparation process. By automating the conversion and standardisation of country names, we were able to improve the efficiency and accuracy of our data handling procedures.

The utilisation of the Country Converter in our project significantly improved the quality and consistency of our country-related data. By standardising country names to coco's "name_short" format, we ensured that our analysis was based on accurate and consistent information, thus enhancing the reliability and interpretability of our findings.

For further technical details and documentation of the Country Converter, refer to the official Country Converter documentation and user guides [12].

3.4 NetworkX

For our visualisations, we have chosen to utilise NetworkX, a prominent Python library, for its exceptional capabilities in handling and analysing complex networks. This choice is strategically aligned with our project objectives for several compelling reasons:

Firstly, NetworkX is renowned for its user-friendly nature, offering a straightforward and intuitive interface. This accessibility is particularly beneficial in simplifying the processes involved in network analysis, making it easier for our team to focus on the analytical aspects of our project without getting bogged down by complex programming hurdles.

Furthermore, NetworkX is comprehensive in its functionality. It is equipped with a vast array of built-in tools and algorithms that cater to a wide spectrum of network analysis requirements. Whether it is basic operations like adding or removing nodes and edges or more advanced analytical tasks such as calculating network centrality measures, detecting community structures or pathfinding. NetworkX provides an all-encompassing suite of tools that enhances the depth and breadth of our analysis.

Another significant advantage of NetworkX is its seamless integration with the broader Python ecosystem. It works harmoniously with other Python libraries, such as Matplotlib for data visualisation or Pandas for data manipulation [13].

Network Metrics

For our network analysis, we used various metrics to gain insight into the structure and dynamics of the network. These metrics include:

Number of Nodes: Represents the total number of distinct entities (such as companies, industries or countries) in a network.

Number of Edges: Indicates the total number of connections or relationships between the nodes in the network.

Average Degree: Reflects the average number of connections per node. A higher average degree typically indicates a more interconnected network.

Density: Measures how close the network is to being fully interconnected. A higher density implies more connections between the nodes relative to the total possible connections.

Average Clustering Coefficient: This metric assesses the degree to which nodes in a network tend to cluster or form tightly knit groups. A higher value suggests a greater tendency for nodes to form local clusters.

Diameter: The longest shortest path between any two nodes in the network. A larger diameter can indicate a more spread out network.

Average Shortest Path Length: Represents the average number of steps required to connect any two nodes in the network. A higher value can suggest a less interconnected network.

Connected Components: The number of distinct subnetworks or groups within the network that are not connected to each other. A single connected component means the network is entirely interconnected.

Modularity: This measures the strength of division of a network into modules or communities. Higher modularity indicates a network with well-defined clusters or groups of nodes.

Assortativity Coefficient: Reflects the tendency of nodes to connect to other nodes that are similar or dissimilar to themselves. Positive values indicate a preference for similar nodes to connect (assortative), while negative values suggest a tendency to connect with dissimilar nodes (disassortative).

Average Edge Weight: Indicates the average strength or importance of the connections in the network. Higher weights can imply stronger or more significant relationships.

Degree Centrality: Measures the number of direct connections a node has. Nodes with higher degree centrality are more central in the network.

Betweenness Centrality: Reflects the number of times a node acts as a bridge along the shortest path between two other nodes. High betweenness centrality indicates a node with significant influence over the flow of information in the network.

Closeness Centrality: Represents the average length of the shortest path from a node to all other nodes in the network. Nodes with lower closeness centrality can reach other nodes more quickly.

Eigenvector Centrality: A measure of a node's influence in the network, considering the influence of its neighbors. Nodes with high eigenvector centrality are connected to many well-connected nodes.

PageRank: A measure of a node's importance based on the quantity and quality of links to it. It considers both the number and the significance of the connections to determine the rank of a node.

3.5 Louvain Method

The Louvain method is an algorithm used for detecting communities in large networks. Developed in 2008, it's effective and efficient for analysing large datasets [14].

The Louvain method, renowned for its efficiency in network analysis, is characterised by several key features and advantageous attributes that make it a compelling choice for our project. Central to its functionality is the optimisation of 'modularity,' a process that identifies communities within a network by ensuring densely connected groups. This is achieved through a hierarchical approach, where nodes are progressively combined into larger communities, with modularity optimised at each step. This method proves particularly effective for large networks, as it can handle extensive datasets rapidly and scales well with network size [14].

The decision to utilise the Louvain method in our project is driven by several factors. Primarily, its capability to process large volumes of network data swiftly aligns well with our project's requirements, especially considering the scalability needed for larger networks. The method is also renowned for its accuracy in community detection, reliably identifying distinct groups within a network. Moreover, the simplicity of its implementation is a significant advantage; it can be easily integrated into our project and adapted for various types of networks. This versatility is further demonstrated by our ability to create diverse networks

by varying nodes (such as companies, countries) and edges (like organisations, people), offering a wide range of analytical possibilities [14].

We chose the Louvain method for our project because it's fast, accurate, easy to implement and versatile, making it a great tool for analysing complex network data. However, there are 2 main reasons for choosing this technique to cluster. First, although our network is not very complex, BDO can easily apply bigger and more complex networks for clustering as this algorithms works well with large networks. Secondly, after testing different clustering techniques, the Louvain method seemed to work best on our type of data/network.

3.6 K-means

Introduction to K-Means Clustering:

K-Means clustering stands out for its simplicity, efficiency and versatility, making it a widely used technique in clustering tasks. Its straightforward nature contributes to its popularity, as it is easy to understand and implement. This simplicity doesn't come at the cost of performance; K-Means is particularly efficient, handling large datasets effectively. Its versatility allows application across various fields, such as market segmentation, document clustering and image compression [15].

The elbow method, associated with K-Means, is a strategy to identify the optimal number of clusters in a dataset. This involves executing K-Means clustering for a range of cluster values (k) and then calculating the sum of squared distances from each point to its cluster center for each k. By plotting these values, the 'elbow point' can be identified – a point where the rate of decrease in the sum of squared distances sharply changes. This elbow point is significant as it suggests the most appropriate number of clusters for the dataset [15].

The combination of K-Means and the elbow method is chosen for its effectiveness in accurately determining the number of clusters in a dataset. The simplicity and clarity of both techniques make them not only easy to implement but also straightforward in interpreting results. Furthermore, they are instrumental in extracting and understanding meaningful patterns from the data, thereby providing valuable insights [15].

K-means clustering, augmented with the elbow method, offers a powerful and efficient approach for data segmentation in machine learning. Its simplicity, effectiveness and interpretability make it a valuable tool in the field, especially the financial data we received from the project partner BDO.

4 Data

The datasets for our project are provided by our partner BDO Austria GmbH. In total, we have nine different data sets.

Four JSON files: The datasets contain various companies featured in the Euro Stoxx 600 index, detailing their interactions with diverse entities, including companies, locations, and individuals, for the years 2011, 2013, 2016 and 2022. These datasets are pivotal for the project, serving primarily as the foundation for constructing the networks.

Four Excel files with financial data: As our project evolved, we realised the need for more detailed financial data. In response to this, BDO provided us with four additional datasets containing financial data for the companies listed in the JSON files. Corresponding to the same years (2011, 2013, 2016 and 2022), these Excel sheets provide specific financial metrics for each company, which include:

1. `IQ_TOTAL_REV`: Total Revenue
2. `IQ_COST_REV`: Cost of Revenue
3. `IQ_COGS`: Cost of Goods Sold
4. `IQ_GP`: Gross Profit
5. `IQ_OPER_INC`: Operating Income
6. `IQ_NI`: Net Income
7. `IQ_TOTAL_ASSETS`: Total Assets
8. `IQ_TOTAL_LIAB`: Total Liabilities
9. `IQ_TOTAL_EQUITY`: Total Equity
10. `IQ_EBITDA`: Earnings Before Interest, Taxes, Depreciation and Amortization
11. `IQ_NET_DEBT`: Net Debt
12. `IQ_CURR_TAXES`: Current Taxes
13. `IQ_RETURN_ASSETS`: Return on Assets
14. `IQ_RETURN_EQUITY`: Return on Equity
15. `IQ_GROSS_MARGIN`: Gross Margin
16. `IQ_EBITDA_MARGIN`: EBITDA Margin
17. `IQ_NI_MARGIN`: Net Income Margin
18. `IQ_TOTAL_DEBT_EQUITY`: Total Debt to Equity Ratio
19. `IQ_TOTAL_DEBT_CAPITAL`: Total Debt to Capital Ratio
20. `IQ_INT_EXP_LTD`: Interest Expense on Long-Term Debt
21. `IQ_CAPEX`: Capital Expenditures
22. `IQ_CASH_OPER`: Cash from Operations
23. `IQ_MARKETCAP`: Market Capitalization

This allows us to track and analyse the financial performance and trends of these companies across different years. With these data, we could explore deeper insights. We were not just looking at financial stability, but also at how these companies are interconnected within their business networks, offering a more holistic view of corporate health and relations.

One Excel file with labelled data: The most recent dataset, from October 2023, provides labelled information about the Stoxx 600 companies. It includes a range of attributes for each company: Name, ISIN, Location (Address, City, Country), Primary Industry, Exchange, Employee Count, CEO (Chief Executive Officer), CFO (Chief Financial Officer), Rating and up to ten each of Suppliers, Customers and Board Members. This data set enables us to enrich the JSON files with specific labelled attributes. We deliberately selected four attributes — home country, home city, exchange and industry - assuming that these will generally not change for a company over the years.

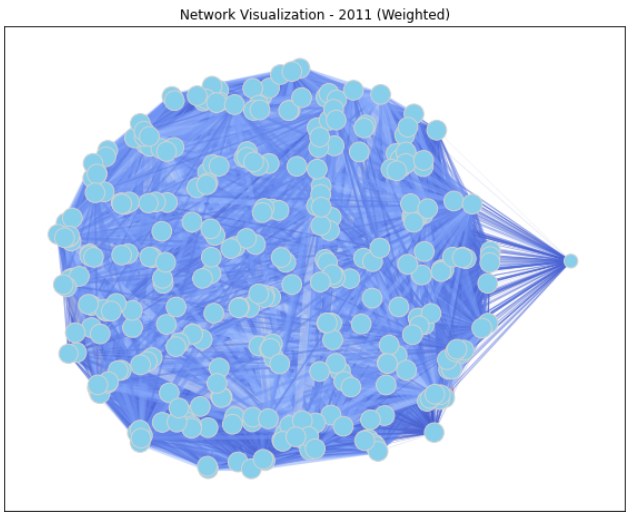

Figure 1: Network Graph for 2011

The graph above represents the 2011 network in its raw, unfiltered state. Our nodes, each representing a distinct company, are connected by an immense amount of entities. The weighting is constructed of multiple ranked labels, which we will discuss later. In its current form, the graph's high density significantly hinders our ability to discern the individual connections and understand the broader relationships among the companies depicted. The tangled array of lines and points effectively masks the true structure of the network, thus complicating our efforts to unravel the complex ways in which these companies interact. recognising this issue, it becomes imperative to systematically clean and preprocess the data. This preparatory step is key to cutting through the clutter and revealing a clearer, more navigable map of corporate interrelations. Without this essential data management, the graph remains overly complex, obstructing our path to gain insights and draw well-founded conclusions from our networks.

4.1 Labelling the Data

In our network analysis, labelling the attributes allows for the identification of specific attributes or types of connections. By assigning different labels and weights, we can prioritise certain connections over others. This is essential to dissect the complex web of interactions in a network and to understand how different entities influence each other.
labelling the data facilitates the segmentation of the overall network into subnetworks, each representing a different type of connection or interaction. This segmentation enables a more targeted analysis, allowing us to delve deeper into specific aspects of the network.

To achieve the best result, we divided the labelling process into three parts (three loops):

1. **Named Entity Recognition**: First, we used Spacy's NER as a basis on which we built our further analysis.

2. **Overwriting Spacy's labels for countries, cities, organisations and departments**: In the next step, we used the Country Converter, a list of cities, typical company endings and typical department names to further enhance Spacy's NER.

3. **Filtering the attributes labelled as persons**: Persons are a very important category in our network analysis, but are difficult to identify for an NLP model. Therefore, we used additional methods to improve the identification of persons.

We will go into further detail for each of the three steps in the following chapter.

4.1.1 Spacy's Named Entity Recognition (NER)

Spacy's English TRF model was chosen for its efficiency and accuracy in recognising various types of entities. The model was applied to each attribute within our dataset. For each company in the dataset, we iterated over its attributes. The Spacy model processed each attribute string, returning an object containing identified entities. These entities were then extracted and their labels (e.g., 'GPE' for geographical locations, 'PERSON' for people's names) were collated into a string. The output of the NER process was a list of dictionaries, each containing the company name, the original attribute and the derived label. To facilitate further analysis and reporting, this list was transformed into a more structured format using Pandas DataFrame. The utilisation of Spacy's English TRF model for Named Entity Recognition proved to be a pivotal step in our data processing pipeline. By applying this advanced NLP technique, we were able to efficiently categorise the unlabelled attributes of our dataset, thus enriching the data with meaningful labels that can be utilised for further analysis and insights. The resulting labelled dataset now serves as a foundational element for our ongoing project, allowing a deeper understanding of the relationships and characteristics of companies within the Euro Stoxx 600 index. [10]

4.1.2 Country Converter

In our pursuit to further refine and standardise our dataset for network analysis, we integrated the use of the Country Converter (coco) tool. This step was crucial in identifying and standardising country names, ensuring consistency and accuracy in our data, especially for analyses based on shared attributes among companies. The Country Converter, a versatile and user-friendly Python package, allowed us to convert various country names into a standardised short format (name_short). We iterated over each row of our labelled data, which we had previously processed using Spacy's Named Entity Recognition. In this process, every attribute identified as a potential country name was passed through the Country Converter. When a match was found and the conversion to the standardised short name was successful, we updated the attribute value to this standardised name. Simultaneously, we changed the associated label to 'country', thus achieving uniformity in representing countries across our dataset. This method played a key role in improving the reliability of our network analysis. By standardising country names, we minimised the discrepancies that often arise from various representations of the same country (e.g., United States of America and USA are both converted to United States). This uniformity allowed for more accurate associations and comparisons between companies based on their country attributes, thus enriching our analysis of the interconnectedness within the Euro Stoxx 600 index. [12]

4.1.3 Cities Dataset

To further enhance the accuracy and granularity of our labelling algorithm, we integrated an external dataset specifically focused on cities. This dataset, sourced from all-countries-and-cities-json on GitHub, provided a comprehensive list of cities around the world. Incorporation of this city dataset was a strategic move to identify and label city names accurately within our dataset, thereby supplementing our earlier steps involving country standardisation. After the initial step of country name conversion using the Country Converter, we proceeded to validate and label city names. In this process, each attribute not identified as a country was cross-referenced with the city dataset. When a match was found, we updated the label to 'city', thereby distinguishing it from other entity types. The addition of city-level data significantly increased the depth of our network analysis. By being able to distinguish between cities and other types of attributes, we could draw more nuanced connections and insights, especially in understanding geographical influences and relationships among the companies. This city dataset not only increased the precision of our attribute labelling but also enriched our dataset with a layer of geographical specificity.

4.1.4 Company Endings

Building on our previous steps of categorising countries and cities, we advanced to the crucial task of identifying organisations within our dataset. recognising that company names often follow specific naming conventions, we employed a method to detect organisation names based on a list of typical company endings. To facilitate this process, we compiled a list of common company suffixes, such as 'GmbH', 'SA', 'Ltd',

'Limited', 'AG' and 'Group'. These suffixes are widely used across various countries and legal systems to denote a business entity. For attributes not categorised as countries or cities, we checked if they ended with any of the listed company endings. This method proved highly effective in distinguishing organisational entities from other types of data. By focusing on the structural characteristics of company names, we were able to accurately label a significant portion of our attributes as organisations. This step was particularly important because it allowed us to identify and analyse the network of relationships between companies and other entities within our dataset.

4.1.5 List of Departments

The final step in our second labelling iteration involved categorising attributes as company departments. This step was crucial for attributes that had not been previously identified as countries, cities, or organisations. It aimed to recognise and label various functional areas within companies, thus adding another layer of detail to our dataset. To achieve this, we compiled an extensive list of department names commonly found within organisations. This list encompassed a wide range of departments, from 'human resources' and 'legal affairs' to 'marketing', 'logistics' and 'information technology'. The source for some of these department names was simplicable.com, which provided a comprehensive overview of typical corporate divisions. We then examined each remaining unlabelled attribute in our dataset. If an attribute contained a department name from our list and had not yet been assigned a label, we labelled it as 'dep' (short for department). This meticulous process of department labelling enriched our dataset by providing insights into the internal structures and functions of the companies. It allowed for a more nuanced analysis of the companies' compositions and the potential relationships between different departments across various organisations.

4.1.6 First and Last Names Dataset

The subsequent phase of our project focused on refining the accuracy of person identification within the dataset. This process involved utilising datasets that comprise the first and last names obtained from data.europa.eu, which significantly aided in discerning individual names from other entities. We developed two distinct approaches for updating the 'person' label, one considering both first and last names and another considering only first names.

First and Last Name Approach:

In this approach, we employed two separate datasets: one for first names and another for last names. These datasets were filtered based on the frequency of names to ensure the consideration of more common names, thereby increasing the likelihood of accurate identification. The process involved iterating over each attribute in our dataset. We first eliminated any attribute containing parts of a company's name, assuming that such attributes were unlikely to represent a person. Next, we checked whether the attribute contained both a first and a last name from our lists. If both were present, the original label was retained; otherwise, the label was updated to 'other'.

First Name Only Approach:

This approach was similar to the first but solely focused on first names. It was particularly useful in instances where last names were not available or the context made it difficult to determine last names.

In both approaches, we used Spacy's NER model to extract potential person names from each attribute. This added a layer of natural language processing, enhancing the capability to recognise names even in more complex or unstructured data. After a lot of different approaches - using the combination of first and last names or only the first names and fine-tuning the parameters - we came to the conclusion that a combination of first and last name detection works best. Regarding the parameters, we filtered the first name dataset to count over 350 and the last name dataset to count over 100. The reason for the lower count in the last names dataset is that last names do not occur as often as first names. The column 'count' in the datasets indicates how often the respective name appeared in the survey.

This enhancement in person labelling significantly increased the accuracy of our dataset, particularly in identifying individuals. By distinguishing personal names from other entities more effectively, we were able to provide a clearer analysis of individual involvement and connections within the network. This level of precision was vital for in-depth network analysis, especially when exploring the roles and influences of individuals in corporate structures and relationships.

4.2 Customizing Label Weights for the Networks

The assignment of different weights to various labels plays an important role in highlighting the relative importance of connections between our nodes. This approach is instrumental in our analysis, as it allows us to quantify the significance of different types of relationships within the network.

We specifically focused on the labels "person", "country", "city", "org" (organisation) and "other." The weight distribution was deliberately chosen based on the perceived importance of these connections in our network context. Entities labelled as "person" were assigned the highest weight of 10, underscoring the critical role individuals play in business networks, often acting as key decision makers or influencers. Organisations were given a weight of 7, reflecting their significant, despite slightly lesser, impact compared to individual persons. "country" and "city" labels received a weight of 5, acknowledging the importance of geographical factors in corporate interactions. The label "other" was assigned the lowest weight of 1, indicating that there is a connection between the two nodes but rather negligible.

Label	**Weight**
person	10
org	7
country	5
city	5
other	1

Table 1: Label Weights within our Networks

This differential weighting approach is rooted in the understanding that not all connections in a network have the same value or influence. By assigning higher weights to more critical connections (such as "Person" or "Org"), we can more accurately model the real-world complexities and hierarchical nature of business networks.

4.3 Descriptive Analysis

By analysing our metrics of the networks 3.4, we could uncover patterns, identify key influencers, understand group dynamics and predict how changes within the network might unfold. This holistic view is invaluable for strategic decision-making, risk assessment and identifying opportunities within the business landscape.

Metric	2011	2013	2016	2022
Number of Nodes	433.000000	4.26e+02	413.000000	594.000000
Number of Edges	92883.000000	9.0051e+04	84958.000000	175945.000000
Average Degree	429.020785	422.7746	411.418886	592.407407
Density	0.993104	0.9947639	0.998590	0.999001
Average Clustering Coefficient	0.994904	0.9963700	0.998729	0.999141
Diameter	2.000000	2.000000	2.000000	2.000000
Average Shortest Path Length	1.006896	1.005236	1.001410	1.000999
Connected Components	1.000000	1.000000	1.000000	1.000000
Modularity	0.000875	1.110223e-16	0.000000	0.000000
Assortativity Coefficient	-0.017292	-0.010933	-0.011194	-0.006756
Average Edge Weight	19.294984	19.88767	22.233198	26.280156

Table 2: Network Metrics across the Years

The table above presents an overview of the network metrics for the years 2011, 2013, 2016 and 2022. While these metrics offer general insights into the structure of our networks, the practical application of these metrics in our analysis seems challenging.

Primarily, the sheer size and density of the network posed considerable obstacles. The network's high density, as indicated by values approaching 1 and the substantial number of nodes and edges, especially in the year 2022, resulted in a highly interconnected network. This extreme level of interconnectivity, while indicative of a robust network structure, made it difficult to extract meaningful patterns or identify distinct clusters within the network.

Additionally, metrics such as modularity, which ideally would help in understanding the network's division into modules or communities, were near zero or insignificant, suggesting a lack of clear modular structure due to the network's density. Consequently, despite the rich dataset and detailed metrics, the ability to utilise these metrics to derive actionable insights was limited, underscoring the need for alternative analytical approaches or techniques to manage and interpret the complexity inherent in such dense networks.

D-A-CH network

Metric	2011	2013	2016	2022
Number of Nodes	102.000000	100.000000	97.000000	131.0000
Number of Edges	5151.000000	4950.000000	4656.000000	8515.0000
Average Degree	101.000000	99.000000	96.000000	130.0000
Density	1.000000	1.000000	1.000000	1.0000
Average Clustering Coefficient	1.000000	1.000000	1.000000	1.0000
Diameter	1.000000	1.000000	1.000000	1.0000
Average Shortest Path Length	1.000000	1.000000	1.000000	1.0000
Connected Components	1.000000	1.000000	1.000000	1.0000
Modularity	0.000000	0.000000	0.000000	0.0000
Assortativity Coefficient	NaN	NaN	NaN	NaN
Average Edge Weight	0.085447	0.088289	0.066101	0.0469

Table 3: D-A-CH Network Metrics

The updated table offers a focused examination of the D-A-CH network, isolating the connections across the years 2011, 2013, 2016 and 2022. It reveals a refined understanding of the structure and dynamics of the

subnetwork. Unlike the broader network previously discussed, the D-A-CH subnetwork showcases a more manageable number of nodes, ideal for in-depth analysis. The quantity of edges is also substantial, striking a balance that allows for a meaningful examination of the network's connectivity without the overwhelming density seen in earlier analyses. This configuration lends itself to a more nuanced interpretation of the Average Edge Weight metric 3.4, which now provides valuable insights into the connection strength between nodes, as the network avoids the pitfalls of excessive density.

Despite these improvements, certain network metrics continue to offer limited utility even when dissecting the network's complexities. The constant high density and average clustering coefficient, with values persistently at 1, paint a picture of a network where every node is directly connected to every other node. This level of interconnectivity, while indicative of a closely-knit economic or corporate structure within the D-A-CH region, renders metrics like modularity (which remains at 0) ineffective for identifying distinct community structures or sectors within the network. The lack of discernible clusters complicates efforts to analyse the network's response to economic or social changes in a granular manner.

Furthermore, the diameter and average shortest path length of the network remain fixed at 1, underscoring the exceptionally integrated nature of the D-A-CH economic landscape. This characteristic, though beneficial for rapid communication and collaboration, further diminishes the relevance of certain analysis metrics that thrive on identifying hierarchical or layered structures within a network. The assortativity coefficient (NaN) remains not applicable, reflecting the unchanged status of the network's complete connectivity, which eliminates the potential for observing patterns of preferential attachment or assortative mixing.

In conclusion, the D-A-CH network's revised structure, with an optimal number of nodes and a significant number of edges, facilitates a more meaningful analysis than previously possible, especially in terms of connection strengths.
However, the persisting uniformity in several network metrics highlights the challenges of using traditional analysis approaches to fully comprehend the intricacies of the D-A-CH region's economic and corporate landscape, pointing to the need for alternative methods to capture the dynamics within this integrated network.

For a more detailed exploration of the D-A-CH network's structure and its implications, readers are directed to the appendices. These sections contain comprehensive graphs and extended analyses that delve into the nuances of the network. The graphical representations provide a visual account of the network's evolution over the selected years, offering insights into the connectivity patterns, strength of relationships and the overall network dynamics (see section: 10).

5 Data Analysis

When discussing data analysis, it is crucial to highlight the various approaches we employed to analyse our datasets. The availability of labelled files introduced new methods to construct our networks. In the upcoming chapters, we will first emphasise our targeted exploration of specific regions and present the results derived from each distinct approach.

5.1 Selective Focus: D-A-CH Region

The decision to concentrate our network analysis predominantly on the D-A-CH region, comprising Germany (D), Austria (A) and Switzerland (CH), was driven by several factors. Given that the project partner is BDO Austria, there is an inherent interest in delving into the corporate and industrial interconnections within this geographical cluster.

Furthermore, the time constraints inherent in a project of this scope required a more focused approach. By narrowing most parts of our network analysis to the D-A-CH region, we are able to conduct a more thorough exploration of the networks and their structures, rather than a superficial sweep across a broader, yet less relevant, global landscape. This targeted analysis allows for a deeper understanding of the regional nuances that shape the business environment.

To give you a short overview, the D-A-CH region stands as a powerhouse in Europe, known for its robust economy and high living standards. This area showcases a vibrant e-commerce sector with Germany at its heart, holding the title of Europe's top consumer market [16]. When we look at mergers and acquisitions, Germany emerges as a central hub within the region, with a marked increase in deal values and volumes, especially in 2021. This surge in M&A (Mergers and Acquisitions) activities, powered by significant international interest, underscores the global confidence in the region's economic prospects. The key sectors driving these deals include real estate, industrials and chemicals [17]. Despite the global disruptions caused by the COVID-19 pandemic, the D-A-CH region demonstrated exceptional recovery capabilities, aided by efficient public health systems and strong governmental trust, particularly in Germany. This resilience is mirrored in the rapid bounce-back of the M&A market, suggesting a robust and optimistic economic forecast [18].

5.2 Companies as Nodes

When coming to our network visualisations, we first will have a look at the companies as nodes graphs. This approach is already particularly focusing on companies located in the D-A-CH region. In our visualisations, each node signifies a distinct company, capturing the essence of their interactions and connections exclusively within the D-A-CH context.

Moreover, the choice to limit our visualisations for companies to the D-A-CH region not only simplifies the complexity of our visualisations, but also provides a clear and focused lens through which to understand the business dynamics prevalent in these countries.

5.2.1 Construction

In constructing the network visualisations, particularly for the countries as nodes representation, we developed different functions for various networks, which transform our dataframes into insightful visual graphs. The underlying code structure for these functions remains consistent, necessitating an initial explanation of its fundamental components:

The layout of the networks is determined using 'nx.spring_layout', a function that positions nodes using a force-directed algorithm, simulating a physical system to arrange the nodes in an aesthetically pleasing manner. This method provides a clear and intuitive spatial representation of the network, with the 'k' parameter controlling the spread of nodes [13].

One of the distinctive features of our visualisation is the usage of colour and size to convey information. Node sizes are dynamically adjusted based on their degree, scaled against the maximum degree in the network, providing an immediate visual cue to the node's relative importance. For edge visualisation, we employ a dual-colour scheme. The edge widths are proportionate to their weights, enhancing the visual emphasis on stronger connections. Edges with weights above a certain threshold are coloured using a gradient from the 'Reds' colourmap, visually indicating the intensity of connections (the stronger the connection, the redder it is). Conversely, less significant edges (below the threshold) are depicted in 'gainsboro', a light grey colour, ensuring they remain visible but do not dominate the visualisation.

Figure 2: Colourmap: "Reds"

The node colouration is particularly innovative, utilising the 'coolwarm' colourmap to represent a combined metric of degree and the sum of edge weights. This colour mapping is normalised, allowing for a comparative view of node influence based on both their connectivity and the strength of their connections.

Figure 3: Colourmap: "coolwarm"

In the rendering process, we selectively display node labels based on a degree threshold to maintain clarity and avoid clutter. This thoughtful approach to labelling ensures that the visualisation remains informative without being overwhelmed by text.

This visualisation function, with its careful consideration of layout, colour and scale, provides a clear and detailed representation of the various networks, highlighting key relationships and patterns within the data. It is important to note that this function is tailored for the visualisations of the companies and part industries as nodes networks. We have this core implementation of the function for each important subnetwork ("person", "org", "country" and "city") with adjusted parameters and thresholds, ensuring each visualisation is optimally configured for its specific context.

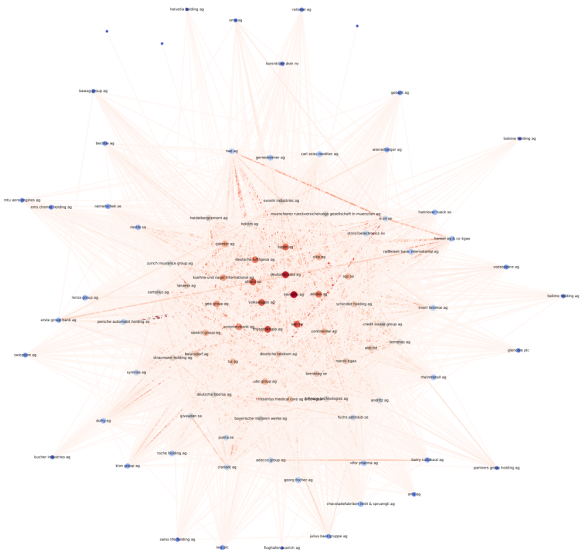

Figure 4: D-A-CH Region, Companies as Nodes Network - 2011

The graph presented above demonstrates the network's effectiveness in identifying key players within its structure. By visually distinguishing nodes with higher degrees of connectivity, the graph effectively highlights the most influential or central entities in the network through the colours (coolwarm), signifying their important roles. However, despite this clarity in pinpointing key players, the graph also reveals a limitation in its current form: the high density of connections between nodes. This density results in a visually complex and intertwined network, making it challenging to discern specific connections or extract detailed insights about the relationships between individual nodes. Thus, while the network excels in revealing prominent members, it simultaneously necessitates further refinement to allow for a clearer understanding of the inter-nodal relationships. For a detailed overview of all the visualisation, we refer to the appendices (10).

5.2.2 Analysis of Label-Specific Subnetworks

In our comprehensive analysis of network structures, one of the most important aspect is the dissection of the network into various subnetworks based on distinct labels. This approach is instrumental in unraveling the nature of our network data, allowing us to delve deeper into specific types of interactions and relationships. Each subnetwork is defined by a particular label, such as "person", "organisation", "country", "city", or "other", providing a focused view on how these individual elements interact within the broader network. This segmentation facilitates a more granular and insightful examination of the network, enabling us to identify and analyse patterns and trends that might be obscured in a more generalised network view.

By visualising these subnetworks separately, we gain clarity and a better understanding of the distinct dynamics and characteristics that each label brings to the network. This method not only enhances the specificity of our analysis but also aids in generating more targeted and relevant insights, crucial for understanding the complex interplay of elements within the network.

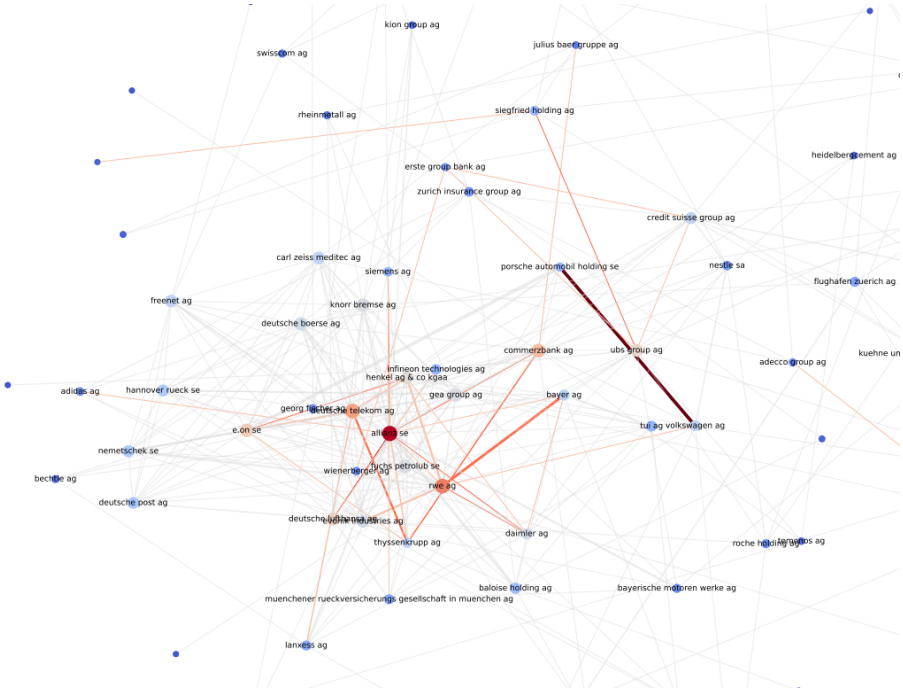

Figure 5: D-A-CH Region, Companies as Nodes Network (Label: Person) - 2011

The snipped from the graph above serves as an illustration of the benefits of segmenting the network into subnetworks. The subnetwork which is displayed above is the "person 2011 network". In this refined visualisation, we focus on the label 'person', providing a clearer view of the network's structure as it stood in 2011 within this subnetwork. The reduction in overall density is clearly visible. This targeted approach now not only highlights the key players within this particular subnetwork but also brings out the most significant connections between nodes.

This visualisation validates our strategy of dividing the network into subnetworks, demonstrating that such segmentation is crucial for a deeper and more nuanced analysis of both the nodes and their connections. In our analysis, we included all subnetworks for the important labels "person", "org", "city" and "country" across the four years. This section displays the key findings of our exploratory data analysis.

Person Networks: Findings

Among the different subnetworks, the Person network consistently emerged as the least densest one. Even though this low density makes it challenging to discern significant connections, we came across some interesting attributes.

A key observation is the recurrent presence of Klaus Heubeck (Actuary) across different networks and years, establishing him as the most important figure. Alongside Klaus Heubeck, other individuals frequently appeared in these networks, serving in diverse roles such as chairpersons, advisors (Andreas Pohlmann), politicians (Olaf Scholz, Donald Trump), investors (Paul Achleitner) or high-tier managers (Wolfgang Mayrhuber, Ulrich Lehner).

However, it is Klaus Heubeck's consistent presence in all networks and his pronounced relevance that stands out distinctly. His involvement across various years and networks underscores his pivotal role within these networks. Furthermore, he is the only important person which still appears in the most recent year 2022, which makes him the only really important person in the hole network from a present perspective.

Organisation Networks: Findings

For the "org" subnetwork, our findings highlighted a recurring theme of banks as central players within these networks over various years. Prominent among these are institutions like Credit Suisse, UBS Group AG and Deutsche Bank AG, which have consistently emerged as key entities in the networks. These banks notable presence underscores their significant influence within the corporate landscape of the D-A-CH region. It also makes sense, that the banking sector performs best when the connections between the companies are only based on organisations.

Another aspect of the organisation subnetwork is the prevalence of shared attributes that transcend mere corporate identities. We observed that different stock exchanges and commissions frequently appeared as common links between various companies. Additionally, the presence of major global companies such as BlackRock Inc., KPMG AG, Goldman Sachs or McKinsey & Company was noted.

Geographical Networks - Cities and Countries: Findings

Our analysis of the geographical subnetworks, reveals that these are the densest among all of the subnetworks we explored. The most frequently occurring attributes within these networks are those associated with the most economically significant countries and cities, which serve as vital nodes in the network due to their global economic influence.

For the country-based subnetwork, the United States, China, Germany, the United Kingdom and Japan consistently emerge as key attributes, reflecting their roles in the global economy. Interestingly, in 2022, Ukraine appears as the fourth most common shared attribute within the country network. This notable presence is likely to the increased global focus on Ukraine due to its conflict with Russia, highlighting how geopolitical events can impact network dynamics.

In the city-based subnetworks, hubs such as Munich, Paris, Zurich, Vienna, Tokyo and Berlin are featured prominently. These cities, known for their significant contributions to global and regional economies, seemed to be most relevant for the network. Their recurring presence underscores the importance of urban centers in shaping economic and corporate linkages, both within the D-A-CH region and beyond.

In general, the geographical subnets provide a comprehensive view of how global economic centres and key urban areas interconnect and influence the broader network. Their density and the prominence of certain countries and cities underscore the critical role these geographical entities play in the fabric of global economic relations.

5.3 Industries as Nodes

For the industries as nodes networks, we applied the same visualisation logic as used in the company nodes networks. The focus was on ensuring that the visual representations accurately reflected the underlying data, with an emphasis on clarity and comprehensibility. With the provided datasets from BDO, we were able to merge the industries for each companies to our datasets, which enabled us to create networks, looking at the relation between industries.

Given this consistent approach in visualisation, we will now delve into the most salient findings from the industry networks analysis. These findings provide insights into the inter-industry dynamics and highlight key trends and patterns that emerged over the years. This section will showcase how the different industries are interconnected, which industries are most central in the network and how these relationships have evolved over time, providing a deeper understanding of the industrial fabric of the D-A-CH region.

Persons Networks: Findings

In 2011, while there were notable instances of interconnectivity among certain industries like multi-line insurance and automobile manufacturers, these connections did not follow a strong, discernible pattern. The network's structure was such that clear and consistent industry dominance was not evident, reflecting the unique characteristics of personal connections within the corporate landscape.

The same trend continued in 2013 and 2016. Although certain industries, such as diversified capital markets and speciality chemicals, appeared to be more connected in some instances, these connections did not exhibit a consistent pattern that could signify a significant shift or trend within the network. The connections were present but not dense enough to establish a dominant industry or a clear relationship.

Coming to the year 2022, the network landscape appeared more balanced with several industries showing higher levels of connectivity but without a single dominant industry. Pharmaceuticals continued to be a central player, exhibiting strong connections with both specialty chemicals and diversified capital markets. This pattern reflects a diversification in industry relationships, with no single sector overwhelmingly dominating the network.

This lack of a significant pattern in the 'person' networks for these years highlights the challenges in drawing broad conclusions from networks based on less frequent but potentially influential labels like 'person.' The rare occurrences and unique nature of personal connections within the corporate network lead to a less dense network structure, making it difficult to identify dominant trends or significant inter-industry relationships as clearly as in other types of networks.

Organisation Networks: Findings

In contrast to the person network, the organisation network exhibits a visible shift in the prominence of specific industries, pointing towards evolving economic dynamics within the network.

In 2011, multi-line insurance emerged as the standout industry in terms of connectivity. Its significant presence underscores the pivotal role insurance played in the corporate interactions and economic structure of that period (after the financial crisis).

However, from 2013 onwards, we observed a shift, with diversified capital markets ascending to become the most connected industry across the network. This change is one in the network's core, highlighting the growing influence and centrality of capital markets. Unlike the transient patterns observed in the person network, this shift in the organisation network signifies a lasting transformation, with diversified capital markets maintaining their dominance throughout the subsequent years.

Alongside this, other industries such as pharmaceuticals have progressively increased their connectivity within the network. This trend suggests not just a static change but a gradual evolution toward a more interconnected industry landscape, with pharmaceuticals and others building stronger ties across the network.

The general trend observed in the organisation network is one of increasing connectivity over the years. This trend indicates a deepening of economic interdependencies and a broadening of the network's scope as more industries become integral to the network's structure.

Geographical Networks - Cities and Countries: Findings

In exploring the geographical networks, we observed a general trend towards increased connectivity. This growing interconnectedness is indicative of a more integrated economic landscape where geographical boundaries are becoming less of a barrier to industry relationships. Despite this overall increase in connections, pinpointing key shifts or trends within the network proves challenging due to the network's comprehensive connectivity.

Within this densely connected framework, certain industries stand out for their higher levels of interconnection. In particular, the Pharmaceutical industry, both in terms of countries and cities, along with the Automobile Manufacturing industry (at the city level), exhibited more connections than other sectors. This observation suggests that these industries play a central role in the network's economic interactions, potentially driven by their global significance and the intricate supply chains that span multiple regions.

Despite the clear evidence of a highly connected network, the dense nature of these connections makes it difficult to discern significant shifts or identify clear trends over the years. The complexity of the network and the uniform increase in connectivity across various sectors and geographical regions highlight the challenges in isolating specific changes or pointing to particular industries or locations as pivotal nodes within the broader economic context.

5.4 Countries as Nodes

In contrast to the previous chapters, all the data was used for the countries as nodes network and is therefore not limited to the DACH region. Each country represents all companies that have their headquarter within its boarders. Therefore, all attributes of each company were assigned to its home country. After identifying shared attributes between the countries and filtering the data by industry, we built the following networks for the years 2011, 2013, 2016 and 2022. In the following chapters, we will only provide the most recent plots from 2022. To see the progression from 2011 to 2022, we refer to our jupyter notebook 'Countries and Industries as Nodes.ipynb' where we implemented interactive plots with a dropdown menu that makes it possible to switch between the years.

5.4.1 All Industries

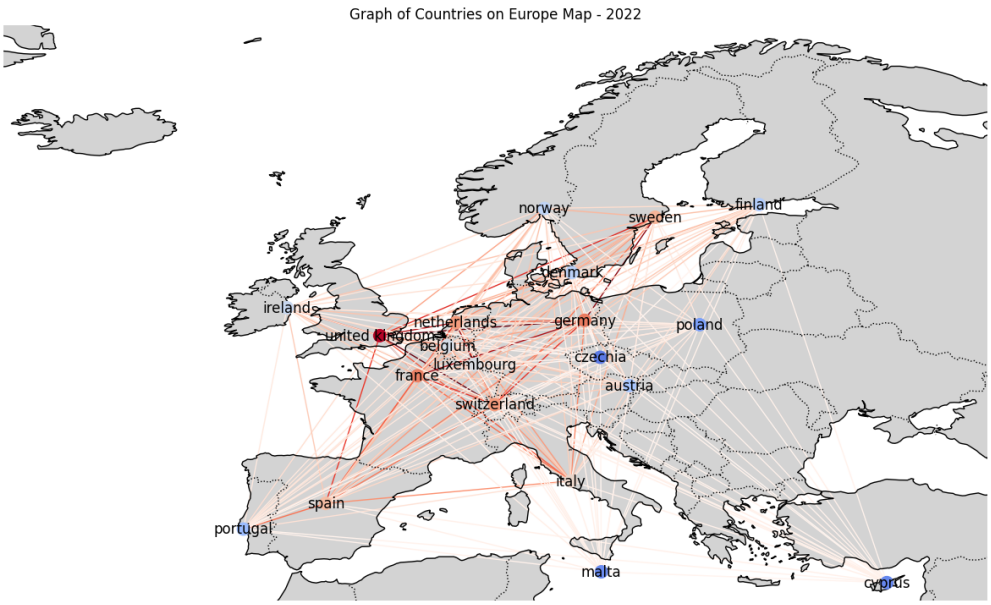

Figure 6: Countries as Nodes - All Industries - 2022

The United Kingdom's central role in the European economic network can be attributed to several key factors, primarily its robust financial sector, diverse economy and historical influences.

1. **Policy Initiatives**: The UK government has implemented several policies aimed at stimulating economic growth. This includes increasing the minimum wage, raising state pensions and enhancing working age benefits. Additionally, a reduction in National Insurance and potential cuts in income tax are expected to boost disposable incomes significantly. These measures are anticipated to improve living standards for many Britons [19].

2. **Labor Market Resilience**: The UK's labor market has shown considerable resilience. The unemployment rate, though slightly increased, remains relatively low. This stability in the job market is crucial for economic growth. Moreover, wage growth, which has been high, is beginning to show signs of slowing down, aligning with the declining inflation rates [20] [21].

3. **Economic Growth Forecast**: While the UK's economic growth is forecasted to be modest in 2024, it is expected to show improvement. Factors contributing to this include an increase in real disposable income and a diminishing impact from previous monetary tightening. However, it's important to note that the UK's growth rate is still expected to lag behind that of the US and the Euro area [22].

4. **Inflation and Interest Rates**: Inflation rates are dropping faster than anticipated, which positively impacts the economy, particularly the poorer households. The Bank of England is expected to maintain steady interest rates, with potential rate cuts anticipated later in the year. This scenario is likely to benefit mortgage borrowers and consumers in general [19] [20].

5. **Energy Costs and Consumer Prices**: There is an expected decrease in household energy bills, which will reduce overall expenses for households. This, coupled with the falling consumer price inflation, is contributing to a more favorable economic environment [19].

6. **Challenges**: Despite these positive aspects, the UK still faces challenges like stickier inflation compared to other regions and issues like long-term sickness in the workforce impacting labor force participation [20].

Each of these elements contributes to the UK's unique role and influence in the global economy. However, it is important to mention that this is only a fraction of all the factors that influence the UK's leading position.

5.4.2 Diversified Banks

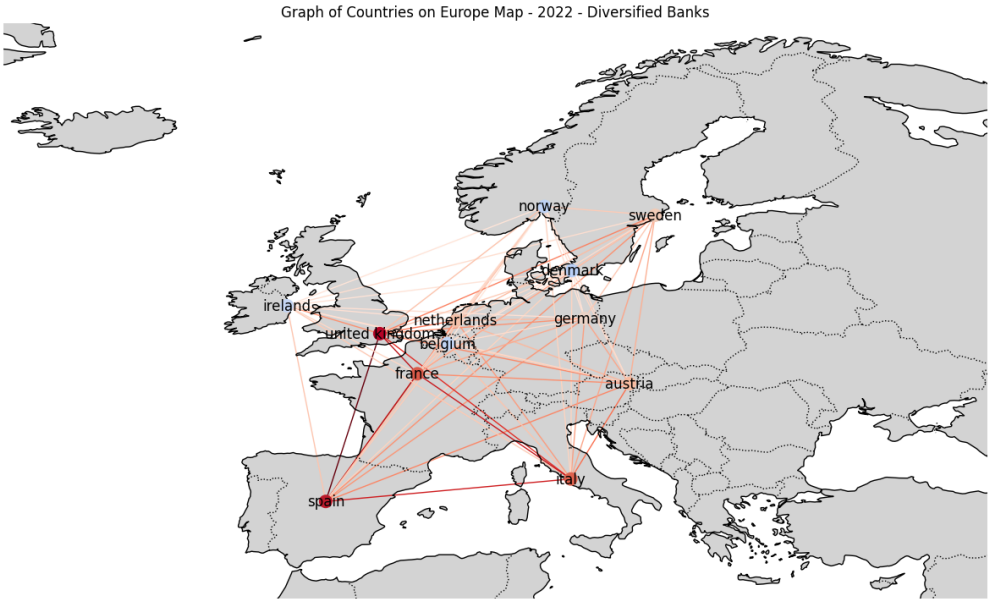

Figure 7: Countries as Nodes - Diversified Banks - 2022

From the previous analysis, we already know that the United Kingdom has a strong economy, especially a leading financial factor. Therefore, we decided to focus the analysis for the industry 'Diversified Banks' on Spain which performed surprisingly well in this sector and - according to our analysis - outperformed all the other countries except the UK. Spain's financial sector's strength in recent years can be attributed to several key factors:

1. **Economic Recovery and Growth**: Following a significant economic downturn and recession in 2009, Spain has experienced a slow yet steady recovery. This recovery has been supported by growth in various industries such as tourism, manufacturing and the service sector. Spain's GDP is forecasted to grow by 3.8% from 2020 to 2025, indicating a resilient and recovering economy [23].

2. **Diverse and Competitive Industries**: Spain has a diverse economy with strong sectors in tourism, manufacturing (notably automotive and pharmaceuticals), services and agriculture. This diversification provides a robust foundation for economic stability and growth [23].

3. **Labor Market Resilience**: Despite challenges, Spain's labor market has shown resilience. The unemployment rate is projected to decrease gradually and there is sustained job creation. This stability in the job market supports overall economic health [24].

4. **Innovative and Open Economy**: Spain has undergone significant changes over the decades, moving away from protectionist policies and opening up its economy to foreign investment. This has been facilitated by political stability and macroeconomic policy adjustments. The Spanish economy's transformation has been marked by increased productivity growth and reduced economic contraction frequency. Importantly, Spain's development in various sectors, such as the automotive industry, was supported by protection and regulation but coupled with enforced competition and openness to foreign technology and investment [25].

5. **Strong Infrastructure and Digital Progress**: Spain has invested in improving and modernising its infrastructure, making it more competitive than some of its Southern European neighbors. The country's digital infrastructure is also notable, offering significant opportunities in the tech sector [23].

6. **Government Policies and Legal Framework**: Post-economic crisis and pandemic, the Spanish government has implemented various policies to encourage investment, including incentives in the private sector and a favorable tax and legal framework [23].

7. **Access to Markets and Trade Agreements**: As a member of the European Union, Spain benefits from access to a market of over 500 million people. Additionally, the country has numerous bilateral agreements with various countries, enhancing its trade options [23].

These factors collectively contribute to the strength of Spain's financial sector, indicating a positive trajectory for future economic development and stability.

5.4.3 Industrial Machinery and Supplies and Components

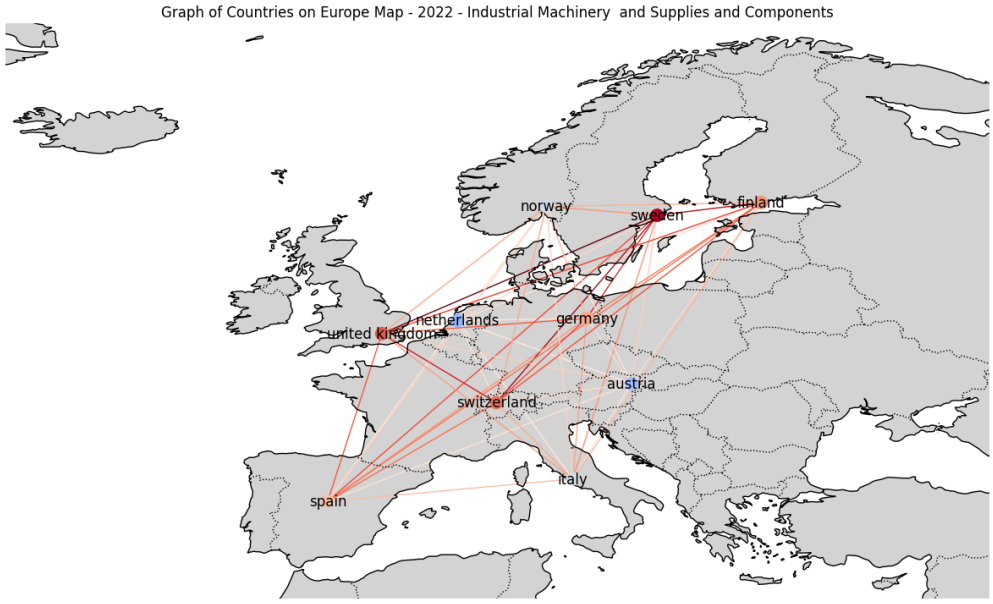

Figure 8: Countries as Nodes - Industrial Machinery and Supplies and Components - 2022

In this chapter, regarding the industry 'Industrial Machinery and Supplies and Components', we want to focus our analysis on Austria. Our investigation showed that Austria's industrial machinery sector experienced a decline in relevance from 2011 to 2022. Reasons for this development could be:

1. **Economic Challenges and COVID-19 Impact**: The Austrian economy, which has generally performed well, encountered a decline in productivity growth and faced challenges in environmental sustainability and skill mismatch despite a strong vocational training system. The COVID-19 pandemic hit Austria particularly hard due to its significant tourism and hospitality sectors, which increased vulnerability to mobility restrictions [26].

2. **Inflation and Competitiveness Issues**: Austria's inflation rate remained persistently high compared to other eurozone economies, impacting the service sector, including tourism, which is crucial for Austria. High inflation led to increased costs in public services and sectors like hotels and restaurants, making Austria less competitive as a tourist destination. Additionally, high inflation amplified wage increases in sectors like accommodation and food services, contributing to the competitiveness challenges [27].

3. **Market Challenges**: The Austrian industrial machinery sector faced various challenges, including the need for adoption of advanced manufacturing technologies to enhance productivity and competitiveness. Most Austrian manufacturers are small- and medium-sized, family-owned companies, necessitating more cutting-edge innovative technologies for their growth and development [28].

4. **Decline in Industrial Output**: There was a noticeable decline in industrial production, with significant contractions in capital goods output. This decline is indicative of challenges within the industrial sector, including machinery and equipment [29].

5. **Job Trends in the Industrial Goods and Machinery Sector**: There was fluctuation in the number of active job postings and new jobs in the industrial goods and machinery sector, indicating volatility and changes in the employment landscape within the industry [30].

6. **Advanced Manufacturing and Innovation Efforts**: Despite the decline, Austria has been a global leader in additive manufacturing companies per capita. The country has focused on modernising important industrial sectors with Industry 4.0 solutions, showcasing its efforts in improving productivity and product quality through intelligent production processes and innovative automation solutions [31].

Overall, Austria's decline in the industrial machinery sector can be attributed to a combination of macroeconomic challenges, inflationary pressures, competitiveness issues, technological advancements and employment trends in the sector.

5.5 Clustering

For the cluster analysis of our networks, organisational affiliations were chosen to define the connections, with companies as nodes and shared organisational links as edges. We assigned weights to these connections using six labels: person, country, city, exchange, organisation and gpe, prioritizing them based on relevance with connections via common persons ranked highest. Each node was further defined by financial and industry attributes. This structuring led to a more interesting network, notably different from the more uniform distribution observed in networks based on countries and cities label. In 2011, the organisational network showed a high level of connectivity, with only a few companies not linked to the rest.

Our clustering analysis employed the Louvain method and MLP kmeans clustering. The Louvain method allowed for immediate application to the network, facilitating swift cluster identification. Conversely, the MLP kmeans required network transformation into a dataframe, using financial data to perform clustering, followed by reintegrating the cluster identifiers as node attributes. For 2011, the optimal number of clusters was identified as two using the elbow method; three clusters were considered but dismissed due to one cluster having too few companies, resulting in a less meaningful division.

5.5.1 K-means

The kmeans clustering, based solely on financial data, divided the companies into two clusters: Cluster 1 with 11-16 companies and Cluster 2 with 81-120 companies. In 2011, Cluster 2 is characterised by significantly lower assets (Figure below), while Cluster 1 has a lower equity-to-asset ratio. Despite differences in size, both clusters exhibit similar performance metrics, with Cluster 1 showing a 10 percent lower gross margin. Notably, ratios like Return on Equity and Return on Assets only slightly differed. This suggests kmeans clustering primarily grouped companies by size, with major firms like Allianz, Volkswagen, UBS and Deutsche Bank in Cluster 1.

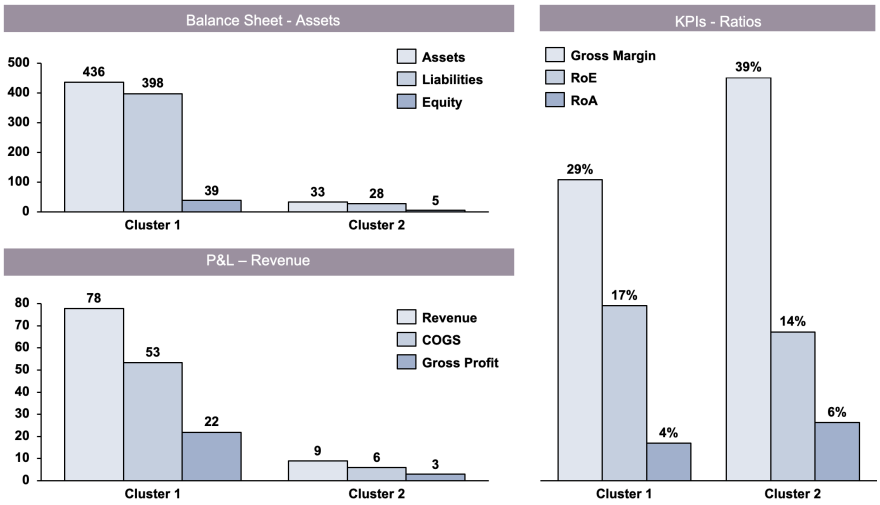

Figure 9: Financial Clusters - K-means - 2011

The network graph below displays the 2011 network with kmeans cluster split based on the node colour with green as cluster 1 and pink as cluster 2. The visualisation was constructed in Gephi with the *Fruchterman Reingold* algorithm.

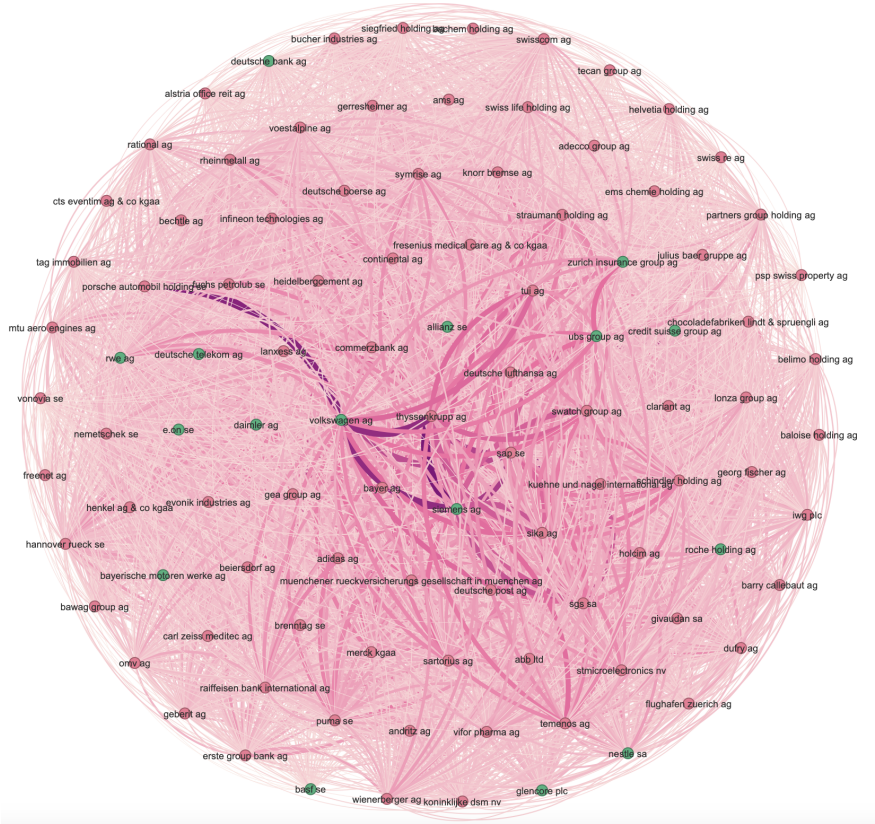

Figure 10: Network K-means for 2011

Findings:

It can be seen that there is a strong connection between volkswage ag, thyssenkrupp ag, siemens ag and porsche automobil holding based on the predefined weights. However, there is no clear pattern between cluster 1 and cluster 2 connections or among them. The kmeans clustering can identify companies based on size when using all financial data as input in the algorithm. It can then be used to classify those companies in the network, to see if there are any connections between the large companies and smaller companies or connections within large companies and within small companies. However, for the data we were given, there were not any meaningful connections between clusters. But, we did find that some companies, especially banks, tend to be more central than others (irrespective of the cluster) since they have many organisational links.

5.5.2 Louvain Method

The Louvain method consistently identified three clusters within the network for each year analysed. For the 2011 network, the clusters are visually distinguishable by colour in Figure 12: Germany is represented by purple, Switzerland by orange and Austria by green, based on the companies' headquarters. The original network composition was 50 percent German, 42 percent Swiss and 8 percent Austrian companies. Cluster analysis revealed a strong geographical correlation: the first cluster consisted of 93 percent Swiss companies, the second was entirely German and the third was exclusively Austrian. This indicates that the Louvain method effectively grouped companies into clusters based on geographic location, influenced by shared organisational attributes. For instance, in the Swiss cluster, there are German and Austrian companies that have significant ties to Swiss organisations.

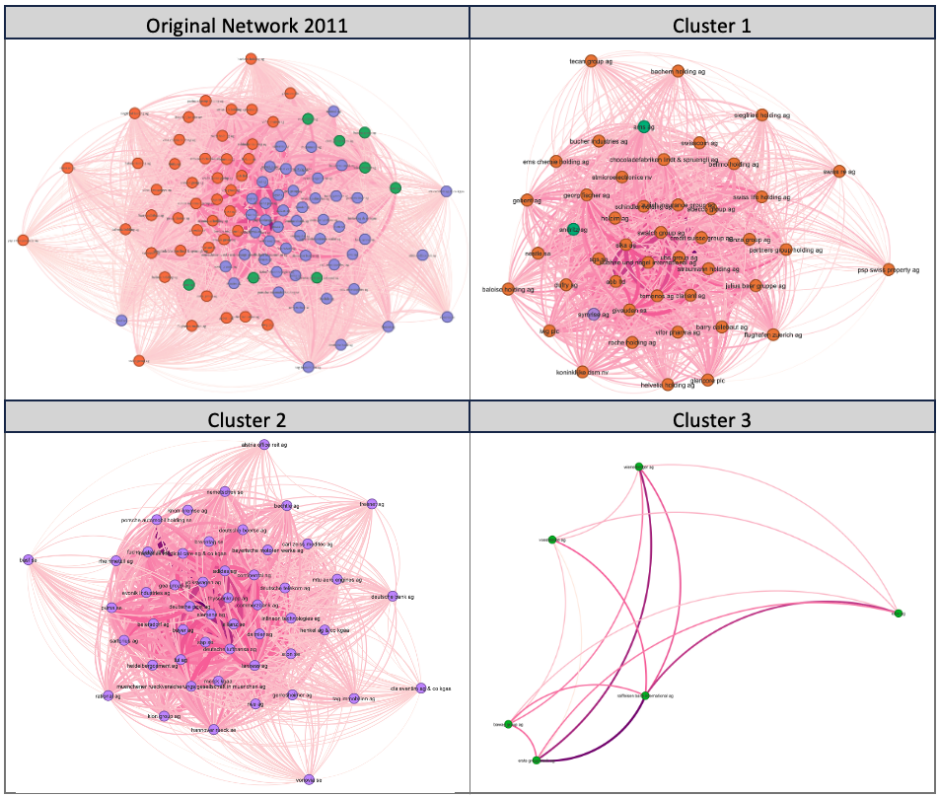

Figure 11: Network Clustered for 2011

Findings:

This geographical clustering provides practical insights, particularly for risk analysis. For example, if Swiss companies face a specific risk, German and Austrian companies in the same cluster may be similarly at risk due to their organisational connections. Further investigation into these common organisations could reveal the nature of these connections and inform a more nuanced risk assessment strategy. This can be particularly valuable for understanding cross-border economic exposures and developing robust risk mitigation plans.

Furthermore, the Louvain algorithm excels in handling large datasets, facilitating efficient clustering even with substantial volumes of data. This allows BDO to import and cluster extensive company datasets effectively. The algorithm can then locate a target company within these clusters. This process enables deeper analysis of the specific cluster, uncovering the rationale for the community's formation and the criteria used for company categorisation.

6 Link Prediction

- **Jaccard Coefficient**: This algorithm measures the similarity between the sets of neighbors of two nodes. A high Jaccard score means the nodes share many neighbors, suggesting a strong likelihood of a connection.
- **Adamic Adar Index**: This algorithm also considers common neighbors but weighs less-connected neighbors more heavily. The idea is that sharing a less-common neighbor is more significant than sharing a highly-connected one.
- **Preferential Attachment**: This algorithm is based on the notion that the more connections a node has, the more likely it is to form new connections. Higher scores indicate a greater likelihood of a new link.
- **Common Neighbor**: Similar to the Jaccard Coefficient, this method counts the number of common neighbors between two nodes. Higher scores indicate a greater likelihood of a link.
- **Katz Score**: Our own implementation of the Katz Index considers not just direct neighbors but also the paths of length up to 3 between nodes, with a damping factor reducing the influence of longer paths.

In summary, each algorithm sheds light on different aspects of potential connections in the graph. The Jaccard Coefficient and Common Neighbor methods highlight nodes with shared immediate environments, the Adamic Adar Index brings in the aspect of unique shared connections, Preferential Attachment points to the influence of highly connected nodes and the Katz Score explores more distant or indirect connections [32].

Conclusion: Challenges in Fully Connected Networks

Despite the effectiveness of these link prediction algorithms in many scenarios, their applicability becomes limited in the context of fully connected networks and subnetworks. In a fully connected network, every node is directly connected to every other node. This characteristic fundamentally alters the underlying assumptions on which these algorithms are based.

Loss of Neighborhood Diversity: Algorithms like the Jaccard Coefficient and the Common Neighbor method lose their effectiveness in fully connected networks. Since every node is connected to every other node, the concept of 'neighborhood' becomes uniform across the network, negating the ability to discern meaningful similarities or differences between node connections.

Reduced Significance of Unique Connections: The Adamic Adar Index thrives on the uniqueness of shared connections. In fully connected networks, the uniqueness of connections is nullified, rendering this algorithm ineffective for predicting meaningful links.

Irrelevance of Preferential Attachment: The core idea of Preferential Attachment is that nodes with more connections are more likely to form new connections. However, in a fully connected network, where each node already connects to every other node, the concept of forming 'new' connections becomes irrelevant.

Diluted Impact of Path Lengths: The Katz Score, which takes into account the paths of varying lengths between nodes, also faces challenges. In a fully connected network, the direct path between any two nodes makes longer paths less relevant, diminishing the Katz Score's ability to provide additional insights.

In conclusion, while these link prediction algorithms are powerful tools for understanding and predicting connections in many types of networks, their utility diminishes in the context of fully connected networks. The intrinsic properties of such networks - uniformity of connections and the absence of unique or indirect paths - lead to a scenario where traditional link prediction methodologies offer limited additional insights into the network's structure or the likelihood of future connections [32]. Due to the fact that our networks and subnetworks are all fully connected, link prediction was not possible for all the reasons mentioned before. Nonetheless, we fully implemented the listed link prediction algorithms and included a section in the associated notebook where we demonstrate them with filtered versions of our networks where we set individual cutoff points for the edge weights to arrive at networks with some missing edges. Moreover, we, as an alternative approach, tried to predict the future change in edge weight instead, however, unsuccessfully so since our ML models never yielded satisfactory results. Furthermore, we even went as far as implementing a Graph Convolutional Network (GCN) ML model using the Deep Graph Library with PyTorch as a backend to additionally exploit the graph structure of the networks but again with no satisfactory results.

7 Conclusion

Simon Böck:

Starting this project, I was stepping into something completely new because I had never worked with networks in any sense before. Facing challenges became part of the project that taught me not just about networks, but also about pushing through tough spots. Getting to grips with network analysis felt like piecing together a big puzzle. Bit by bit, everything started to click, transforming abstract ideas into real, working solutions. This process was exciting. The project really put my and the team's problem-solving skills to the test, encouraging us to think out of the box and come up with creative solutions when things did not go as planned. In short, this project was more than just learning about network analysis. It was about growing, facing challenges and experiencing the joy of turning an idea into reality. This whole experience has given me valuable skills and insights that I can use in the future.

Manuel Petschinger:

Like Simon, network analysis was a completely new topic for me. It therefore took a very long start-up phase at the beginning to get to grips with the project. Due to the unclassified attributes, we tried for a very long time to make the network clearer or to subdivide it, but without success. Accordingly, I have learnt from this to think of methods to overcome such major obstacles at an early stage instead of searching in vain for solutions to make the best of the available resources. Even if you often do not realize that systems like Spacy are available for free, it's definitely worth looking into. I also learnt a lot about the technical component. My programming and data science skills have improved considerably, I have learnt a lot about networks and how to analyse them, I have become much more creative in developing solutions. All in all, I was able to add a bunch of knowledge and new methods to my repertoire.

Jakob Koller:

Starting network analysis was a new and challenging task for me, similar to Simon's and Manuel's experience. In the beginning, I struggled with the complexity of the project, especially with clustering due to my limited coding skills. Working with clustering algorithms was particularly tough, as it was a complex task that pushed my problem-solving skills and technical knowledge to their limits. I also encountered difficulties in analysing the clusters because I was inexperienced with such analyses. Despite these issues, I learned a lot from the experience. It encouraged me to discover and use tools like Gephi, which were powerful and available but not widely used. This project improved my programming and data science skills significantly, gave me a solid grounding in network construction and analysis and enhanced my ability to think creatively to solve problems. It was a tough but ultimately rewarding process that added new techniques and insights to my skill set.

Markus Weiss:

In contrast to my other team members, Network Analysis was not a completely new topic to me, which is why I started with the initial processing of network files and their analysis and quickly tried to gain as much insight as possible and to share what I knew along the way. This phase was, however, not just about understanding the data, but about seeing the bigger picture it formed and realising that we faced many challenges with the data at hand. Thus, together with new data, we later shifted our focus the D-A-CH region, which added needed clarity to our project and allowed us to proceed. The following Link Prediction aspect was also intriguing, despite the lack of proper results even with various alternative approaches due to the fully-connected nature of our networks. It was a challenging yet rewarding journey, pushing me to adapt and think critically. This project was a profound learning experience through dealing with the various challenges that arose in our Data Science project and an advancement of my skill set in Network Analysis and programming in general.

8 Team

Manuel Petschinger - IT-Specialist in Charge of Technology (Data Processing, NLP and Countries as Nodes) - Oversees the overall project timeline, ensures that milestones are met and coordinates team meetings. Manuel is also responsible for resource allocation and ensuring that all team members have the tools and information they need. Furthermore, Manuel's tasks are data processing, developing NLP-algorithms to label the data with high accuracy and building the 'countries as nodes' networks for insights into relations between countries.

Jakob T. Koller - Specialist in Cluster Analysis and Data Insights - Jakob leads the focus on exploratory data analysis with a particular emphasis on cluster and subnetwork analysis. His primary objective is to simplify complex datasets, extracting meaningful insights. He applies advanced techniques to segment data into clusters, enabling a deeper understanding of inherent patterns and relationships. This approach facilitates the discovery of valuable insights, helping to inform strategic decisions and identify key trends within the data with respect to risk management and corporate finance.

Markus C. Weiss - IT-Specialist in charge of Technology (Data Science, General Network construction, Analysis and Link Prediction) - Focused on laying the code foundation and providing coding assistance to all managers, creating the initial general data analysis workflow. Markus worked closely with the data analysis manager to ensure that visuals accurately represent the data insights and are presented in a manner that is easy to understand. Ultimately, Markus focused on the general network construction, analysis and the link prediction.

Simon Böck - IT-Specialist in charge of Technology (Exploratory Data Analysis) - As the IT Specialist in charge of Technology with a focus on Exploratory Data Analysis in our project, Simon's role was key in delving deep into the data to uncover patterns, possible trends and insights. Simon was responsible for the analysis of networks, particularly those structured around countries and industries as nodes. Collaborating closely with Manuel Petschinger was a crucial aspect, given his expertise in pre-processing the data. Our teamwork ensured a seamless integration of the pre-processing efforts with the exploratory data analysis phase, allowing us to efficiently bridge the gap between raw data and actionable insights.

9 References

[1] BDO Austria GmbH. About bdo. https://www.bdo.at/de-at/uber-bdo/gute-grunde-fur-bdo/zahlen-und-fakten, 2023. Retrieved from BDO Austria GmbH.

[2] Irene Aldridge and Marco Avellaneda. *Big data science in finance*. John Wiley & Sons, 2021.

[3] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of big data*, 2(1):1–21, 2015.

[4] Rudolf Volkart and Alexander F Wagner. *Corporate Finance: Grundlagen von Finanzierung und Investition*. Versus Verlag, 2018.

[5] Andrew Hiles. Enterprise risk management. *The definitive handbook of business continuity management*, pages 1–21, 2012.

[6] Hui Chu. An empirical analysis of corporate financial management risk prediction based on associative memory neural network. *Computational Intelligence and Neuroscience*, 2021:Article ID 4383742, 2021.

[7] Ming-Fu Hsu, Ying-Shao Hsin, and Fu-Jiing Shiue. Business analytics for corporate risk management and performance improvement. *Annals of Operations Research*, 315:629–669, 2022.

[8] Fabio Caccioli, Paolo Barucca, and Teruyoshi Kobayashi. Network models of financial systemic risk: a review. *Journal of Computational Social Science*, 1:81–114, 2018.

[9] Daning Hu, Gerhard Schwabe, and Xiao Li. Systemic risk management and investment analysis with financial network analytics: research opportunities and challenges. *Financial Innovation*, 1(2), 2015.

[10] Explosion AI. spacy: Industrial-strength natural language processing, 2024. Accessed: 2024-01-15.

[11] Met Office. *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon, 2010 - 2015.

[12] K Stadler. The country converter coco - a python package for converting country names between different classification schemes. *The Journal of Open Source Software*, 2017.

[13] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[14] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[15] Cássia Sampaio. K-means clustering with the elbow method. *Stack Abuse*, 2023.

[16] Kinga Edwards. The complete guide to understanding the dach market. *ecommercegermany.com*, 2021. https://ecommercegermany.com/blog/the-complete-guide-to-understanding-the-dach-market.

[17] MA Community Portal. Dach ma trends report for the first half of 2021. *mnacommunity.com*, 2021. https://mnacommunity.com/insights/dach-ma-trends-report-h1-2021/.

[18] MA Community Portal. Dach region ma trends: Full year report for 2021. *mnacommunity.com*, 2021. https://mnacommunity.com/insights/dach-ma-trends-report-fy-2021/.

[19] Uk economy set to escape 2024 'hard landing' in boost for sunak — evening standard. https://www.standard.co.uk/business/uk-economy-set-to-escape-2024-hard-landing-in-boost-for-sunak-b1129294.html, 2023. Accessed: 2024-01-31.

[20] Will the uk economy keep up with the rest of europe in 2024? https://www.goldmansachs.com/intelligence/pages/will-the-uk-economy-keep-up-with-the-rest-of-europe-in-2024.html. Accessed: 2024-01-31.

[21] William Schomberg. Uk economy picks up speed but red sea crisis hits factories — reuters. https://www.reuters.com/markets/europe/uk-services-firms-grow-faster-factories-feel-red-sea-hit-pmi-2024-01-24/, 2024. Accessed: 2024-01-31.

[22] European economics analyst: Uk outlook 2024: Not so different after all. https://www.goldmansachs.com/intelligence/pages/uk-outlook-2024-not-so-different-after-all.html, 2023. Accessed: 2024-01-31.

[23] Why you should invest in spain in 2022 — ins global. https://ins-globalconsulting.com/news-post/invest-spain-2022/. Accessed: 2024-01-31.

[24] Economic forecast for spain - european commission. https://economy-finance.ec.europa.eu/economic-surveillance-eu-economies/spain/economic-forecast-spain_en. Accessed: 2024-01-31.

[25] Oscar Calvo-Gonzalez. Unexpected lessons from spain's economic rise — europp. https://blogs.lse.ac.uk/europpblog/2021/10/22/unexpected-lessons-from-spains-economic-rise/, 2021. Accessed: 2024-01-31.

[26] Key policy insights — oecd economic surveys: Austria 2021. https://www.oecd-ilibrary.org/sites/62cf975b-en/index.html?itemId=/content/component/62cf975b-en. Accessed: 2024-01-31.

[27] Sticky inflation hurts austria's competitiveness — article — ing think. https://think.ing.com/articles/sticky-inflation-hurts-austrias-competitiveness, 2023. Accessed: 2024-01-31.

[28] Jan Bruckner. Austria - advanced manufacturing. https://www.trade.gov/country-commercial-guides/austria-advanced-manufacturing, 2022. Accessed: 2024-01-31.

[29] Austria industrial production march 2023 - focuseconomics. https://www.focus-economics.com/countries/austria/news/industrial-production/industrial-output-falls-at-sharpest-rate-since-july-2022-in-march/, 2023. Accessed: 2024-01-31.

[30] Austria: Job trends in the industrial goods and machinery sector (march 2023 - june 2023) - globaldata. https://www.globaldata.com/data-insights/industrial-goods-and-machinery/austria-job-trends-in-the-industrial-goods-and-machinery-sector-2095491/. Accessed: 2024-01-31.

[31] Advanced manufacturing — invest in austria. https://investinaustria.at/en/industries-functions/industry/advanced-manufacturing/. Accessed: 2024-01-31.

[32] Song C. Ge Y. et al. Wu, H. Link prediction on complex networks: An experimental survey. *Data Sci. Eng. 7, 253–278 (2022)*, 2022.

10 Appendices

https://github.com/CoulombTunnel/DS_Lab_BDO