Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Master project of the Design Thinking and Innovation Research Chair
in cooperation with the Artificial Intelligence and Intelligent Systems Chair

Uncovering Global Supply Chains by the Integration of Company Data into a Graph Database Analyzing Multi-Tiers.

Information integration, prediction, and analysis of supplier relationships in the context of sustainability and resilience, based on publicly available data.

Motivation

Companies face the issue of delivering products and services in times of highly interconnected and complex value chains. This comes with many dependencies and risks. Decision makers like Apple know with which companies they directly collaborate (so called Tier 1 suppliers) but are unable to retrace the full spectrum of the value chain. For this, there currently isn't a publicly or privately available complete dataset in research nor in industry. To increase efforts in sustainability and resilience, such a dataset yields very high impact. Many companies already publish their direct Supplier Lists (see Figure 1) based on own sustainability efforts and regulatory pressure. Those who currently do not, can be surveyed, get predicted, or derived from other data.

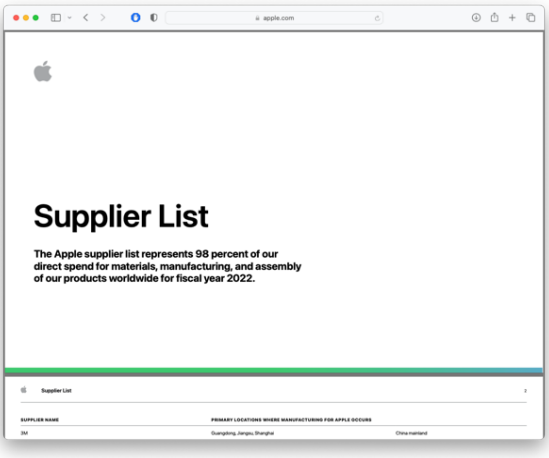

Figure 1: Supplier List of Apple

Connecting suppliers over multiple levels is not yet tackled thoroughly. Building such a dataset, enables deriving risks from supplier networks. Which value
chain paths are more, which less risky? For this every supplier in the network must receive a certain risk score based on its specific data. A sophisticated algorithm could give optimized advice to choosing improved paths in this supplier network.

Background

This project is based on a sustainability challenge originally coming from a corporate partner of HPI. The HPI students of this project discovered that most sustainability analyses only take a very local view aspect but do not grasp the full picture. Leveraging the global view on holistic value chains is one of the most pressing topics in sustainability. Most emissions of a company are in

the Scope 2 and Scope 3 which only really can be grasped if a full value chain retracing is possible. Full global dependencies of companies can only get understood if a network dataset is produced.

Goal

Analyzing publicly available data is a challenge itself. Here many different information integration tasks are to be tackled. Crawling and saving the data, harmonizing, and integrating it. Also matching differently spelled company names, e.g.: based on similarity metrics or with large language models remains a well-known challenge in research and industry. After this comes the task of building up the supplier network with state-of-the-art Graph Database technology and research. Deriving sustainability and resilience path optimizations from this graph-network can be seen as a bonus task.

Possible concrete questions and technologies could include:

- Data **Crawling** and Data **Collection**
- **Information integration**, like cleansing and harmonizing this data with **big data** technologies, **large language models** or other matching techniques
- Building the **Graph Database** network of this supplier data
- If no concrete naming of direct suppliers for a certain company can be found, use **machine learning** to **predict its Tier 1** suppliers, based on similar companies and assign confidence scores to those predictions
- **Improving** current supply chain paths by algorithmic suggestions, with **neural networks** or other state of the art machine learning technologies
- … and more!

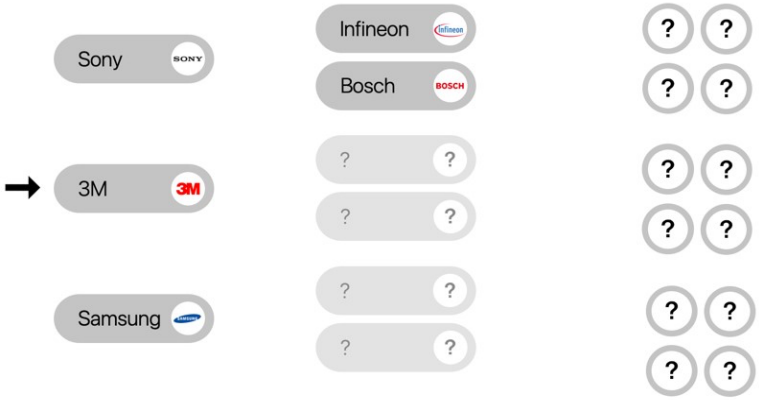

Figure 2: Graph Network of Tiers of suppliers in a value chain

About You

- You are interested in working in a multidisciplinary team connecting research and industry tackling sustainability challenges
- Working with graph databases and publicly available data sounds fun to you
- You have sufficient knowledge in information integration and databases
- You have experience in training and deploying machine learning models
- You are studying Data Engineering or IT Systems

Contacts

Do you have any questions? Please send us an e-mail.
We are looking forward to hearing from you!

Prof. Dr. Falk Übernickel
E-Mail: falk.uebernickel@hpi.de

Tolga Buz
E-Mail: Tolga.Buz@hpi.de