



POLITECNICO DI TORINO
DEPARTMENT OF CONTROL AND COMPUTER ENGINEERING (DAUIN)

Master Degree in Data Science and Engineering

Master Degree Thesis

**A Knowledge Graph Approach to
Critical Raw Materials Risk
Monitoring: Developing a Strategic
Digital Twin with Neo4j**

Author: Muhammed Emin ORAL

Advisor: Prof. Alessandro SAVINO

Co-Advisor: Andrea BERNARDINI, Nicolò MAUNERO, Francesco GIANCATERINI

July, 2026

Abstract

European industrial economies face a structural vulnerability in their dependence on external suppliers of critical raw materials (CRMs). These minerals are indispensable to the green and digital transition, advanced manufacturing, and defense industries. The European Union has responded by enacting the Critical Raw Materials Act (CRMA), which requires companies to trace their supply chains back to the point of origin and to assess disruption risks at every stage. However, existing monitoring tools rely on aggregate country-level indicators that treat entire nations as single supply nodes. These tools are unable to capture firm-level dependencies or the indirect, multi-hop paths through which localized disruptions actually propagate across trade networks.

To address this gap, this thesis draws on two technical foundations: the regulatory architecture of the EU CRMA and the computational properties of graph database systems. It establishes why traditional relational databases are structurally inadequate for tracing multi-tier supply dependencies, and why a Labeled Property Graph engine — specifically Neo4j — provides the efficiency required for topology-aware risk analysis. Core graph algorithms, including weighted Betweenness Centrality and PageRank, are formalized as instruments for identifying brokerage bottlenecks and systemic vulnerabilities within trade networks.

The primary contribution of this thesis is a *Knowledge-Graph-based Strategic Digital Twin* for CRM supply chain risk monitoring. A synthetic firm generation procedure is developed to reconstruct market structures from publicly available UN Comtrade trade flow data, calibrated against the observed Herfindahl-Hirschman Index. The resulting entities are integrated into an ontology-driven Knowledge Graph in Neo4j. A novel network metric, *Nodal Elasticity*, is introduced to quantify the short-run sensitivity of a downstream economy to localized supply shocks under fixed trade dependencies. The thesis also proposes a generalized multi-tier conceptual framework that bridges corporate Bill of Materials structures with macroeconomic risk indicators through a Composite Structural Fragility Index, providing a regulatory-aligned blueprint for future implementation.

The framework is validated on the global tungsten supply chain as an empirical case study, with the Italian industrial system as the target economy. Stress-testing simulations reveal that the firms generating the largest downstream impact on Italy are not the dominant Chinese producers, but East and Southeast Asian intermediaries that supply Italy in negligible direct volumes yet sit at structurally pivotal positions within regional trade ecosystems. The most critical intermediaries produce supply closure losses approaching 44% under high propagation regimes — a result that remains entirely invisible to country-level concentration metrics. Furthermore, the analysis shows that structural control within

the network is driven by brokerage position rather than trade volume: German and U.S. intermediaries dominate the critical routing paths despite China holding the largest share of aggregate production.

Contents

List of Tables	6
List of Figures	7
1 Introduction	9
2 Background	13
2.1 The EU CRMA Regulation and Macroeconomic Risk Indicators	13
2.2 Graph Databases and Knowledge Graph Architecture: Why Property Graphs?	15
2.3 Graph Data Science and Fundamentals of Network Analysis	16
2.4 Software Infrastructure and Data Sources	18
3 Related Work	21
3.1 Stress tests	21
3.2 Economic synthetic data generation	22
3.3 Knowledge Graphs and Digital Twins for Supply-Chain Risk	23
3.4 Research Gap: Macro-Criticality Frameworks vs. Micro-Corporate Manage- ment	23
3.5 Systematic Literature Review Methodology via LLM-Powered Pipelines . .	24
4 Methodology and Contributions	29
4.1 Design Choices and Contributions	29
4.2 A Granular Downscaling Methodology for CRMs	33
4.3 Knowledge Modeling Framework	37
4.4 Nodal Elasticity	40
4.5 Proposed Multi-Tier & Multi-Commodity Conceptual Framework	42
5 Case Study: Tungsten	47
5.1 Tungsten: Industrial Applications and Strategic Criticality	47
5.2 The Tungsten Value Chain and Commodity Classification	48
5.3 Market Structure and Geopolitical Concentration	51
5.4 The Global Tungsten Trade Ecosystem	51
6 Quantitative Analysis of the Tungsten Supply Network	55
6.1 Network Statistics of the Reconstructed Tungsten Graph	55
6.2 Structural Centrality Analysis	61

6.2.1	Computing Network Metrics with the Graph Data Science Library .	62
6.2.2	Weighted Betweenness Centrality	63
6.2.3	PageRank Centrality	65
6.2.4	Joint Analysis of Betweenness and PageRank	66
6.3	Geopolitical Risk Exposure	67
6.4	Cross-Metric Synthesis and Country-Level Aggregation	70
7	Trader-Level Elasticity and Stress-Testing Results	73
7.1	The Direct Exposure Profile of the Italian Economy	74
7.2	Firm-Level Stress Testing of Chinese Suppliers	75
7.3	Global Elasticity Ranking and Indirect Amplifiers	76
7.4	Nodal Elasticity versus Structural Centrality	78
8	Conclusions, Limitations and Future Work	81
8.1	Conclusions	81
8.2	Limitations	82
8.3	Policy Implications and Replicability	82
8.4	Future Work	83
	Bibliography	85

List of Tables

4.1	<i>Core entities and associated attributes of the ontology-driven Strategic Digital Twin</i>	38
5.1	<i>Mapping between tungsten value-chain stages, representative HS commodity codes, and their inclusion in the empirical trade-network reconstruction. The three codes retained for analysis span the mid-to-downstream metallic processing stages, where economic value density and strategic content are highest.</i>	50
6.1	<i>Baseline composition of the reconstructed tungsten Strategic Digital Twin graph.</i>	56
6.2	<i>Distribution of synthetic Trader nodes across the ten most represented countries in the tungsten supply network.</i>	57
6.3	<i>Structural connectivity metrics of the trader-level subgraph.</i>	59
6.4	<i>Country-level aggregation of trader centrality. Aggregate structural control (betweenness) does not track the number of national traders: Germany leads with only three traders, while China, with the largest trader count and dominant physical production, ranks third.</i>	72
7.1	<i>Top ten traders by nodal elasticity of Italian imports, with elasticity and closure-loss measures under the high ($\lambda = 0.9$) and intermediate ($\lambda = 0.7$) propagation regimes.</i>	76

List of Figures

3.1	<i>Full architecture of the semi-automated Systematic Literature Review pipeline, from automated Google Scholar scraping to LLM-based classification and curated output generation.</i>	25
4.1	<i>Synthetic firm-share reconstruction for China under the Pareto-calibrated downscaling procedure. The dark bars denote the real-world anchor firms, whose approximate market shares are retained from observed data, while the lighter bars denote the synthetic traders generated from the calibrated Pareto distribution to populate the long tail of the national firm-size structure.</i>	35
4.2	<i>Ontology-driven conceptual model of the Knowledge-Graph-based Strategic Digital Twin</i>	37
4.3	<i>Database schema of the Strategic Digital Twin, rendered by the Neo4j <code>db.schema.visualization()</code> procedure. The diagram shows the five node labels (Trader, Country, CriticalRawMaterial, ConcentrationRisk, and GeopoliticalRisk) and the four directed relationship types (<code>SUPPLIES_TO</code>, <code>LOCATED_IN</code>, <code>TRADES</code>, and <code>HAS_RISK</code>) constituting the property-graph ontology.</i>	39
4.4	<i>Main operational stages and regulatory benchmarks of the strategic raw material value chain under the EU CRMA framework. Percentage thresholds ($\geq 10\%$, $\geq 40\%$, $\geq 25\%$) refer to the domestic capacity targets established by Regulation (EU) 2024/1252, Article 5 [2].</i>	43
5.1	<i>Main stages of the tungsten value chain, from ore extraction to downstream finished products, and the associated HS classifications considered in the analysis. The dashed re-entry arrow indicates the reintegration of recycled scrap into the reduction and consolidation stages.</i>	49
6.1	<i>Sample subgraph of the reconstructed tungsten network, rendered in the Neo4j Desktop. Yellow nodes are synthetic Traders and orange nodes are Countries, linked by <code>LOCATED_IN</code> edges. Most countries host a single trader (isolated pairs), while the dense star-shaped cluster represents a single highly concentrated country populated with multiple synthetic traders.</i>	58
6.2	<i>Degree rank-size distribution of the trader network on doubly logarithmic axes. The heavy upper tail is consistent with a power-law-like firm-size structure induced by the Pareto-calibrated downscaling procedure.</i>	60
6.3	<i>Reconstructed tungsten supply-chain network with synthetic trader-level entities.</i>	61

6.4	<i>Top twenty traders by weighted Betweenness centrality. Only 42 of the 192 traders record a non-zero score, and the five leading brokers account for roughly 68% of total betweenness.</i>	64
6.5	<i>Top twenty traders by PageRank centrality. The distribution is less concentrated than betweenness, reflecting the broader diffusion of recursive systemic importance across the dense trading core.</i>	65
6.6	<i>Joint distribution of weighted Betweenness and PageRank. Dashed lines mark the median of each metric. Upper-right: structural super-brokers (high on both). Lower-right / pure brokers: high betweenness, moderate PageRank (e.g. Xiamen Tungsten). Upper-left: recursively central but zero-brokerage traders (e.g. <code>MEX_Trader_02</code>, <code>USA_Trader_08</code>).</i>	67
6.7	<i>Geopolitical Risk Exposure decomposed into its structural (normalized PageRank) and geopolitical (country risk) components, with point color encoding the composite GRI. Risk-driven nodes (e.g. <code>COD_Trader_01</code>, <code>RUS_Trader_01</code>) occupy the upper-left; the structurally dominant but geopolitically stable <code>USA_Trader_01</code> occupies the lower-right.</i>	69
6.8	<i>Top fifteen traders by Geopolitical Risk Exposure Index under the equal-weight ($w = 0.5$) configuration.</i>	70
6.9	<i>Spearman rank-correlation matrix of the centrality metrics and total degree. The moderate betweenness–PageRank correlation and the strong PageRank–degree correlation confirm that the metrics capture distinct structural dimensions.</i>	71
7.1	<i>Composition of Italy’s direct (Tier-1) tungsten imports by supplier country. European suppliers dominate the direct import structure, while China (highlighted) accounts for only about 12.8% of direct imports.</i>	74
7.2	<i>Nodal elasticity of Italian imports to disruptions at the six largest Chinese traders, as a function of the propagation parameter λ. Elasticity rises sharply and convexly with λ, indicating that the systemic importance of these suppliers is driven predominantly by indirect, multi-hop propagation rather than direct bilateral exposure.</i>	75
7.3	<i>Top fifteen traders by nodal elasticity at $\lambda = 0.9$, colored by region. The highest-elasticity nodes are East and Southeast Asian intermediaries rather than the dominant Chinese producers.</i>	77
7.4	<i>Nodal elasticity ($\lambda = 0.9$) versus weighted Betweenness centrality. The highest-elasticity amplifiers (<code>MYS_Trader_02</code>, <code>BRN_Trader_01</code>) occupy the upper-left region — maximal propagation impact with negligible or zero brokerage — making them entirely invisible to structural metrics.</i>	78

Chapter 1

Introduction

The growing geopolitical fragmentation and the repeated disruptions in global commodity markets have intensified strategic concerns about the structural resilience of industrial supply chains. The European Union, in particular, remains heavily dependent on external suppliers for strategic minerals and metals essential to the green and digital transition, defense industries, and advanced manufacturing. Recent disruptions, including the European energy crisis and the 2026 Strait of Hormuz conflict, have demonstrated how geographically localized shocks can rapidly cascade through interconnected trade networks, generating systemic failure far beyond directly exposed economic actors [1].

Precisely because of this critical dependence, the European Union has progressively strengthened its monitoring and governance framework. This process culminated in the adoption of the Critical Raw Materials Act (CRMA) [2], which establishes ambitious 2030 targets for domestic extraction, processing, recycling, and import diversification. Beyond baseline capacity targets, the CRMA explicitly emphasizes the need to identify the exact origin of critical raw materials, monitor sub-tier supply-chain dependencies, evaluate structural concentration risks, and develop advanced stress-testing mechanisms.

However, a profound operational gap exists between regulatory goals and current technical capabilities. Existing monitoring frameworks rely predominantly on aggregate country-level concentration indicators, import dependency ratios, and macroeconomic trade statistics. Although informative at the macro level, these measures provide no visibility into firm-level dependencies, downstream intermediary actors, and multi-hop transmission channels through which localized disruptions spread across industrial networks [3]. Furthermore, standard industrial management practices operate in complete isolation from these macroeconomic metrics. While manufacturing enterprises manage internal production using a localized Bill of Materials (BOM) mapping direct Tier-1 procurement, they possess no structural visibility over the raw chemical or mineral constituents embedded within purchased components, let alone their primary extraction and refinement trajectories.

To address this structural fragmentation, this thesis shifts the analytical focus from aggregate macroeconomic statistics to an integrated knowledge engineering approach. This research develops an end-to-end data science methodology to reconstruct, model, and simulate firm-level supply networks under conditions of limited data transparency. By bridging macro-level trade aggregates with calibrated synthetic micro-level structures, this work

establishes a scalable framework capable of translating policy-driven indicators into actionable operational intelligence.

The core implementation of this research is a Knowledge-Graph-based Strategic Digital Twin (SDT) developed within a Neo4j Property Graph database environment. The structural requirement drives the choice of a graph-native architecture to capture multi-hop dependencies and indirect risk propagation paths that traditional relational databases cannot trace efficiently. Within this graph environment, we introduce *Nodal Elasticity*, designed to quantify the short-run sensitivity of downstream manufacturing economies — with particular focus on the Italian industrial system — to targeted supplier disruptions under rigid trade dependencies.

To validate the scalability and empirical fidelity of the proposed knowledge modelling architecture, the global tungsten trade network is deployed as an empirical pilot case study. Tungsten serves as an ideal material for validation due to its extreme upstream production concentration, heavy dominance by vertically integrated upstream groups, and strategic criticality across advanced manufacturing sectors. By simulating targeted shocks within the reconstructed tungsten network, this thesis demonstrates how firm-level topology uncovers systemic vulnerabilities and indirect risk-amplifying intermediaries that remain entirely hidden within aggregate country-level indicators.¹

This thesis contributes to the literature in five main dimensions:

1. **A Constraint-Based Synthetic Firm Generation Pipeline for Trade Data Reconstruction:** A programmatic downscaling workflow combining aggregate UN Comtrade statistics with constrained Pareto-based synthetic firm generation to reconstruct plausible oligopolistic market structures under acute data scarcity.
2. **An Ontology-Driven Knowledge Graph Architecture:** The structural engineering of a Strategic Digital Twin in Neo4j, seamlessly integrating countries, synthetic traders, trade dependencies, and multi-dimensional risk scores into a semantic property graph environment.
3. **The Formalization of Nodal Elasticity:** A network-native structural metric capable of capturing heterogeneous shock propagation and higher-order diffusion effects across relatively rigid, short-run procurement relationships.
4. **An Empirical Micro-Level Assessment of Tungsten Risk:** A comprehensive network case study demonstrating how firm-level concentration structures in China act as systemic risk carriers, identifying politically stable intermediary hubs as hidden risk amplifiers for the European network.
5. **A Generalized Regulatory-Aligned Conceptual Model:** A theoretical multi-layer graph framework spanning five functional layers (mining, refining, component

¹The results of this thesis are part of an article currently under consideration for an international conference, co-authored with Francesco Giancaterini, Tiziano Bacaloni, Alessio Abeltino, and Andrea Bernardini of the Ugo Bordoni Foundation (FUB), titled *Firm-Level Stress Testing of the Tungsten Supply Chain: A Nodal Elasticity Analysis*. Although improved modeling of the tungsten network was among the shared objectives, the novel contribution of this thesis with respect to that work lies in the generalized multi-tier and multi-commodity conceptual framework introduced in Section 4.5.

manufacturing, OEM assembly, and recycling) that mathematically integrates corporate BOM allocations via a Composite Structural Fragility Index to support broad multi-commodity tracking.

The structural and analytical investigation is guided by the following core research questions:

- **RQ1:** *How can firm-level concentration structures be reconstructed from aggregate raw material trade data, and how do they differ from country-level indicators such as the Herfindahl-Hirschman Index (HHI)?*
- **RQ2:** *What network-level vulnerabilities and resilience patterns emerge from a trader-level graph topology that cannot be identified using aggregate macroeconomic trade statistics?*
- **RQ3:** *How do localized supply shocks affecting dominant upstream traders propagate through a graph-native supply network under alternative short-run propagation regimes?*
- **RQ4:** *How can a graph-native conceptual framework effectively bridge micro-level corporate BOM properties with macroeconomic supply risk factors to fulfil the comprehensive supply chain tracking mandates established by the EU CRMA?*

The remainder of the document is organized as follows. In Chapter 2, the regulatory framework of the EU CRMA and the computational foundations of graph database systems are established; in Chapter 3, a systematic review of the related literature across supply chain stress testing and synthetic data generation is presented; in Chapter 4, the granular downscaling methodology, the network-native Nodal Elasticity metric, and the generalized multi-tier conceptual framework are formulated; in Chapter 5, the global tungsten trade ecosystem is deployed as an empirical case study to validate the proposed architecture; in Chapter 6, the quantitative graph analysis encompassing centrality and geopolitical risk metrics is detailed; in Chapter 7, the stress-testing simulation results and trader-level elasticity rankings are reported; finally, Chapter 8 concludes the thesis by discussing structural limitations and delineating primary trajectories for future work.

Chapter 2

Background

This chapter establishes the regulatory foundations necessary to contextualize the knowledge engineering methodology developed in this thesis. First, it reviews the strategic imperatives introduced by the European Union Critical Raw Materials Act (CRMA) [2], examining the macroeconomic indices currently utilized to monitor supply vulnerabilities. Second, it evaluates the architectural paradigms of graph database systems, formalizing the structural and computational rationale behind selecting Labeled Property Graphs (LPG) [47] over traditional relational architectures and semantic web triplestores. Finally, it outlines the core mathematical formulations of the centrality metrics available in the Neo4j Graph Data Science (GDS) library [49] and network topology, focusing on how these structural centrality metrics can be leveraged to quantify systemic risk and multi-hop dependency propagation.

2.1 The EU CRMA Regulation and Macroeconomic Risk Indicators

The structural vulnerability of modern industrial economies is increasingly dictated by their reliance on non-energy, non-agricultural raw materials that are highly concentrated at the primary extraction and refining stages. To institutionalize a coordinated response to these geopolitical dependencies, the European Union enacted Regulation (EU) 2024/1252, formally known as the CRMA. The regulation establishes a rigid dual classification system, distinguishing between "*Critical*" and "*Strategic*" raw materials based on their cross-sectoral economic importance, supply risk profile, and systemic relevance to the green, digital, defense, and aerospace transitions.

To safeguard the European industrial system, the CRMA defines specific macroeconomic capacity benchmarks to be achieved across the Union's supply architecture by 2030:

- **Domestic Extraction:** At least 10% of the Union's annual consumption of strategic raw materials must originate from domestic mining.
- **Domestic Processing:** At least 40% of the Union's annual consumption must undergo primary refining and chemical processing within Member States.

- **Domestic Recycling:** At least 25% of the Union’s annual consumption must be satisfied by secondary recovery and recycling infrastructures.
- **Import Diversification:** No more than 65% of the Union’s annual consumption of any strategic raw material at any stage of processing may originate from a single third country.

Before the operational enforcement of these benchmarks, the empirical identification of critical materials by the European Commission Joint Research Centre (JRC) [53] relied on a static, macro-level evaluation matrix structured around two primary axes: Economic Importance (EI) and Supply Risk (SR) [5].

The Economic Importance (EI) metric quantifies the systemic value of a material by mapping its end-use allocations across distinct manufacturing sectors, weighted by the gross value added (GVA) of those industries. It is functionally represented as:

$$EI = \sum_s A_s \cdot GVA_s \quad (2.1)$$

where A_s denotes the proportion of the specific raw material consumed within economic sector s , and GVA_s represents the corresponding macroeconomic value added of that sector relative to the overall economy. This structural weighting ensures that materials embedded in high-value manufacturing trajectories (such as advanced microelectronics or defense equipment) receive a proportionally higher criticality score, even if their absolute physical volume remains limited.

Concurrently, the Supply Risk (SR) framework measures the probability of a localized supply disruption and its subsequent transmission into the internal market. The JRC methodology models this risk by combining global production concentration with the governance quality of supplier nations, modified by recycling substitution rates and import reliance factors. The mathematical abstraction can be formalized as:

$$SR = \left(\sum_c S_c^2 \cdot WGI_c \right) \cdot (1 - EOL_{RRI}) \cdot IR \quad (2.2)$$

where S_c is the global market share of country c in the production of the material, WGI_c is the normalized Worldwide Governance Indicators score of country c (reflecting political stability, regulatory quality, and rule of law), EOL_{RRI} represents the End-of-Life Recycling Input Rate, and IR denotes the Union’s net import reliance.

The global production concentration component $\sum S_c^2$ directly mirrors the Herfindahl-Hirschman Index (HHI), a standard economic metric used to identify oligopolistic market architectures. HHI values range from close to 0 in perfectly atomistic markets to 1 in absolute monopolies. In the context of the CRMA, elevated HHI scores signify severe structural vulnerabilities, where localized regulatory shifts, environmental disasters, or geopolitical conflicts in a dominant country can paralyze entire global downstream value chains.

However, while these macro-level indicators are indispensable for high-level policy formulation, they suffer from a fundamental structural limitation: they are inherently static, country-level aggregates. They treat sovereign nations as homogenous, monolithic nodes, completely masking the firm-level dependencies, corporate ownership networks, and multi-hop trade relationships that actually execute international commerce. A country may

appear diversified based on trade statistics, yet its domestic trading entities might remain completely dependent on a handful of tightly integrated corporate actors abroad. Bridging this granular operational gap requires transitioning from aggregate country-level indicators to flexible, graph-native data structures capable of modeling micro-level trade networks.

2.2 Graph Databases and Knowledge Graph Architecture: Why Property Graphs?

Traditional data management paradigms are inherently unsuited for representing and querying multi-tier, highly dynamic supply chain networks. Relational Database Management Systems (RDBMS) enforce rigid schemas structured around tabular entities. When modeling a global procurement ecosystem within an RDBMS, tracing a material flow across multi-hop paths requires executing multiple nested relational join operations across foreign-key dependencies.

As the depth of the supply chain path increases, the computational complexity of these join operations degrades exponentially. This phenomenon, known as "*join pain*", occurs because the database engine must perform global index lookups and large-scale table scans to compute the intersections between tables at each sequential tier. For graph-structured questions — such as tracking the systemic exposure of a local manufacturer to a specific sub-tier supplier — RDBMS queries become computationally prohibitive and structurally opaque.

To overcome these limitations, modern knowledge management systems leverage graph-native architectures, which can be broadly categorized into two competing paradigms [47, 48]: Resource Description Framework (RDF) triplestores and Labeled Property Graphs (LPG).

RDF triplestores are rooted in the semantic web standards governed by the W3C [50]. They represent information as a collection of standardized statements structured as directional triples: Subject-Predicate-Object (e.g., `TraderA -[Supplies]-> TraderB`). RDF frameworks rely on formal ontologies designed using the Web Ontology Language (OWL) [51] and are queried via SPARQL [52]. Because every element within an RDF graph is bound to a unique Internationalized Resource Identifier (IRI), triplestores excel at ensuring global data interoperability, semantic consistency, and automated logical reasoning.

However, RDF architectures suffer from significant computational overhead when applied to complex network simulations or heavy algorithmic workflows. In their baseline implementation, standard RDF triples cannot naturally accommodate complex metadata directly on the relationships (predicates) themselves. Representing attributes such as trade value, transaction year, or physical mass requires a process called reification — creating a separate intermediary node to describe the connection — which multiplies the overall node count and dramatically reduces path traversal efficiency.

Conversely, the LPG architecture, as implemented in engines like Neo4j [48], provides an optimal balance between structural semantics and algorithmic execution speed. In an LPG model, entities are represented as distinct **Nodes** that can possess multiple labels (e.g., `:Trader`, `:Country`), and interactions are modeled as explicit, typed, directed **Relationships** (e.g., `[:EXPORTS]`). Crucially, both nodes and relationships can store arbitrary key-value pairs, known as **Properties**. This capability is fundamental for supply chain

modeling, as it allows financial volumes, physical mass, and temporal flags to be stored directly inside the relationship edge:

```
// Conceptual LPG relationship structure in Cypher notation
(s:Trader {id: "DEU_Trader_01"})
  -[e:SUPPLIES_TO {usd: 5400000, kg: 32000, year: 2024}]->
(t:Trader {id: "ITA_Trader_01"})
```

From a computational engineering perspective, the defining advantage of a native LPG engine is **Index-Free Adjacency (IFA)** [48]. In a database utilizing IFA, each node acts as a micro-index, holding direct physical memory pointers to its adjacent neighbouring nodes and relationships. The database engine can traverse the graph structure by simply dereferencing these memory pointers, completely bypassing the need for global index lookups during path execution. Consequently, the computational complexity of traversing a multi-hop procurement path depends strictly on the size of the subgraph being isolated, rather than the total volume of the global dataset. Traversal occurs in $O(1)$ constant time per hop, enabling rapid real-time execution of deep structural queries, geopolitical disruption routing, and systemic stress-testing across hundreds of simulated sub-tier layers.

2.3 Graph Data Science and Fundamentals of Network Analysis

GDS extends the structural utility of a Knowledge Graph by applying advanced mathematical and topological algorithms directly to the underlying network architecture. Within a property graph environment, a network can be formally defined as a directed weighted graph $G = (V, E, W)$, where V represents the set of heterogeneous nodes (Traders, Countries), $E \subseteq V \times V$ represents the set of directed trade relationships, and $W : E \rightarrow \mathbb{R}^+$ is a weighting function assigning an empirical value (such as economic trade volume) to each directed edge. The specific Knowledge Graph instance developed in this thesis — the Strategic Digital Twin, populated with Trader, Country, and Critical Raw Material nodes — is formally introduced in Section 4.3; the metrics formalized below are first presented in general graph-theoretic terms and are subsequently applied directly to that structure in Chapter 6.

To operationalize supply chain resilience monitoring within this graph space, we leverage structural network metrics to uncover hidden vulnerabilities, brokerage bottlenecks, and recursive dependencies that remain completely invisible to classical macroeconomic concentration indexes.

Weighted Betweenness Centrality

Betweenness Centrality measures the extent to which a node serves as an information or material broker by calculating the proportion of all shortest paths that pass through it. In a supply chain context, material flows do not traverse networks randomly [54]; they follow optimized procurement pathways. Let σ_{st} denote the total number of shortest paths from node s to node t , and let $\sigma_{st}(k)$ denote the number of those paths that pass through node

k . The standard betweenness centrality $C_B(k)$ is defined as:

$$C_B(k) = \sum_{s \neq t \neq k} \frac{\sigma_{st}(k)}{\sigma_{st}} \quad (2.3)$$

In our Strategic Digital Twin framework, we implement a **Weighted Betweenness Centrality**. To align the topological shortest path with economic realities, the edge weights are transformed using an inverse formulation, where $w_{ij} = 1/V_{ij}$, and V_{ij} represents the financial value of the trade flow from node i to node j .

By defining edge weights inversely to trade volume, relationships characterized by higher trade density are treated as low-resistance paths, minimizing the mathematical "distance" between nodes in shortest-path computations. Nodes that exhibit an elevated weighted betweenness score under this configuration function as critical structural brokers. If these specific trading entities are disrupted, the network lacks immediate alternative low-cost routing paths, forcing downstream economies to redirect procurement through highly fragmented, high-resistance channels.

PageRank Centrality

While betweenness captures brokerage control, PageRank measures recursive systemic importance [55]. Originally designed for web page ranking, the algorithm operates on the structural principle that a node's centrality is directly determined by the centrality of the neighbouring nodes pointing to it. In a global trade graph, a supplier node is central not merely because it has a high direct export volume (high in-degree), but because it supplies downstream entities that are themselves critically central to major industrial economies.

The PageRank score $PR(i)$ for a node i within a directed network is calculated iteratively according to the following steady-state distribution formula:

$$PR(i) = \frac{1-d}{n} + d \sum_{j \in M(i)} \frac{PR(j)}{L(j)} \quad (2.4)$$

where $d \in [0,1]$ is the damping factor (typically initialized at 0.85, representing the probability that a flow continues along the dependency path rather than randomly jumping), n is the total number of nodes in the network, $M(i)$ is the set of all nodes pointing to node i (its direct buyers), and $L(j)$ is the total out-degree of node j (the total number of suppliers node j buys from).

By executing PageRank to convergence over the trader-level dependency graph, we map how systemic risk and structural dependency accumulate across multi-hop links. A small upstream processor that feeds a highly connected European distributor will naturally accumulate a disproportionately high PageRank score, exposing it as a hidden systemic single-point-of-failure long before any material reaches the final target market.

Geopolitical Risk Exposure Index

To bridge internal network properties with external structural instability, GDS allows the mathematical fusion of topological centrality scores with country-specific macroeconomic risk indices. In isolation, a node's high PageRank score signifies high systemic integration,

but it does not specify whether that integration is safe or volatile. Conversely, a high country-level risk score means little if the firm operating within that nation is isolated from the global grid.

We formalize this integration by designing a composite **Geopolitical Risk Exposure Index** (GRI_k) for each trading entity k . The metric dynamically blends the node’s recursive structural prestige with the institutional stability of its sovereign location:

$$GRI_k = f\left(PR(k), WGI_{\text{country}(k)}\right) \quad (2.5)$$

By parameterizing the graph computation to scale a node’s risk exposure based on its topological centrality, the index separates structurally critical nodes operating in stable jurisdictions from central nodes exposed to severe political fragmentation. This multi-metric approach forms the analytical bedrock of the Strategic Digital Twin, ensuring that subsequent shock propagation models and Nodal Elasticity simulations are driven by a structurally grounded understanding of network topology.

2.4 Software Infrastructure and Data Sources

Before proceeding to the methodological framework, this section provides a concise overview of the software tools, libraries, and data sources employed throughout this thesis. The purpose is to equip the reader with sufficient context to understand the implementation decisions documented in Chapter 4, without requiring prior familiarity with any specific tool.

Neo4j Desktop

Neo4j is a native graph database management system built around the Labeled Property Graph model described in Section 2.2. For this thesis, Neo4j Desktop (version 2.1.4) was used as the primary database environment [56]. Neo4j Desktop is a locally installed application that provides a self-contained workspace for creating and managing graph database instances, installing plugins such as the GDS library and *Awesome Procedures On Cypher* (APOC), and executing Cypher queries through an integrated browser interface. Unlike cloud-hosted Neo4j deployments, the Desktop environment operates entirely offline, which was essential for the iterative experimental workflow of this thesis, where the database was repeatedly cleared and repopulated across multiple simulation runs.

Cypher Query Language

Cypher is the declarative query language native to Neo4j. It is designed specifically for graph-structured data and allows the user to express pattern-matching queries over nodes and relationships in an intuitive, ASCII-art-like syntax. For example, the pattern `(a:Trader)-[:SUPPLIES_T0]->(b:Trader)` retrieves all directed supply relationships between Trader nodes. Throughout this thesis, Cypher is used for three distinct purposes: importing synthetic node and relationship data into the database, executing structural queries such as shortest-path traversals and subgraph extractions, and invoking Graph Data Science algorithms such as weighted Betweenness Centrality and PageRank directly on the in-memory graph projection.

Neo4j Graph Data Science (GDS) Library

The GDS library is an official Neo4j plugin that exposes a catalogue of graph algorithms — including centrality measures, community detection, and path-finding algorithms — directly within the Cypher query environment [49]. GDS operates on in-memory graph projections: before running an algorithm, the user defines a projection that specifies which node labels and relationship types to include and how to weight the edges. The algorithm is then executed on this projection without modifying the underlying database. In this thesis, GDS is used to compute weighted Betweenness Centrality and PageRank scores across the synthetic tungsten trader network, as described in Section 2.3.

Python and the Neo4j Python Driver

Python (version 3.10+) [57] is the primary programming language used for the data processing and graph population pipeline. The official Neo4j Python driver (`neo4j` package) [58] is used to establish a connection between the Python runtime and the local Neo4j Desktop instance, and to execute batched Cypher `CREATE` and `MERGE` statements programmatically. Additional Python libraries used in the pipeline include `pandas` [59] for tabular data manipulation, `numpy` [60] for numerical operations, and `scipy` [61] for the Pareto distribution sampling and HHI calibration procedures.

Selenium and Undetected ChromeDriver

Selenium WebDriver is a browser automation library that allows Python scripts to control a Chrome browser instance programmatically [36]. In this thesis, Selenium is used exclusively within the Systematic Literature Review pipeline described in Section 3.5, where it automates the collection of bibliographic records from Google Scholar. The `undetected_chromedriver` variant is used in place of the standard ChromeDriver to reduce the likelihood of bot-detection triggering by Google’s anti-scraping mechanisms.

LMStudio and Qwen3-VL-8B

LMStudio is a desktop application that enables local inference of open-source large language models without requiring cloud API access [62]. For the purpose of this thesis, we employed the `Qwen3-VL-8B` model, which served as the classification agent within the automated literature review pipeline. The model runs entirely on local hardware, exposes a REST API (Application Programming Interface) endpoint compatible with the OpenAI chat completion format, and can be queried from Python using standard HTTP requests. This setup ensures that the classification process is fully reproducible and independent of any external service.

UN Comtrade Database

The United Nations Comtrade database is the primary empirical data source for this thesis. Comtrade is the world’s largest repository of international merchandise trade statistics, maintained by the UN Statistics Division [63]. It provides annual bilateral trade flow records at the HS (Harmonized System) commodity code level, reported separately by

exporting and importing countries. For this thesis, export-reported flows for tungsten-related HS codes (8101.10, 8101.94, 8101.99) for the reference year 2024 are extracted and used as the empirical foundation for the synthetic firm generation procedure described in Section 4.2. The decision to rely exclusively on export-reported flows, rather than import-reported flows, is motivated by the greater reporting reliability of export statistics for high-value industrial commodities such as tungsten, as discussed in Section 4.1.

Chapter 3

Related Work

This section reviews the key areas of literature and methodology that shaped the design decisions made throughout this thesis. The review is organized around four interrelated themes: supply-chain stress testing, which provides the analytical foundation for modeling disruption propagation; economic synthetic data generation, which motivates the Pareto-based firm reconstruction approach; the operational gap between macroeconomic criticality frameworks and micro-level corporate risk management, which defines the core problem this thesis addresses; and the automated literature screening pipeline developed as part of this research. A concluding section reflects on the specific tool and design choices made during the thesis, connecting the literature to the practical decisions encountered during implementation. These contributions collectively motivate the methodological approach developed in Chapter 4.

3.1 Stress tests

As systematized by Ivanov et al. (2023) [6], Supply Chain Stress Tests (SCSTs) can pursue multiple objectives, including ripple-effect analysis, disruption impact assessment, disruption prediction, and resilience evaluation. Existing approaches can also be broadly distinguished between reactive frameworks, which analyze post-shock propagation and recovery dynamics, and proactive approaches, which aim to identify vulnerabilities and anticipate disruptions before they materialize.

From a methodological perspective, the SCST literature has evolved along four main directions: simulation-based models, network-theoretic approaches, optimization frameworks, and artificial intelligence techniques. Simulation approaches mainly study disruption propagation and operational performance degradation under material flow shocks [7, 8, 9]. Network-based approaches focus on structural vulnerability, centrality, robustness, and hidden interdependencies within supply networks [10, 11, 12]. Optimization models analyze resilience and viability under disruption scenarios through stochastic and robust programming methods [13, 14, 15]. Finally, AI-based approaches employ machine learning and probabilistic techniques to predict disruptions and uncover latent supply-chain risks [16, 17, 18, 19, 20].

Recent studies [21] have increasingly extended these approaches to geopolitical and critical raw material disruptions, highlighting the importance of integrating network analysis, resilience assessment, and stress-testing methodologies within highly interconnected supply ecosystems. In this context, the proposed approach combines network-oriented stress testing with a graph-based reconstruction of firm-level trade dependencies, enabling topology-aware analysis of disruption propagation within critical raw material supply chains.

3.2 Economic synthetic data generation

The generation of synthetic firm-level trade networks is increasingly framed as a challenge of balancing statistical realism with structural economic consistency. A key stylized fact in this context is the granular nature of international trade, where aggregate trade outcomes are driven by a limited number of large firms [22]. Empirical evidence consistently shows that a small number of firms account for a disproportionate share of global trade flows [23, 24], consistent with power-law and Zipf-like regularities in firm-size distributions [25].

Recent advances in synthetic data generation include LLM-based tabular models [27], conditional GANs (Generative Adversarial Networks) such as CTGAN (Conditional GAN) [28], and diffusion models [29]. These approaches can reproduce complex, high-dimensional dependencies and realistic marginal distributions. However, they remain primarily data-driven and do not explicitly encode economic structure. As a result, they may fail to preserve key structural properties such as concentration patterns, oligopolistic market organization, and consistency between macro-level aggregates and micro-level network representations.

In addition, these methods are often difficult to interpret and control from a structural perspective, since economic constraints emerge implicitly from training data rather than being imposed by design. This limits transparency and reproducibility, particularly in applications requiring explicit modeling of supply-chain organization and disruption propagation mechanisms. In contrast, the proposed methodology explicitly incorporates concentration constraints, Pareto-based firm heterogeneity, and macro-level trade consistency by design, enabling interpretable and structurally coherent reconstruction of firm-level supply networks under limited data availability.

More directly related to the firm-level reconstruction pursued in this thesis, recent work has sought to infer unobserved supply-network structure from partial data using machine-learning and maximum-entropy techniques, analyzing the reconstructed networks through standard centrality measures such as degree and betweenness [38]. Complementary efforts have produced semantic data generators that synthesize structurally consistent supply-chain datasets for benchmarking and simulation [39]. These approaches share the present work’s objective of constructing analyzable firm-level networks under limited data transparency, but differ in emphasis: whereas they prioritize predictive inference or benchmarking flexibility, the methodology developed here prioritizes explicit, economically interpretable concentration constraints, calibrating a Pareto firm-size structure directly against an observed market-concentration target.

3.3 Knowledge Graphs and Digital Twins for Supply-Chain Risk

Parallel to the stress-testing and synthetic-data literatures, a rapidly growing body of research has applied knowledge-graph and digital-twin technologies to supply-chain risk management, and it is against this backdrop that the present architecture is positioned. Knowledge graphs have been used to consolidate fragmented supply-chain information and improve visibility of risk in financial, operational, and credit dimensions [40], and to address the disconnect in data that hampers the modeling of cascading risk events [41]. More recent contributions move beyond data integration toward resilience-oriented reasoning, using knowledge graphs and machine learning to identify latent risks [42] and employing ontology-based semantic representations to manage and evaluate supply-chain disruptions [43].

In parallel, digital-twin research has progressed from general modeling paradigms for complex networked systems [44] toward concrete sectoral applications, including the management of disruptions in semiconductor supply chains [45] and the mitigation of trade-disruption and tariff volatility through twin-enabled flexibility models [46]. These contributions demonstrate the maturity of graph-native and twin-based methods for supply-chain analysis. However, they predominantly target either real-time operational synchronization driven by IoT (Internet of Things) telemetry, or domain-specific manufacturing and logistics settings. None couples a firm-level, graph-native representation of a critical-raw-material trade network with quantitative, regulation-aligned stress testing under conditions of acute data scarcity — the specific combination that motivates the Strategic Digital Twin developed in this thesis, and that distinguishes its *structural* orientation from the operational twins prevailing in the literature.

3.4 Research Gap: Macro-Criticality Frameworks vs. Micro-Corporate Management

While the literature on supply chain stress testing and synthetic data generation has advanced significantly, a structural fragmentation remains prominent when looking through the lens of recent regulatory requirements, specifically the European Union’s Critical Raw Materials Act (CRMA) - Regulation (EU) 2024/1252. Article 24 of the CRMA enforces a strict legal obligation on large manufacturing firms to perform comprehensive supply chain risk assessments on a triennial basis. Obligated enterprises are required to trace strategic raw materials across every structural stage of their value chain (extraction, refining, processing, and recycling) down to the primary geographical point of origin.

However, a systematic cross-examination of the current state of the art reveals that the scientific literature is heavily split into two disconnected domains:

- **Macroeconomic Criticality Frameworks:** Sophisticated methodologies such as the EU Critical Raw Materials assessment, the Yale Criticality Framework, and ES-SENZ rely entirely on aggregate indicators like Economic Importance (EI), Supply Risk (SR), and the Herfindahl-Hirschman Index (HHI) [2, 5]. These frameworks function strictly at the system-wide level (continents, country-systems, or broad economic sectors) [2, 5]. They produce macro-level indexes that an individual corporate actor

cannot directly use without complex contextualization.

- **Microeconomic Corporate Operations:** At the operational corporate level, manufacturing firms manage dependencies using a standard Bill of Materials (BOM). A corporate BOM meticulously details part numbers, engineering descriptions, and direct Tier-1 procurement suppliers. However, standard industrial management practices do not trace components back to the raw chemical or mineral constituents. For instance, a manufacturer records a sub-assembly component from an immediate supplier. Still, the ordinary corporate information layout possesses no visibility over the exact mass of critical raw materials (such as tungsten or rare earths) embedded within that component, let alone their primary refinement location.

Consequently, a profound operational gap exists between macroeconomic risk indicators and microeconomic data structures. The literature contains a noticeable deficit regarding integrated conceptual models that can simultaneously bridge a corporate BOM, a multi-tier supply chain graph, aggregate macroeconomic criticality metrics, and the quantitative constraints established by modern regulations. This thesis addresses this specific fragmentation by evaluating a graph-based framework capable of connecting these distinct dimensions.

The corpus curated through the systematic literature review described in Section 3.5 reinforces this conclusion in a systematic rather than anecdotal manner. The retained studies are distributed across the four methodological categories of the screening taxonomy — ontology and knowledge-graph implementation, structural and centrality analysis, digital-twin and simulation, and stress testing and resilience — yet their application domains lie almost entirely outside critical raw materials, spanning agriculture and food systems, construction, plastics and circular-economy settings, logistics, and semiconductor manufacturing. Knowledge-graph and digital-twin methods are thus demonstrably mature in adjacent fields, but no screened study couples a firm-level, graph-native representation of a critical-raw-material trade network with quantitative stress testing aligned to the traceability and risk-assessment mandates of the EU CRMA. This absence, surfacing through a transparent and reproducible screening protocol, is precisely the gap that the present thesis addresses.

3.5 Systematic Literature Review Methodology via LLM-Powered Pipelines

To rigorously confirm the research gap identified in the preceding sections and systematically identify studies mapping to the intersection of Knowledge Graphs, Critical Raw Materials, and Supply Chain Resilience, a Systematic Literature Review (SLR) was conducted following a formalized and transparent protocol. Given the exponential growth of academic publications between January 2022 and March 2026, a traditional manual search strategy was deemed insufficient and prone to human selection bias. To ensure strict auditability and broad temporal coverage, a semi-automated paper curation pipeline was custom-engineered as part of this research workflow.

Pipeline Architecture

The pipeline combines browser automation with local Large Language Model (LLM) inference to execute a targeted taxonomy screening across Google Scholar. It is organized into two independent software modules, each responsible for a distinct stage of the review process.

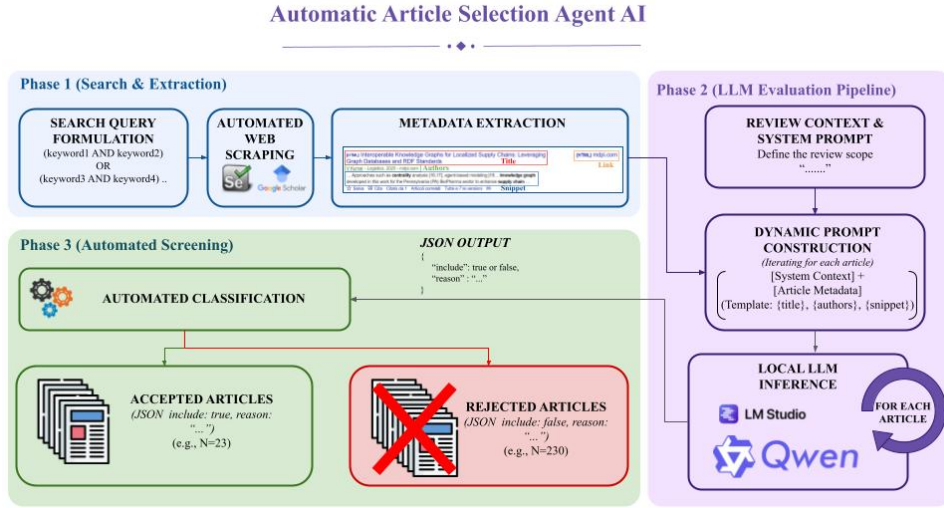


Figure 3.1. Full architecture of the semi-automated Systematic Literature Review pipeline, from automated Google Scholar scraping to LLM-based classification and curated output generation.

The high-level data flow is the following. The Selenium-based scraper¹ queries Google Scholar using a structured Boolean search string and collects, for each result, the title, author list, publication year, and source URL (Uniform Resource Locator), together with the short preview snippet displayed on the results page. Where the full abstract of a paper could also be retrieved, it was collected in addition to the snippet, providing richer context for the classification step described below. All raw bibliographic records are exported to an Excel file, which is then consumed by the LLM classification agent, evaluating each paper independently and producing a structured JSON (JavaScript Object Notation) annotation for each record. The final curated output, containing only the papers that satisfy the inclusion criteria, is exported to a separate Excel file for manual review and integration into the thesis bibliography.

¹Selenium is an open-source browser automation framework that allows a program to control a web browser programmatically, simulating human interaction such as page navigation, clicking, and text extraction. Available at: <https://www.selenium.dev>. Last accessed on July 10, 2026.

Module 1: Automated Literature Scraper

The first module, implemented in `find_papers.py`, automates the collection of bibliographic data from Google Scholar using Python and the `undetected_chromedriver` library², which launches a Chrome browser instance configured to bypass standard bot-detection mechanisms. The scraper navigates search result pages programmatically, collecting the title, author list, publication year, snippet text, and source URL for each result.

The search query used for this thesis was structured using Boolean operators to target the specific intersection of topics relevant to the research:

Listing 3.1. Boolean search query used for the automated Google Scholar literature scrape.

```
("knowledge graph" OR "semantic network" OR
 "supply chain ontology" OR "graph-based modeling")
AND
("supply chain" OR "raw materials" OR
 "logistics network" OR "material flow")
AND
("centrality" OR "betweenness centrality" OR
 "network analysis" OR "graph metrics")
AND
("digital twin" OR "dynamic modeling" OR
 "real-time simulation")
AND
("stress testing" OR "resilience analysis" OR
 "shock simulation" OR "disruption modeling")
```

A key practical challenge encountered during the scraping process is that Google Scholar periodically activates CAPTCHA (*Completely Automated Public Turing-test-to-tell Computers and Humans Apart*) challenges to throttle automated requests. The pipeline handles this through built-in pause states: when a CAPTCHA is detected, the browser window remains open and visible, allowing the researcher to solve the challenge manually. The scraper then resumes automatically once the results page reloads. This semi-automated approach ensures that the data collection process is not fully blocked by anti-scraping measures while still maintaining a high degree of automation. All raw bibliographic records are exported to `data/articles.xlsx` for downstream processing.

Module 2: Local LLM Classification Agent

The second module, implemented in `paper_selection.py`, consumes the raw bibliographic data and classifies each paper according to a predefined taxonomic framework. To avoid introducing personal selection bias and to maintain full reproducibility without relying on commercial API services, a local LLM inference server was deployed using LMStudio³, hosting the open-source Qwen3-VL-8B model⁴.

²<https://pypi.org/project/undetected-chromedriver/>, last accessed on July 10, 2026.

³<https://lmstudio.ai/>, last accessed on July 10, 2026.

⁴<https://lmstudio.ai/models/qwen/qwen3-vl-8b>, last accessed on July 10, 2026.

The decision to use a locally hosted model rather than a commercial API, such as GPT-4 or Claude, was deliberate. Local inference ensures that classification decisions are fully reproducible, independent of external service availability, and not subject to policy changes or rate limits imposed by external providers. Furthermore, because each paper evaluation is independent of the others, a completely fresh chat session is instantiated for every paper before the classification prompt is submitted. This design choice eliminates the risk of hallucination carry-over [64], where the model’s judgment on a previous paper might unintentionally influence its evaluation of the next one.

Each paper is evaluated against the following taxonomic prompt template, into which the title, author list, and snippet are dynamically inserted at runtime:

Listing 3.2. Taxonomic classification prompt template submitted to the local LLM agent for each paper.

```
You assist a literature review about raw material value chain
analysis using knowledge graphs, digital twins, centrality
measures, and stress testing.

Taxonomy categories:
  1. Domain Ontology & Knowledge Graph Implementation
  2. Structural Analysis & Centrality Measures
  3. Supply Chain Digital Twin & Simulation
  4. Stress Testing & Resilience Analysis

Evaluate the following paper in the context of this review.

Title:    {title}
Authors:  {authors}
Snippet:  {snippet}

Return ONLY a JSON object:
{
  "include": true or false,
  "category": "",
  "reason":  ""
}
```

The model is instructed to return a structured JSON object containing three fields: a binary inclusion flag indicating whether the paper satisfies the review criteria, a category assignment drawn from the predefined taxonomy, and a concise textual justification for the classification decision. An example of a negative classification output produced by the agent is the following:

Listing 3.3. Example LLM classification output for a paper that did not meet the inclusion criteria.

```
{
  "include": false,
  "category": "",
  "reason":  "The paper focuses on semantic network analysis
              in the context of urban waste and circular
              economy, but does not address raw material value
```

```
} chain analysis using knowledge graphs, digital  
twins, centrality measures, or stress testing  
as required by the taxonomy."
```

Limitations of the Pipeline

Despite its advantages in terms of reproducibility and bias reduction, the pipeline carries several methodological limitations that must be acknowledged. First, the classification agent evaluates each paper based exclusively on its title, author list, and abstract, rather than the full text. Papers whose relevance is only apparent from their methodology or results sections may therefore be incorrectly excluded. Second, although the full abstract was retrieved and used whenever accessible, this was not possible for every paper; in such cases the classification relied on the truncated preview snippet returned by Google Scholar, which does not always correspond directly to the paper abstract and may introduce noise into the classification input. Finally, as with any LLM-based classification, the agent may produce incorrect category assignments or borderline inclusion decisions, particularly for interdisciplinary papers that partially satisfy multiple taxonomy categories simultaneously.

Chapter 4

Methodology and Contributions

This chapter documents both the engineering decisions that shaped the development of this thesis and the methodological framework that underlies the empirical analysis. The chapter is organized as follows. Section 4.1 reflects on the key design choices made during the research process, connecting the literature reviewed in Chapter 3 to the practical and conceptual decisions encountered throughout the work, and summarizes the individual contributions of the author. Section 4.2 then introduces the synthetic downscaling procedure used to disaggregate country-level trade flows into firm-level structures consistent with observed market concentration and aggregate trade volumes. Section 4.3 describes the Knowledge Graph architecture and its implementation within the Strategic Digital Twin. Section 4.4 formalizes the trader-level network structure, the definition of Nodal Elasticity, and the stress-testing methodology used to simulate supply disruptions under fixed trade dependencies. Finally, Section 4.5 proposes an original generalized multi-tier and multi-commodity conceptual framework that extends the baseline model beyond a single topological layer, bridging micro-level corporate Bill of Materials structures with macro-level criticality indicators in alignment with the regulatory mandates of the EU Critical Raw Materials Act.

4.1 Design Choices and Contributions

This section reflects on the key design decisions made during the development of this thesis, connecting the literature reviewed above to the practical and conceptual choices encountered throughout the research process. Its purpose is to make explicit the reasoning behind the tools, methods, and modeling approaches adopted, and to distinguish the author’s individual contributions from the shared implementation developed within the research group.

Choice of Neo4j as the Graph Database Engine

The decision to implement the Strategic Digital Twin in Neo4j as a Property Graph, rather than in a relational database or an RDF-based triplestore, was motivated by several practical and structural considerations. Supply chain networks are intrinsically graph-structured: entities such as traders, countries, and raw materials are connected by typed, directed relationships whose traversal is the primary analytical operation. Relational databases require expensive multi-table joins to replicate this traversal, while RDF triplestores, although semantically richer, introduce significant overhead in query formulation and lack the native support for weighted graph algorithms available in Neo4j’s Graph Data Science (GDS) library. Neo4j’s Cypher query language also allowed rapid prototyping of centrality metrics (Betweenness, PageRank) and geopolitical risk exposure indices directly on the graph structure, without requiring external processing steps.

Neo4j Desktop as the Deployment Environment

Among the available Neo4j deployment options, this thesis adopts *Neo4j Desktop* (version 2.1.4) as the primary working environment, rather than the web-based Neo4j Browser or a cloud-hosted AuraDB instance. This choice was deliberate and motivated by several practical and technical considerations.

Neo4j Desktop provides a fully self-contained local environment in which the database instance, the Graph Data Science (GDS) library, and the APOC procedure library can all be installed and managed without any external network dependency. For a research setting where the database is being actively designed, iteratively restructured, and repeatedly reloaded with synthetic data across multiple experimental runs, local control over the database instance is essential. A cloud-hosted instance would introduce latency in every query and import operation, and would require continuous internet connectivity during long-running simulations.

Furthermore, Neo4j Desktop includes an integrated project management interface that allows multiple database instances to be maintained in parallel under a single environment. This was particularly useful during the development phase of this thesis, where different versions of the tungsten graph — with varying numbers of synthetic traders, alternative Pareto calibrations, and different shock scenarios — needed to be stored and switched between without corrupting the production graph. The Neo4j Browser, by contrast, is a query interface only and does not provide this level of environment management.

Finally, Neo4j Desktop exposes a locally accessible database endpoint, which is required for the driver-based ingestion procedure used in this thesis. The Python pipeline described in the following section connects directly to this local instance through the official Neo4j Python driver and writes nodes and relationships programmatically via batched Cypher `CREATE` and `MERGE` statements. This direct, low-latency connection between the Python runtime and the database instance is considerably simpler to configure for a locally hosted Desktop instance than for browser-based or cloud deployments, which typically require additional authentication and network configuration.

Data Ingestion Pipeline: From Python to Neo4j

The population of the Strategic Digital Twin with synthetic trade data is handled by a dedicated Python pipeline, referred to as the *CRM-KG Generator*. Rather than manually constructing graph entities through the Neo4j Browser interface, the entire node and relationship generation process is automated end-to-end: from loading raw UN Comtrade trade flow data for writing the final graph structure into the local Neo4j Desktop instance.

The pipeline is organized into the following functional modules, each responsible for a distinct stage of the data processing and graph construction workflow:

- `load_files.py` — Loads raw UN Comtrade trade flow data and auxiliary configuration files into memory.
- `processing.py` — Transforms raw trade volumes into country-level export shares and bilateral trade matrices, applying the directional consistency constraints described in Section 4.1.
- `hhi.py` — Computes the global supplier-side HHI from the processed trade data and executes the Pareto shape parameter calibration procedure described in Section 4.2.
- `generator/nodes.py` — Generates the node population: Country nodes with geopolitical risk attributes, synthetic Trader nodes with Pareto-calibrated market shares, and the Critical Raw Material node with its associated CRMA classification attributes.
- `generator/edges.py` — Constructs the directed relationship set: `LOCATED_IN` edges connecting Traders to their sovereign Country, `SUPPLIES_TO` edges encoding bilateral synthetic trade flows between Traders, `HAS_RISK` edges linking Countries and CRMs to their associated risk nodes, and `TRADES` edges recording the material-level trading activity.
- `risk.py` — Computes the Geopolitical Risk and Concentration Risk node attributes for each Country and for the CRM node respectively.
- `download/import_neo4j.py` — Establishes a connection to the local Neo4j Desktop instance via the official Python Neo4j driver and executes batched `CREATE` and `MERGE` Cypher statements to write all nodes and relationships into the graph database.

The pipeline is executed through a single entry point, `main.py`, which orchestrates the modules in the correct dependency order. Before each experimental run, the database is reset using the following Cypher cleanup procedure executed directly in the Neo4j Desktop query interface:

Listing 4.1. Database cleanup procedure executed before each experimental run.

```
// Delete all nodes and relationships
MATCH (n)
DETACH DELETE n;

// Drop all constraints and indexes
CALL apoc.schema.assert({}, {}, true);
```

Following the cleanup, `main.py` is executed from the terminal. The pipeline connects to the Neo4j Desktop local server using a URI and an encrypted password stored in a `.env` file, reconstructs the full synthetic graph from scratch, and loads it into the database. Once the import is complete, most of the analytical queries — including centrality computations and geopolitical risk scoring — are executed directly in Neo4j using Cypher and the Graph Data Science library, without any further external processing steps.

The database schema resulting from this import procedure can be inspected using the following Cypher command:

Listing 4.2. Schema visualization query for inspecting the imported graph structure.

```
CALL db.schema.visualization()
```

The full graph, including all nodes and relationships across both the Empirical and Synthetic layers, can be retrieved with:

Listing 4.3. Full graph inspection query.

```
MATCH (n)-[r]->(m)
RETURN n, r, m
```

A particularly useful diagnostic query for validating the consistency of the import is the per-country trade flow analysis, which computes international inflow, outflow, and net balance for each country node in the graph:

Listing 4.4. Per-country trade flow analysis query for import validation.

```
MATCH (c:Country)
WITH c.iso3 AS country_iso
CALL {
  WITH country_iso
  MATCH (s:Trader)-[r:SUPPLIES_TO]->(t:Trader)
  WHERE t.country_iso = country_iso
    AND s.country_iso <> country_iso
  RETURN COALESCE(SUM(r.kg), 0) AS inflow_international
}
CALL {
  WITH country_iso
  MATCH (s:Trader)-[r:SUPPLIES_TO]->(t:Trader)
  WHERE s.country_iso = country_iso
    AND t.country_iso <> country_iso
  RETURN COALESCE(SUM(r.kg), 0) AS total_outflow
}
RETURN
  country_iso AS country,
  inflow_international,
  total_outflow,
  (inflow_international - total_outflow) AS balance
ORDER BY country;
```

This query serves as a mass-balance check: the aggregate outflow across all countries must be consistent with the total import volume encoded in the original UN Comtrade bilateral trade matrix, confirming that no trade value has been lost or duplicated during the synthetic firm allocation procedure.

Working with UN Comtrade Data

A significant practical challenge encountered during the thesis was the extraction, cleaning, and aggregation of UN Comtrade trade flow data for tungsten-related Harmonized System (HS) Codes. The raw data presents several structural issues: reporter asymmetries between exporting and importing countries, missing flow records, and inconsistent classification of product subcategories across reporting years. The decision to rely exclusively on export-reported flows (rather than import-reported flows) was made to ensure directional consistency within the trade matrix, as Comtrade export data tends to be more reliably reported for high-value industrial commodities such as tungsten.

Calibration of the Pareto Shape Parameter

The choice to use a Pareto distribution for synthetic firm generation, rather than a data-driven model such as CTGAN or a diffusion model, was deliberate and grounded in the related work reviewed in Section 3. The primary reason is interpretability: the Pareto shape parameter α has a direct economic interpretation in terms of firm-size concentration, and its calibration against the observed global HHI provides a transparent, auditable link between macro-level data and the synthetic micro-level network. Data-driven generative models, by contrast, encode concentration patterns implicitly and are difficult to constrain to specific economic targets. Given the limited availability of firm-level ground truth data for tungsten, structural controllability was prioritized over distributional flexibility.

Motivation for the Multi-Tier Conceptual Framework

The multi-tier and multi-commodity conceptual framework proposed in Section 4.5 emerged directly from the regulatory analysis conducted in Section 3 and from the limitations encountered during the baseline implementation. The EU CRMA Article 24 obligations require traceability across every structural stage of the value chain, from extraction to recycling, for each strategic raw material. The baseline model (Case 1) collapses all these stages into a single flat layer of Traders, which is sufficient for aggregate stress testing but does not support the stage-specific visibility required by regulation. The multilayer graph formalization was therefore developed as a theoretically grounded blueprint to fill this gap, while acknowledging the data scarcity constraints that prevent its immediate empirical implementation.

4.2 A Granular Downscaling Methodology for CRMs

The proposed methodology aims to disaggregate country-level trade flows into synthetic firm-level structures while preserving the concentrated structure observed in the global tungsten market. As a first step, the methodology assigns to each country c a target number of synthetic firms proportional to its share in global tungsten exports:

$$N_c = \min \{ \max [1, \lceil s_c K \rceil], N^{\max} \}, \quad (4.1)$$

where s_c is country c 's share of total global exports, $N^{\max}$ is an upper bound imposed to avoid excessively large synthetic structures, and K is a scaling parameter controlling

the overall density of the synthetic network. Higher values of K increase the number of synthetic firms generated across countries, leading to a more granular representation of the supply chain but also to a larger and computationally more demanding trade network.

While the previous step determines the number of synthetic firms associated with each country, the model must also reproduce the concentration patterns observed in the global tungsten market. Since firm-level trade data are not directly observable, the concentration of the synthetic firm structures is calibrated using the aggregate concentration observed in country-level export data. More specifically, the approach computes a global supplier-side HHI from UN Comtrade data:

$$\text{HHI}^{\text{global}} = \sum_{c=1}^C s_c^2, \quad (4.2)$$

where s_c denotes the share of total global tungsten exports accounted for by country c , and C is the total number of exporting countries included in the market. The observed global HHI is then used as a calibration target for the Pareto shape parameter α , which governs the synthetic firm-level distributions generated within each country. In particular, for a given number of firms N_c , we generate multiple synthetic firm-share distributions under alternative values of α . For each simulated distribution, the corresponding HHI is computed and compared with the observed global supplier-side HHI. The calibration procedure then selects the value of α that minimizes the difference between the simulated and observed concentration levels.

As a result, the methodology uses the aggregate concentration observed at the global market level to infer plausible firm-level concentration structures within each country. The underlying rationale is that highly concentrated country-level export structures are unlikely to emerge from evenly distributed firm-size structures, and instead typically reflect the presence of dominant firms operating within oligopolistic market configurations. In the absence of directly observable firm-level trade data, aggregate concentration observed in international trade statistics therefore provides a structural signal that can be used to constrain the generation of plausible micro-level market structures. Lower values of α generate highly skewed distributions dominated by a few large firms, whereas higher values produce more fragmented structures. The objective of the procedure is not to identify real firms directly, but to reconstruct plausible firm-level concentration structures consistent with observed aggregate trade patterns under conditions of limited micro-level data availability. In this sense, the synthetic traders should be interpreted as structural approximations of market organization rather than as exact representations of observable firms.

To ensure mass-balance consistency with the original macro data, the generated firm-level values $q_{i,c}$ are scaled so that, for each country c :

$$V_c = \sum_{i=1}^{N_c} q_{i,c}, \quad (4.3)$$

where $q_{i,c}$ denotes the trade value associated with synthetic firm i in country c , and V_c denotes the observed aggregate trade value of country c . The resulting synthetic firms are then represented in the network as *Trader* nodes, corresponding to economic agents involved in the import, export, processing, or redistribution of tungsten-related products. Bilateral

trade flows are then allocated across synthetic traders proportionally to exporters’ and importers’ market shares while preserving consistency with the observed aggregate bilateral trade matrix. These synthetic trader-level entities constitute the micro-level components of the Strategic Digital Twin described in the following subsection.

A key advantage of this approach is its interpretability and controllability: firm heterogeneity, concentration, and network structure are explicitly governed by a small set of economically meaningful parameters, rather than emerging implicitly from black-box learning procedures. This makes the model fully reproducible and analytically tractable, which is essential for stress-testing applications where causal propagation mechanisms must be clearly identified and isolated. In addition, the explicit structure of the model allows for direct calibration to observed concentration measures, ensuring coherence with empirical regularities while maintaining computational efficiency in large-scale simulations.

A refinement is applied to the largest producing countries, for which a small number of dominant firms are publicly identifiable. Rather than generating the entire national firm population synthetically, these well-documented major producers are retained as *anchor firms* and seeded into the network with their approximate observed market shares, flagged through the node property `is_anchor`. The remaining firms required to reach the target count N_c are then drawn from the calibrated Pareto distribution and jointly rescaled with the anchors so that the country’s aggregate export value V_c is preserved. This hybrid construction retains the identity and relative weight of the real market leaders — for which reliable information exists — while relying on synthetic generation only for the long tail of smaller traders, where firm-level data are unavailable. Figure 4.1 illustrates the resulting structure for China: the dominant anchor firms (such as Xiamen Tungsten, China Minmetals, and Jiangxi Tungsten) carry the bulk of the national market share, while the synthetic Pareto-generated traders populate the long tail. The approach thus combines empirical fidelity at the top of the firm-size distribution with structurally consistent synthetic coverage of its tail.

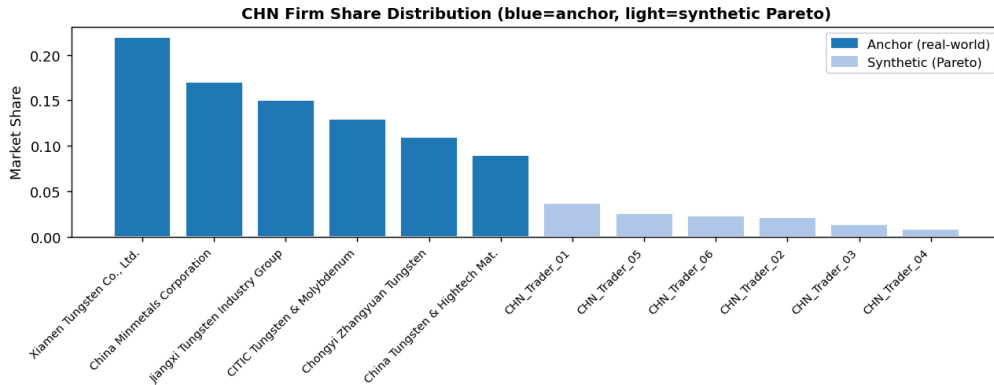


Figure 4.1. *Synthetic firm-share reconstruction for China under the Pareto-calibrated downscaling procedure. The dark bars denote the real-world anchor firms, whose approximate market shares are retained from observed data, while the lighter bars denote the synthetic traders generated from the calibrated Pareto distribution to populate the long tail of the national firm-size structure.*

Step-by-Step Implementation Summary

To make the downscaling procedure fully transparent and replicable, the complete sequence of implementation steps is summarized below. This summary bridges the mathematical formalization presented above with the concrete computational operations executed by the Python pipeline.

1. **Data Loading:** Bilateral trade flow records for tungsten HS codes 8101.10, 8101.94, and 8101.99¹ are extracted from UN Comtrade for the reference year 2024. Only export-reported flows are retained to ensure directional consistency across the trade matrix.
2. **Country Share Computation:** For each exporting country c , the global export share s_c is computed by dividing the country's total tungsten export value by the sum of all reported global tungsten exports.
3. **Global HHI Estimation:** The global supplier-side HHI is computed from the country-level export shares as $\text{HHI}^{\text{global}} = \sum_c s_c^2$. For the 2024 tungsten dataset, this yields a value of 0.157 (with a top-supplier share of approximately 31.7% and a CR3 of approximately 60.4%), indicating a moderately concentrated export market in which a single dominant supplier coexists with a sizeable tail of secondary exporters.
4. **Firm Count Assignment:** The number of synthetic firms N_c assigned to each country is determined according to Equation (4.1), using the scaling parameter K and the upper bound N^{max} specified in the pipeline configuration file.
5. **Pareto Calibration:** Multiple synthetic firm-share distributions are sampled under a grid of candidate α values. For each candidate, the corresponding simulated HHI is computed and compared against the observed global HHI of 0.157. The value of α that minimizes the absolute difference between simulated and observed concentration is selected as the calibrated shape parameter.
6. **Firm Share Sampling:** For each country, identifiable major producers are first seeded as anchor firms with their observed market shares. The remaining firms required to reach the target count N_c are then sampled from the Pareto distribution with the calibrated shape parameter α and minimum scale parameter x_m .
7. **Rescaling:** The sampled firm shares within each country are normalized and rescaled so that their sum equals the observed aggregate export value V_c for that country, ensuring mass-balance consistency with the original Comtrade data.
8. **Bilateral Flow Allocation:** Trade flows between synthetic traders are allocated proportionally to exporter and importer market shares, preserving consistency with the observed aggregate bilateral trade matrix at the country level.

¹The choice of these specific codes will be explained in Section 5.

9. **Graph Population:** The resulting synthetic Trader nodes and `SUPPLIES_TO` relationship edges are written into the Neo4j Desktop instance via the Python pipeline, as described in Section 4.1.

This step-by-step transparency is a deliberate design principle of the methodology. Unlike black-box generative models, every intermediate output of the downscaling procedure can be inspected, validated, and rerun independently. A policy analyst or regulator seeking to apply this framework to a different critical raw material — such as lithium, cobalt, or rare earth elements — would need only to replace the Comtrade input data and adjust the HS code configuration, with no changes required to the core algorithmic pipeline.

4.3 Knowledge Modeling Framework

Synthetic firm-level entities and trade relations are integrated into a Knowledge Graph implemented in Neo4j as a Property Graph. This structure combines Countries, Traders, Critical Raw Materials, and market concentration metrics within an ontology-driven framework (Figure. 4.2). Core entities and attributes are summarized in Table 4.1. In contrast to operational digital twins that depend on real-time IoT telemetry, the proposed SDT emphasizes structural supply-chain analysis, disruption propagation, and resilience-oriented stress-testing. The proposed SDT should therefore be interpreted as a structural and analytical digital twin, rather than a real-time operational twin driven by continuous telemetry or live synchronization. This approach supports explainable modeling of trade dependencies using multiple information sources. The framework distinguishes two complementary layers. The Empirical Layer (gray nodes) includes validated macro-level entities derived from observed data, such as Countries and Critical Raw Materials. The Synthetic Layer (blue nodes) consists of Trader nodes generated through the downscaling procedure described in Section 4.2.

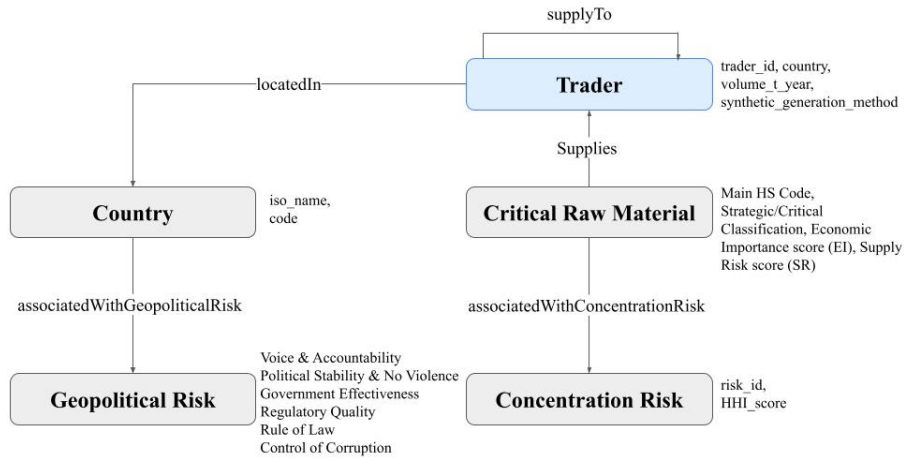


Figure 4.2. *Ontology-driven conceptual model of the Knowledge-Graph-based Strategic Digital Twin*

Entity	Description	Key Attributes (Data Properties)
Trader	Synthetic economic agent involved in the import, export, processing, or redistribution of tungsten-related products.	company_name, country_iso, id, annual_volume_usd, share_pct, role, is_anchor
Country	Sovereign state where a trader operates or raw materials are extracted.	name, iso3, geo_risk_score, wgi_composite, eu_member, export_restriction_active
Critical Raw Material	Strategic mineral or material included in the supply-chain analysis.	name, hs_code, strategic_category, id
Concentration Risk	Market-concentration indicator associated with the global supply distribution of the material.	hhi_score, cr3, top_supplier_iso, top_supplier_share_pct, benchmark_40pct_breach
Geopolitical Risk	Country-level political-stability indicator linked to each sovereign jurisdiction via HAS_RISK .	geo_risk_score, wgi_composite, political_instability, export_restriction_flag

Table 4.1. *Core entities and associated attributes of the ontology-driven Strategic Digital Twin*

Relationship Types and Their Semantic Roles

Beyond the node entities summarized in Table 4.1, the graph structure is defined by four distinct directed relationship types, each encoding a semantically distinct interaction within the supply chain ecosystem:

- **LOCATED_IN**: Connects each Trader node to the Country node corresponding to its sovereign jurisdiction. This relationship enables geopolitical risk scores to be propagated from the country level to the individual trading entity level during the GRI computation described in Section 2.3.
- **SUPPLIES_TO**: The primary supply relationship, connecting a supplying Trader to a receiving Trader. Each **SUPPLIES_TO** edge carries the properties **usd** (the bilateral trade value in USD), **kg** (the traded mass), and **year** (the reference year), enabling value- and volume-weighted network analysis.
- **HAS_RISK**: Connects Country nodes and the Critical Raw Material node to their associated Geopolitical Risk and Concentration Risk nodes respectively. This relationship externalizes risk attributes from the core supply graph into dedicated risk nodes, keeping the primary graph structure clean and enabling modular risk score updates without modifying the supply topology.
- **TRADES**: Connects Trader nodes to the Critical Raw Material node, recording which synthetic trading entities are active participants in the tungsten market. This relationship supports commodity-level filtering queries, allowing the analyst to isolate the

subgraph of traders active in a specific material without traversing the full network.

The conceptual ontology presented in Figure 4.2 is realized in Neo4j as a concrete, queryable database schema. Figure 4.3 shows this implemented schema exactly as returned by the built-in `db.schema.visualization()` procedure, which introspects the live database and renders every node label together with the relationship types that connect them. The diagram confirms that the populated graph contains precisely the five entity classes defined in the ontology — `Trader`, `Country`, `CriticalRawMaterial`, `ConcentrationRisk`, and `GeopoliticalRisk` — and the four directed relationship types through which they interact. The `SUPPLIES_TO` self-loop on the `Trader` label encodes the firm-to-firm trade relationships that form the analytical core of the network; `LOCATED_IN` anchors each `Trader` to its sovereign `Country`; `TRADES` associates `Traders` with the tungsten commodity node; and `HAS_RISK` links each `Country` to its dedicated `Geopolitical risk` node and the commodity node to its `Concentration risk` node. Presenting the schema directly from the database, rather than only as a conceptual diagram, demonstrates that the ontology was faithfully instantiated and provides a verifiable reference for the Cypher and Graph Data Science queries developed in the subsequent chapters.

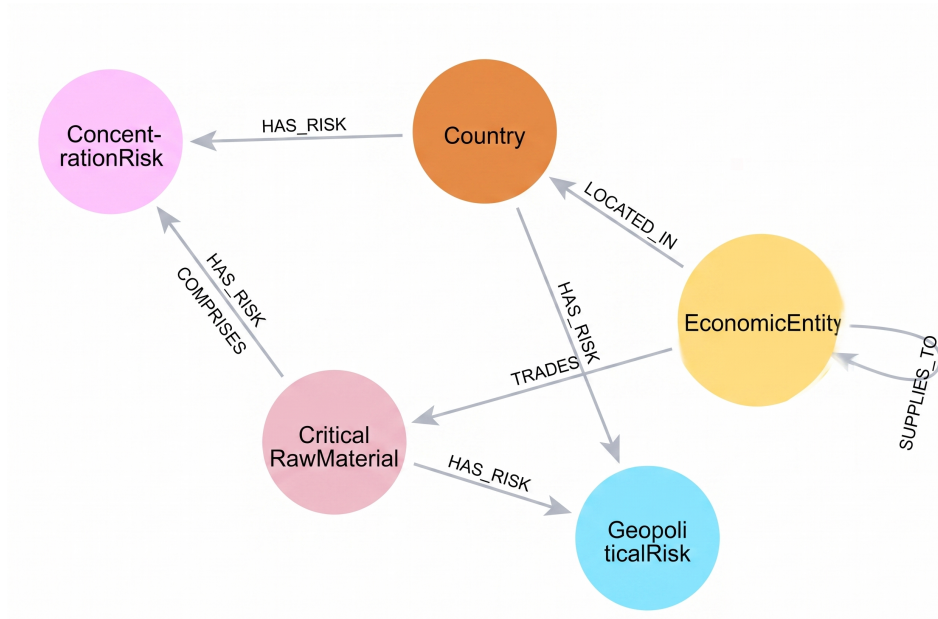


Figure 4.3. Database schema of the Strategic Digital Twin, rendered by the Neo4j `db.schema.visualization()` procedure. The diagram shows the five node labels (`Trader`, `Country`, `CriticalRawMaterial`, `ConcentrationRisk`, and `GeopoliticalRisk`) and the four directed relationship types (`SUPPLIES_TO`, `LOCATED_IN`, `TRADES`, and `HAS_RISK`) constituting the property-graph ontology.

The Two-Layer Architecture in Practice

The distinction between the Empirical Layer and the Synthetic Layer is reflected directly in the structure of the graph rather than assumed only at the conceptual level. The Empirical

Layer comprises the Country and Critical Raw Material nodes, whose attributes are derived from observed data sources; each Country node additionally records the provenance of its indicators through the `data_source` property (for example, UN Comtrade, World Bank WGI indicators, or EU JRC criticality assessments). The Synthetic Layer comprises the Trader nodes, which are generated through the Pareto-calibrated downscaling procedure described in Section 4.2. Within this layer, the boolean `is_anchor` property distinguishes the small set of real, publicly identifiable market leaders — seeded into the graph as anchor firms with their observed market shares — from the purely synthetic traders generated to populate the long tail of the firm-size distribution.

This separation is important for two reasons. First, it allows any downstream analysis to distinguish conclusions that rest on empirically grounded entities (Countries and anchor firms) from those that depend on synthetic approximations. Second, it provides a clear upgrade path for future research: as firm-level customs microdata or corporate disclosure records become available for specific countries or commodity corridors, individual synthetic Trader nodes can be replaced with empirically grounded entities without restructuring the surrounding graph topology.

4.4 Nodal Elasticity

In standard microeconomic theory, elasticity measures the responsiveness of an economic variable to changes in an underlying driver, typically under the assumption that agents can adjust through substitution and re-optimization mechanisms [30].

In highly concentrated critical raw material supply chains, as in our case, however, these adjustment mechanisms are often severely constrained in the short run. Limited substitutability, production rigidities, and long-term contractual dependencies restrict the ability of downstream actors to rapidly reconfigure sourcing patterns. As a result, the relevant analytical problem is not primarily how agents adjust to shocks, but rather how disruptions propagate through an existing network of trade dependencies.

To formalize this mechanism, the supply system is represented as a directed weighted network, where nodes correspond to countries or synthetic traders, and edges encode trade flows. Let $W \in \mathbb{R}^{n \times n}$ denote the matrix of flows, where W_{ij} represents the trade value flowing from node i to node j . Import dependence is captured by the column-normalized matrix $P \in \mathbb{R}^{n \times n}$ defined as:

$$P_{ij} = \frac{W_{ij}}{\sum_{m=1}^n W_{mj}}, \quad (4.4)$$

so that P_{ij} measures the share of imports of node j sourced from supplier i .

A key assumption of the methodology is that the analysis is conducted from a short-run perspective. In this setting, both trade dependencies and prices are assumed to remain fixed throughout the propagation process. More specifically, the dependency matrix P is held constant, implying that countries and firms do not immediately adjust their sourcing patterns in response to disruptions. Each node, therefore, continues to rely on the same suppliers, and the structure of import shares remains unchanged during the simulation horizon. This assumption is particularly appropriate for highly concentrated critical raw material supply chains such as tungsten, where alternative suppliers are limited and substitution is constrained by technological requirements, contractual rigidities, and limited

short-run production capacity. Prices are also assumed to remain constant following the shock. Consistent with the short-run propagation literature [31], changes in trade values are therefore interpreted primarily as changes in effective quantities rather than as endogenous price adjustments. This assumption is particularly plausible in the short run, where contractual rigidities, production constraints, and procurement frictions limit rapid price adjustment.

To operationalize disruption propagation within the supply network, we introduce an exogenous shock vector $s \in [0,1]^n$, where each component s_k represents the fraction of supply capacity lost by node k . Let $r \in [0,1]^n$ denote the vector of effective supply availability, and let $m \in [0,1]^n$ denote the residual supply capacity after the shock, defined as

$$m_j = 1 - s_j. \quad (4.5)$$

The propagation of the shock follows an iterative mechanism in which the availability of each node depends on the availability of its suppliers according to

$$r_j^{(t+1)} = m_j \left[(1 - \lambda) + \lambda \sum_{i=1}^n P_{ij} r_i^{(t)} \right], \quad (4.6)$$

where $\lambda \in [0,1]$ captures the intensity with which disruptions propagate across the network. Lower values of λ attenuate propagation effects, whereas $\lambda = 1$ correspond to full transmission through the dependency structure. Since direct empirical calibration of λ would require firm-level information on inventories, contractual flexibility, substitution capacity, and delivery lags, which are not available in the present setting, we treat λ as a scenario parameter. Varying λ allows us to evaluate how sensitive the results are to alternative assumptions regarding the strength of disruption propagation across the supply network.

The initial condition is given by $r_i^{(0)} = 1$ for all non-shocked nodes and $r_k^{(0)} = 1 - s_k$ for shocked nodes. The update is applied simultaneously across all nodes and is iterated until convergence. In practice, convergence is achieved when successive updates of the availability vector become arbitrarily small, i.e.,

$$\|r^{(t+1)} - r^{(t)}\|_\infty < \varepsilon, \quad (4.7)$$

for a given tolerance level $\varepsilon > 0$. The economic impact of a disruption is evaluated with respect to a target node j^* (e.g., Italy in our empirical investigation). Since the trade matrix is constructed from export-reported Comtrade flows, W_{ij} denotes the value exported by node i to node j . Baseline exposure of the target economy is therefore measured as the total value of incoming export flows received by the target node j^* :

$$M_{j^*}^{\text{baseline}} = \sum_{i=1}^n W_{ij^*}, \quad (4.8)$$

while post-shock imports are computed as

$$M_{j^*}(k; s_k) = \sum_{i=1}^n W_{ij^*} r_i, \quad (4.9)$$

where k denotes the shocked node and s_k the intensity of the associated disruption. The relative loss induced by the shock is then defined as

$$L_{j^*}(k; s_k) = \frac{M_{j^*}^{\text{baseline}} - M_{j^*}(k; s_k)}{M_{j^*}^{\text{baseline}}}. \quad (4.10)$$

Within the proposed framework, nodal elasticity is defined as

$$E_k = \frac{L_{j^*}(k; s_k)}{s_k}, \quad (4.11)$$

which measures the marginal sensitivity of the target node's imports to a supply disruption originating at node k . When evaluated for small perturbations (e.g., a 1% supply shock), E_k provides a finite-difference approximation of the derivative of the loss function with respect to the shock intensity, i.e., $E_k \approx \partial L_{j^*}(k; s_k) / \partial s_k$. Conversely, for large or full disruptions, it naturally becomes a stress-testing framework for scenario-based supply-chain analysis. Larger disruptions may generate non-linear propagation effects and network bottlenecks that are not fully captured by marginal elasticity alone. As a result, nodes with similar elasticity values under marginal perturbations may produce substantially different impacts under large-scale disruption scenarios.

The objective of the proposed stress-testing framework is not to reproduce the full complexity of market adjustment processes, but rather to provide a transparent and easily replicable methodology for assessing structural vulnerabilities in critical supply chains. While more sophisticated approaches may capture additional behavioural and market dynamics, they typically require substantially richer datasets and more extensive calibration procedures. The present framework instead focuses on isolating the propagation mechanisms embedded in the network structure while remaining applicable under limited data availability.

4.5 Proposed Multi-Tier & Multi-Commodity Conceptual Framework

To bridge the operational gap identified in the literature and fully address the regulatory mandates of Article 24 of the EU Critical Raw Materials Act (CRMA), this thesis proposes an original, generalized multi-tier and multi-commodity graph-based conceptual framework. While the practical baseline implemented in this study — hereafter referred to as *Case 1*, denoting a single-tier flat trader network populated with synthetic firms derived from UN Comtrade data — aggregates market entities into a single topological layer of Traders, the comprehensive theoretical design developed herein establishes a vertically integrated, semantic blueprint for a true strategic Digital Twin.

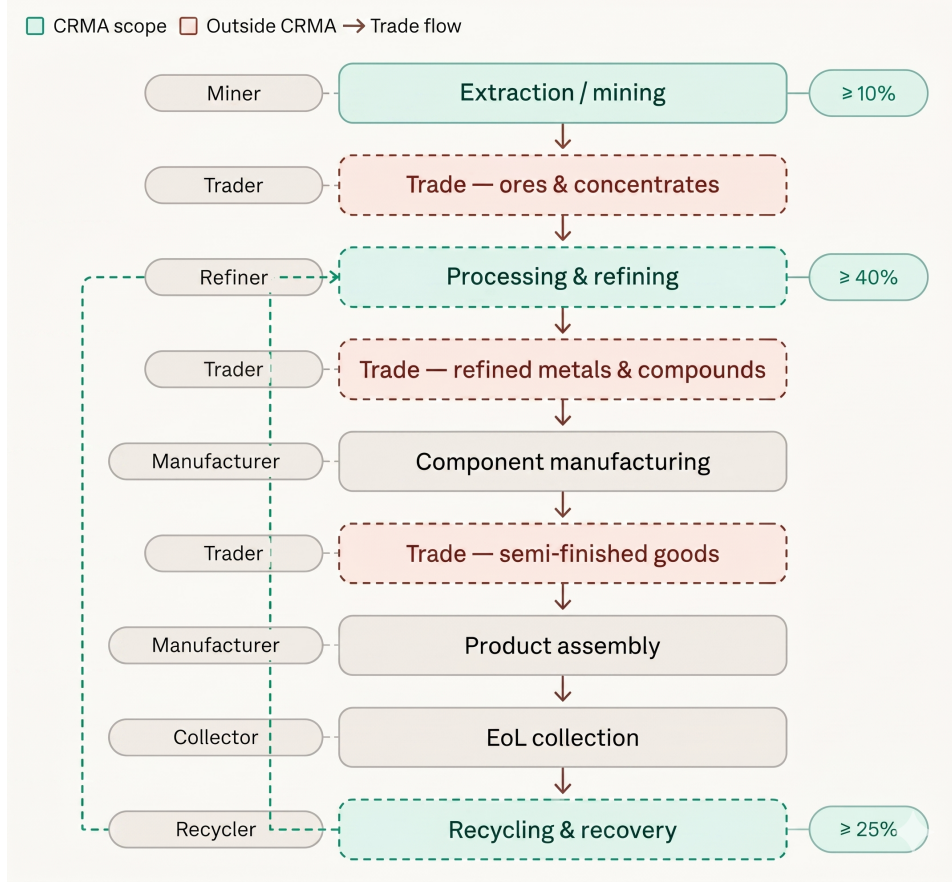


Figure 4.4. Main operational stages and regulatory benchmarks of the strategic raw material value chain under the EU CRMA framework. Percentage thresholds ($\geq 10\%$, $\geq 40\%$, $\geq 25\%$) refer to the domestic capacity targets established by Regulation (EU) 2024/1252, Article 5 [2].

As illustrated in Figure 4.4, a legally compliant monitoring system cannot treat supply chains as flat networks; it must explicitly segregate structural value-added stages and capture cross-commodity dependencies.

Multilayer Graph Formalization

We formalize the generalized supply chain ecosystem as a multi-commodity, multilayer directed graph $\mathcal{M}$. Let M be the set of all Critical Raw Materials (commodities) under analysis (e.g., $M = \{\text{Tungsten, Lithium, Cobalt, } \dots\}$). The global trading architecture is defined as:

$$\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$$

where:

- $\mathcal{L} = \{L_1, L_2, L_3, L_4, L_5\}$ represents the finite set of functional supply chain layers aligned with the CRMA taxonomy shown in Figure 4.4:

$$\begin{aligned} L_1 &= \text{Extraction / Mining} \\ L_2 &= \text{Processing / Refining} \\ L_3 &= \text{Component Manufacturing} \\ L_4 &= \text{Product Assembly (OEM)} \\ L_5 &= \text{Recycling / Recovery} \end{aligned}$$

- $\mathcal{V}$ is the set of heterogeneous nodes. A node $v \in \mathcal{V}$ is defined as a tuple $v = (i, L_\alpha)$, representing a distinct economic actor i operating strictly within layer L_α .
- $\mathcal{E}$ is the set of directed, intra-layer and inter-layer supply edges. An edge $e \in \mathcal{E}$ is formulated as $e = ((i, L_\alpha), (j, L_\beta), m)$, signifying a verified trade flow of a specific raw material $m \in M$ originating from actor i at layer L_α and destined for actor j at layer L_β .

Mathematical Modeling of Corporate Bill of Materials (BOM)

To solve the invisibility of raw materials inside industrial configurations, we mathematically integrate the micro-level corporate Bill of Materials (BOM) into the graph network. Let an assembler node $k \in L_4$ produce a set of finished products P_k . For a specific finished product $p \in P_k$, its internal configuration is governed by a structural BOM mapping matrix, defining the direct material requirements from Tier-1 suppliers:

$$C_p = \sum_{j \in L_3} \omega_{p,j} \cdot q_j$$

where $\omega_{p,j} \geq 0$ represents the physical or monetary volume of component j required per unit of finished product p , and $q_j \geq 0$ denotes the unit production capacity (or trade value) of component manufacturer $j \in L_3$. The scalar C_p therefore measures the total material intake of product p from all Tier-1 component suppliers, expressed in a consistent physical or monetary unit. Crucially, the component manufacturer $j \in L_3$ possesses its own backward-looking dependency profile, which recursively maps through processing nodes (L_2) back to extraction nodes (L_1).

Therefore, the total mass or value contribution Q of a strategic raw material $m \in M$ extracted by mining node $i \in L_1$ that is ultimately absorbed by the finished product p is derived via the multi-hop network path traversal:

$$Q(i \rightarrow p|m) = \sum_{\pi \in \Pi(i,p)} \prod_{e \in \pi} P_e(m)$$

where $\Pi(i,p)$ denotes the set of all active directed supply paths crossing layers $L_1 \rightarrow L_2 \rightarrow L_3 \rightarrow L_4$, and $P_e(m)$ represents the column-normalized import dependence share over edge e for material m , as structurally formalized in Equation (4.4). Since each $P_e(m) \in [0,1]$ and longer chains accumulate products of fractions, $Q(i \rightarrow p | m)$ naturally decreases with path length. This is by design: the quantity measures the *relative*

structural dependency intensity of finished product p on mining node i for material m , not an absolute physical mass. A value close to 1 indicates near-total dependence on a single upstream source across all intermediate layers. In contrast, values close to 0 reflect either negligible bilateral dependency or attenuation through multi-hop intermediaries.

Generalized Cross-Commodity Risk Multipliers

The theoretical expansion to a multi-commodity framework introduces structural risk dependencies. A node j at layer L_3 might require multiple critical raw materials simultaneously (e.g., Tungsten for tooling and Cobalt for coatings) to manufacture a single strategic component. If any single material flow is disrupted, the entire node’s operational capacity collapses.

To model this systemic vulnerability, we define the Composite Structural Fragility Index (SF_j) for any downstream entity as a non-linear combination of its material-specific exposures:

$$SF_j = 1 - \prod_{m \in M} \left(1 - \sum_{i \in \mathcal{V}_{upstream}} P_{ij}(m) \cdot \hat{R}_i \right)$$

where $\hat{R}_i \in [0,1]$ denotes the normalized risk score of upstream supplier i , obtained by linearly rescaling either the Geopolitical Risk or Concentration Risk indicator to the unit interval, and $P_{ij}(m)$ is the column-normalized import dependence share for commodity m as defined in Equation (4.4), satisfying $\sum_i P_{ij}(m) = 1$ for each node j . Since both $P_{ij}(m)$ and $\hat{R}_i$ lie in $[0,1]$, the inner sum is bounded in $[0,1]$, ensuring that each factor in the product remains non-negative and the index SF_j is well-defined on $[0,1]$. Intuitively, SF_j approaches 1 when the node is heavily exposed to high-risk suppliers across multiple commodities simultaneously, and approaches 0 when all upstream risk is negligible.

Implementation Scope and Future Work Alignment

This comprehensive multilayer, multi-commodity mathematical formalization establishes the ideal architecture for tracking systemic vulnerabilities. However, due to severe data scarcity in downstream corporate ownership records and the tight timeline, the practical implementation within this thesis focuses strictly on a single material ($m = \text{Tungsten}$) and aggregates the structural interactions of layers L_1 through L_4 into a unified, flat transaction environment governed by the baseline model (*Case 1*).

Consequently, the empirical network simulation and the calculation of Nodal Elasticity profiles presented in Chapter 6 and Chapter 7 reflect this localized baseline. The software engineering execution of the full multi-tier path traversal matrix and the automated extraction of multi-commodity composite fragilities are formally deferred as primary objectives for *Future Work* (Section 8.4).

Chapter 5

Case Study: Tungsten

This chapter operationalizes the methodological framework developed in Chapter 4 by deploying the global tungsten supply chain as an empirical pilot case study. The chapter is organized into four sections. Section 5.1 reviews the physical, industrial, and strategic properties that make tungsten a particularly demanding material to substitute and a high-priority commodity within the European Union Critical Raw Materials Act (CRMA). Section 5.2 maps the full tungsten value chain and its commodity classification, justifying the decision to focus the empirical analysis on a specific set of Harmonized System (HS) codes. Section 5.3 characterizes the geopolitical and firm-level structure of the global tungsten market, motivating its selection as an ideal stress-testing case. Finally, Section 5.4 describes how the 2024 trade ecosystem is reconstructed, calibrated, and materialized as a property graph within the Neo4j instance that underlies the analysis presented in the following chapters.

5.1 Tungsten: Industrial Applications and Strategic Criticality

Tungsten (chemical symbol W, atomic number 74) is a refractory transition metal whose exceptional physical and chemical properties make it extremely difficult to substitute in a wide range of high-performance applications. It possesses the highest melting point of all metals (approximately 3422°C), the highest tensile strength at elevated temperatures, extreme hardness, high density, and excellent thermal and electrical conductivity. These properties are not merely incremental advantages over alternative materials: in many critical applications they are enabling characteristics for which no technically equivalent substitute exists without a substantial loss of performance.

The dominant industrial use of tungsten is in the production of cemented carbides — also known as hard metals — in which tungsten carbide (WC) particles are bonded within a metallic matrix, typically cobalt. Cemented carbides alone account for the majority of global tungsten consumption and are indispensable for cutting tools, drilling equipment, wear-resistant components, and metal-forming dies that underpin virtually every advanced manufacturing sector. The remaining demand is distributed across tungsten mill products (wire, rod, and sheet used in lighting, electronics, and electrical contacts), tool and

high-speed steels, superalloys for aerospace turbine components, and a range of chemical applications including catalysts and pigments. Beyond civilian manufacturing, tungsten occupies a position of acute strategic relevance in the defense and aerospace domains, where its density and hardness make it a material of choice for armor-piercing munitions, kinetic-energy penetrators, ballast and counterweights, and radiation-shielding components.

This combination of indispensability and supply concentration explains why tungsten is classified by the European Commission as both a *Critical* and a *Strategic* raw material under Regulation (EU) 2024/1252 [2]. As formalized in Chapter 2, the criticality of a material is jointly determined by its Economic Importance (*EI*) and its Supply Risk (*SR*). Tungsten scores highly along both axes: its embedding in high-value-added manufacturing trajectories drives a high *EI*, while its supply risk is amplified by three mutually reinforcing factors — extreme production concentration in a single dominant country, a low end-of-life recycling input rate relative to total demand, and a high net import reliance for the European Union, which possesses negligible domestic primary extraction capacity. The limited substitutability of tungsten further exacerbates its risk profile: although substitution is technically conceivable in selected applications (for example, certain ceramics or molybdenum-based alloys), such alternatives typically entail measurable penalties in performance, durability, or cost, and are therefore not viable at scale in the most demanding end-uses. Tungsten thus represents a paradigmatic example of a material for which country-level diversification strategies alone are insufficient, and for which a finer-grained, firm-level understanding of supply dependencies is required.

5.2 The Tungsten Value Chain and Commodity Classification

The transformation of tungsten from raw ore into finished industrial components proceeds through a sequence of distinct value-added stages, each characterized by a markedly different technological content and economic value density. Figure 5.1 illustrates the complete value chain, from primary mining and ore beneficiation through to finished products and end-use, together with the associated HS commodity classifications and the recycling loop that feeds secondary material back into the intermediate processing stages.

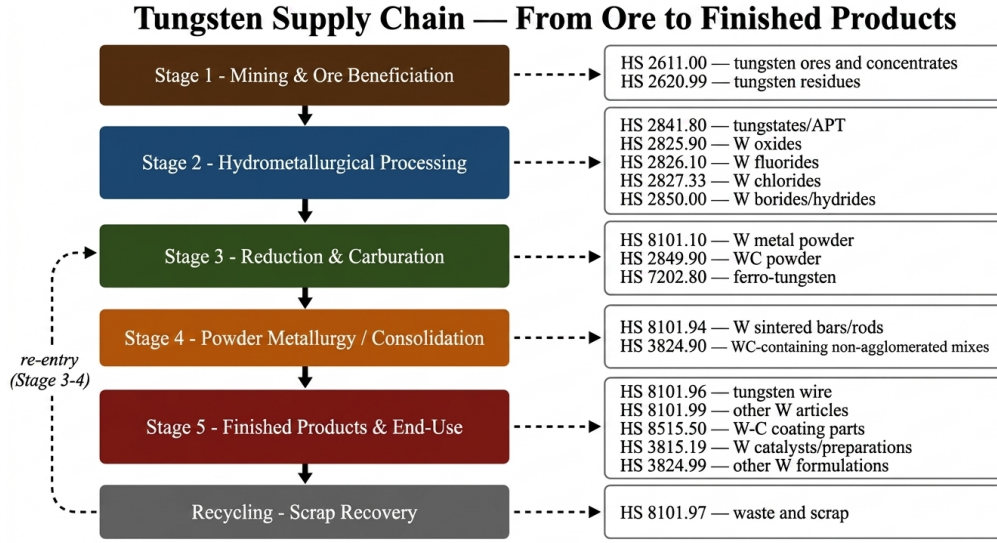


Figure 5.1. Main stages of the tungsten value chain, from ore extraction to downstream finished products, and the associated HS classifications considered in the analysis. The dashed re-entry arrow indicates the reintegration of recycled scrap into the reduction and consolidation stages.

The first stage, *Mining and Ore Beneficiation*, covers the extraction and primary concentration of tungsten-bearing ores, predominantly wolframite and scheelite, captured under HS 2611.00 (tungsten ores and concentrates) and HS 2620.99 (tungsten-bearing residues). The second stage, *Hydrometallurgical Processing*, transforms these concentrates into intermediate chemical compounds — most notably ammonium paratungstate (APT, HS 2841.80), together with tungsten oxides, fluorides, chlorides, and borides. The third stage, *Reduction and Carburation*, yields metallic tungsten powder (HS 8101.10), tungsten carbide powder (HS 2849.90), and ferro-tungsten (HS 7202.80). The fourth stage, *Powder Metallurgy and Consolidation*, produces sintered bars and rods (HS 8101.94) and carbide-containing mixtures, while the fifth stage, *Finished Products and End-Use*, delivers tungsten wire (HS 8101.96), other tungsten articles (HS 8101.99), and a range of coating, catalyst, and specialty formulations. A dedicated recycling and scrap-recovery channel (HS 8101.97) reintroduces secondary material into the reduction and consolidation stages, partially offsetting primary demand.

A defining feature of this value chain is that economic value density increases sharply along the processing trajectory. A unit mass of finished tungsten articles embodies substantially more economic value than the equivalent mass of raw ore or concentrate, reflecting the cumulative technological transformation applied at each stage. This property has a direct and consequential implication for supply-chain monitoring: indicators and stress-testing exercises based on traded weight or physical quantity systematically understate the strategic importance of the downstream segments, which account for a comparatively limited physical volume but a disproportionately large share of economic value. This same logic applies to the geographic structure of trade: a country's apparent weight in tungsten flows can shift substantially depending on whether trade is measured by physical mass or

by economic value, since high-value-added downstream products are not distributed in the same proportions as raw tonnage. Adopting a value-based representation of trade flows, as discussed in Chapter 4, is therefore essential to preserve the heterogeneity in technological content and strategic relevance across the different stages of refinement.

Given the breadth of the full classification, the empirical analysis in this thesis focuses on three HS codes that jointly span the mid-to-downstream metallic processing stages of the chain: HS 8101.10 (tungsten metal powder, Stage 3), HS 8101.94 (unwrought and sintered tungsten, Stage 4), and HS 8101.99 (other tungsten articles, Stage 5). This selection is motivated by three considerations. First, these stages concentrate the highest value density and the greatest strategic content, which is precisely where weight-based indicators are least informative. Second, these metallic-product codes exhibit the most reliable and complete export reporting within the Comtrade dataset, as discussed in Section 4.1. Third, restricting the analysis to a coherent band of adjacent processing stages preserves the structural comparability of the reconstructed trade flows, avoiding the distortions that would arise from mixing chemically and economically heterogeneous product categories. Table 5.1 summarizes the mapping between the value-chain stages, the corresponding HS codes, and their inclusion in the empirical analysis.

Value-Chain Stage	HS Code	Product Description	In Analysis
Mining & Beneficiation	2611.00	Tungsten ores and concentrates	—
	2620.99	Tungsten-bearing residues	—
Hydrometallurgical Processing	2841.80	Tungstates / APT	—
	2825.90	Tungsten oxides	—
Reduction & Carburization	8101.10	Tungsten metal powder	✓
	2849.90	Tungsten carbide (WC) powder	—
	7202.80	Ferro-tungsten	—
Powder Metallurgy / Consolidation	8101.94	Unwrought / sintered tungsten	✓
Finished Products & End-Use	8101.96	Tungsten wire	—
	8101.99	Other tungsten articles	✓
Recycling / Scrap Recovery	8101.97	Tungsten waste and scrap	—

Table 5.1. *Mapping between tungsten value-chain stages, representative HS commodity codes, and their inclusion in the empirical trade-network reconstruction. The three codes retained for analysis span the mid-to-downstream metallic processing stages, where economic value density and strategic content are highest.*

5.3 Market Structure and Geopolitical Concentration

The global tungsten market is among the most geographically concentrated of all critical raw materials. China alone accounts for approximately 80% of world mine production and over 90% of global refining capacity, a dominance documented in successive editions of the USGS Mineral Commodity Summaries [4] and the European Commission Joint Research Centre raw-materials assessments [5]. Beyond primary extraction, China also imports substantial volumes of tungsten ore and concentrate for further processing and subsequent re-export, thereby extending its control over the higher-value-added downstream segments of the chain. This structural position is only partially visible in conventional weight-based trade statistics, which, as noted above, understate the strategic weight of the refined products in which Chinese dominance is most pronounced.

Crucially, concentration in the tungsten market is not confined to the country level: it is reproduced, and in some respects intensified, at the firm level. The market is dominated by a small number of vertically integrated Chinese groups that operate across multiple stages of the value chain simultaneously — from extraction and refining through to advanced manufacturing. The most prominent among these include Xiamen Tungsten Co., Ltd. and China Tungsten & High-tech Materials Co., alongside other major actors such as China Minmetals Corporation, the Jiangxi Tungsten Industry Group, and CITIC Tungsten & Molybdenum [65]. The vertical integration of these groups means that a disruption affecting a single corporate entity can propagate simultaneously across several processing stages, a form of systemic fragility that country-level indicators are structurally incapable of capturing.

This concentration has been further reinforced by recent policy developments. Over the 2025–2026 period, the Chinese Ministry of Commerce (MOFCOM) introduced a series of export-control measures targeting tungsten and related critical minerals [33], tightening the conditions under which strategic tungsten products may be exported. These measures have effectively consolidated the already dominant position of the leading Chinese groups and heightened the exposure of downstream European economies to potential supply shocks. The combination of extreme upstream concentration, deep vertical integration at the firm level, and an evolving regulatory environment makes tungsten an exceptionally informative case study for the purposes of this thesis: it is precisely the kind of supply chain in which aggregate country-level metrics offer a misleadingly reassuring picture, and in which a firm-level, graph-based analysis is most likely to reveal otherwise hidden structural vulnerabilities.

5.4 The Global Tungsten Trade Ecosystem

To illustrate the proposed methodology, the Strategic Digital Twin (SDT) was populated with a synthetic reconstruction of the global tungsten market for the reference year 2024. Bilateral, export-reported trade flows for the three selected HS codes (8101.10, 8101.94, and 8101.99) were extracted from the UN Comtrade database and processed according to the granular downscaling methodology formalized in Section 4.2. The global supplier-side Herfindahl–Hirschman Index (HHI), computed from these Comtrade export values, is 0.157, with a top-supplier (China) share of approximately 31.7% and a three-firm concentration

ratio (CR3) of approximately 60.4%. These figures describe a *moderately* concentrated downstream export market, in which a single dominant supplier coexists with a broad tail of secondary re-exporting economies. Importantly, this moderate export-value concentration stands in sharp contrast to the upstream mining and refining stages, where China alone accounts for roughly 80% of world production [4, 5]. This divergence is itself analytically significant: the aggregate export HHI conveys a reassuringly moderate picture, even though the underlying market is dominated upstream by a limited number of vertically integrated Chinese groups and is further exposed to the consolidating effect of the MOFCOM export restrictions [33]. It is precisely this gap — between a moderate aggregate indicator and the structural vulnerabilities concealed beneath it — that motivates the firm-level, graph-based analysis developed in this thesis.

Consistent with the theory of economic granularity [22, 32], this concentration is modeled through a fat-tailed firm-size distribution. The observed global HHI is used to calibrate the Pareto shape parameter α , which in turn governs the synthetic firm-size structure generated within each exporting country. The reconstruction proceeds through the following steps, detailed in Chapter 4:

1. **Market Concentration Estimation:** The global supplier-side HHI is computed from the Comtrade export data for tungsten.
2. **Concentration Calibration:** The observed global HHI is used as a calibration target to determine the Pareto shape parameter α governing the synthetic firm population.
3. **Stochastic Sampling:** For each country, N_c firm shares are sampled from a Pareto distribution,

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, \quad (5.1)$$

where α controls the concentration of the firm-size distribution and x_m denotes the minimum scale parameter.

4. **Rescaling for Consistency:** The generated firm shares are rescaled so that their sum reproduces the observed aggregate export value of each country, ensuring mass-balance consistency with the original Comtrade data.

This procedure preserves the Pareto-based concentration structure observed at the global level while remaining strictly consistent with the aggregate trade statistics for tungsten. The economic impact of supply disruptions is evaluated with respect to a designated target economy, which in this study is the Italian industrial system. Italy constitutes a particularly suitable target node for the analysis: its manufacturing base is deeply integrated into European advanced mechanical engineering, tooling, and precision-component supply chains that rely heavily on tungsten-based cemented carbides, while Italy possesses no domestic primary tungsten extraction capacity and is therefore structurally dependent on imported metallic tungsten products. The Italian case thus exemplifies the broader European condition of high downstream exposure combined with negligible upstream control.

Once reconstructed, the synthetic firm-level network — comprising countries, synthetic traders, the tungsten commodity node, and the associated risk entities — was materialized

as a Labeled Property Graph within the local Neo4j instance through the Python ingestion pipeline described in Section 4.1. With the graph fully populated and validated, most of the subsequent analytical operations such as the centrality computations and geopolitical risk scoring, were executed directly on this graph using Cypher and the Graph Data Science library. The quantitative characterization of the resulting network is presented in Chapter 6.

Chapter 6

Quantitative Analysis of the Tungsten Supply Network

Having reconstructed and materialized the synthetic tungsten trade network as a Labeled Property Graph in Neo4j (Chapter 5), this chapter subjects that network to a systematic quantitative analysis. The objective is twofold: first, to characterize the global structural properties of the reconstructed graph and confirm that the downscaling procedure has produced a topology consistent with the oligopolistic nature of the tungsten market; and second, to identify the individual trading entities that carry systemic importance, using the family of Graph Data Science (GDS) metrics formalized in Section 2.3. The chapter is organized as follows. Section 6.1 presents the baseline structural statistics of the graph, including density, connectivity, and the degree distribution. Section 6.2 analyzes structural importance through weighted Betweenness and PageRank centrality, both individually and jointly. Section 6.3 introduces the Geopolitical Risk Exposure dimension, and Section 6.4 synthesizes the three metrics through a correlation analysis and a country-level aggregation. All centrality computations were executed directly on the in-memory GDS projection of the trader-level subgraph, using the inverse-value weighting scheme ($w_{ij} = 1/\text{value}_{ij}$) for shortest-path metrics described in Section 2.3.¹

6.1 Network Statistics of the Reconstructed Tungsten Graph

Before applying any analytical metric, it is essential to characterize the scale and structural complexity of the reconstructed network. These baseline statistics validate that the synthetic downscaling procedure has generated a graph of realistic size and connectivity, and they provide the reader with a concrete sense of the object on which the subsequent

¹A single synthetic entity associated with an unspecified-origin trade area (Comtrade code «XX») is excluded from the centrality visualizations and rankings presented in this chapter, since it does not correspond to an identifiable sovereign jurisdiction. It is retained in the underlying graph but omitted from the firm-level interpretation, leaving 191 attributable traders.

analysis operates. Table 6.1 summarizes the node and relationship composition of the full Strategic Digital Twin, while Table 6.2 reports the distribution of synthetic traders across the most represented countries.

Statistic	Value	Interpretation
Total nodes	502	Traders, Countries, risk nodes, and the CRM node
of which Trader nodes	192	Synthetic firm-level trading entities
of which Country nodes	154	Sovereign states represented in the network
of which GeopoliticalRisk nodes	154	One risk node per country
of which CriticalRawMaterial nodes	1	Tungsten as the single modeled commodity
of which ConcentrationRisk nodes	1	Global HHI-based concentration risk node
Total relationships	6,107	All directed edges across all relationship types
SUPPLIES_TO	5,568	Bilateral synthetic trade flows between Traders
LOCATED_IN	192	Trader-to-Country membership edges
TRADES	192	Trader-to-CRM activity edges
HAS_RISK	155	Country and CRM to risk node edges

Table 6.1. *Baseline composition of the reconstructed tungsten Strategic Digital Twin graph.*

Country (ISO3)	Number of Synthetic Traders
CHN	12
USA	8
JPN	5
MEX	4
DEU	3
KOR	3
CAN	3
GBR	2
IND	2
ISR	2
All others	148
Total	192

Table 6.2. *Distribution of synthetic Trader nodes across the ten most represented countries in the tungsten supply network.*

To convey the concrete structure of the populated graph beyond aggregate counts, Figure 6.1 shows a representative subgraph rendered directly in the Neo4j Desktop. Each yellow node denotes a synthetic Trader and each orange node a Country, connected through `LOCATED_IN` membership edges. The visualization makes the firm-level disaggregation immediately tangible: the majority of countries are represented by a single trader and therefore appear as isolated trader–country pairs, whereas the dense star-shaped cluster corresponds to a single country hosting a large number of synthetic traders. This cluster is the visual signature of the Pareto-calibrated downscaling applied to the most concentrated producing economies — most prominently China, which is populated with twelve synthetic traders, consistent with the country-level distribution reported in Table 6.2. The rendering thus exposes, at the level of the raw graph, the same oligopolistic concentration that the centrality and elasticity analyses subsequently quantify.

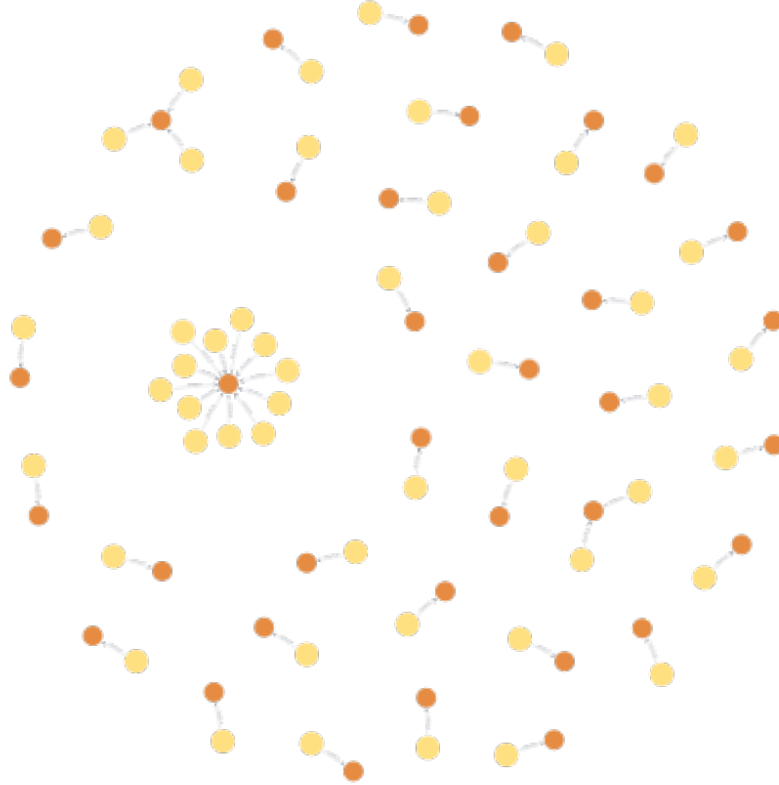


Figure 6.1. *Sample subgraph of the reconstructed tungsten network, rendered in the Neo4j Desktop. Yellow nodes are synthetic Traders and orange nodes are Countries, linked by `LOCATED_IN` edges. Most countries host a single trader (isolated pairs), while the dense star-shaped cluster represents a single highly concentrated country populated with multiple synthetic traders.*

Beyond these compositional counts, Table 6.3 reports a set of structural metrics that characterize the connectivity and cohesion of the trader-level subgraph. The network comprises 192 trader nodes connected by 5,568 directed `SUPPLIES_TO` edges, yielding a directed density of approximately 0.152. This is a high density for a supply-chain graph: on average, each trader maintains active supply relationships with roughly 29 counterparties, both as a supplier and as a recipient. This dense interconnection reflects the re-export-intensive nature of the global tungsten trade, in which dominant upstream processors distribute material to a broad and overlapping set of downstream importers across many geographies.

Structural Metric	Value
Trader nodes	192
Directed SUPPLIES_TO edges	5,568
Directed density	0.152
Average degree (in = out)	29.0
Maximum out-degree	128
Maximum in-degree	66
Weakly connected components	1 (size 192)
Strongly connected components	98
Largest strongly connected component	94
Singleton components	96

Table 6.3. *Structural connectivity metrics of the trader-level subgraph.*

Three structural observations emerge. First, the network forms a single weakly connected component encompassing all 192 traders: ignoring edge direction, every trader is reachable from every other, confirming that the downscaling procedure has not produced isolated sub-markets. Second, and in sharp contrast, the strong-connectivity structure is markedly polarized. The largest strongly connected component contains 94 traders — roughly one half of the network — while the remaining traders are distributed across 96 singleton components and a single component of size two. A strongly connected component is a set of nodes within which every member can reach every other following edge directions; its existence here indicates a cohesive trading «core» of approximately 94 mutually reachable intermediaries, surrounded by a large directed periphery of traders that are either pure recipients (importers) or pure originators (exporters) and therefore cannot be returned to once left. This core–periphery asymmetry is a direct structural manifestation of the oligopolistic market organization imposed by the Pareto-calibrated downscaling.

Third, the degree structure is strongly heterogeneous. While the average degree is 29, the maximum out-degree reaches 128 and the maximum in-degree reaches 66, revealing the presence of a small number of hyper-connected exporters that supply a disproportionately large fraction of the network. This heterogeneity is visualized in Figure 6.2, which plots the total degree of each trader against its rank on doubly logarithmic axes. The approximately linear decay over the upper portion of the distribution is the signature of a heavy-tailed, power-law-like degree distribution, consistent with the theory of economic granularity [22, 32] and with the fat-tailed firm-size structure deliberately induced by the Pareto sampling described in Section 4.2. The empirical recovery of this heavy tail constitutes an internal validation of the synthetic generation procedure: the firm-level topology reproduces the concentration regularities expected of a real oligopolistic commodity market.

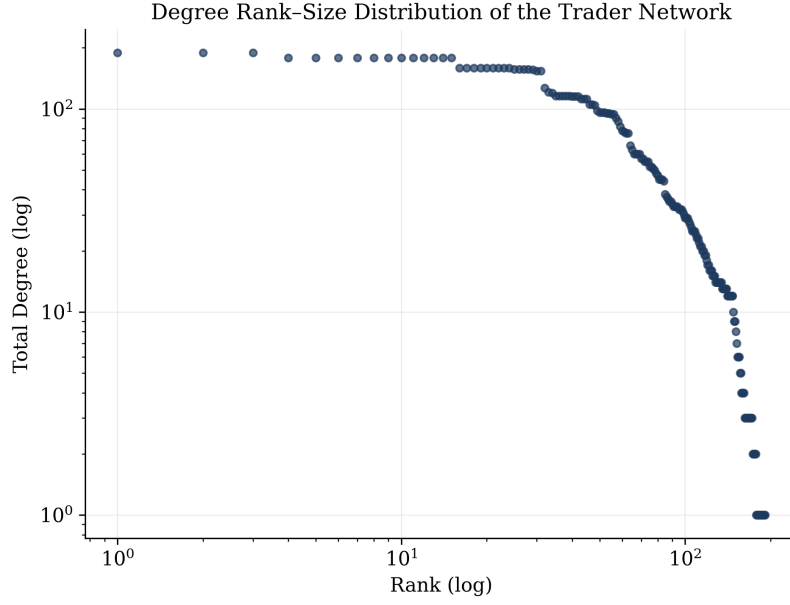


Figure 6.2. *Degree rank-size distribution of the trader network on doubly logarithmic axes. The heavy upper tail is consistent with a power-law-like firm-size structure induced by the Pareto-calibrated downscaling procedure.*

A multi-scale visualization of the reconstructed network is presented in Figure 6.3, which combines aggregate country-level structure with the surrounding synthetic trader-level clusters generated through the downscaling procedure. Countries constitute the macro-level backbone of the network, while the peripheral clusters represent the firm-level entities used to reconstruct granular supply-chain interactions within the Strategic Digital Twin.

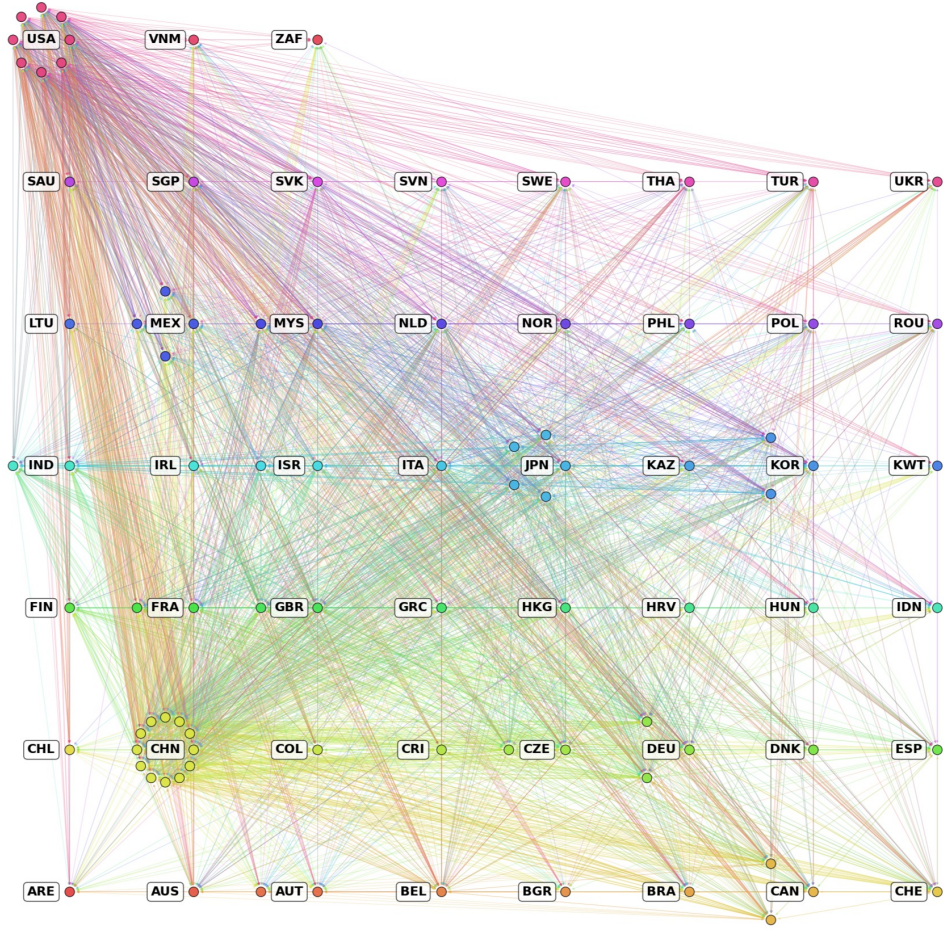


Figure 6.3. *Reconstructed tungsten supply-chain network with synthetic trader-level entities.*

6.2 Structural Centrality Analysis

To analyze structural importance and systemic vulnerability within the tungsten supply chain, the network is modeled as a directed, weighted trader-level graph in which nodes represent individual trading entities and weighted edges encode the value of supply relationships. The analysis relies on two complementary centrality metrics, drawn from the taxonomy of network control measures of [34]: weighted Betweenness centrality, which captures brokerage and routing control, and PageRank, which captures recursive systemic importance. These two metrics are first examined individually and then jointly, since their divergence is itself analytically informative.

6.2.1 Computing Network Metrics with the Graph Data Science Library

All structural metrics reported in this chapter were computed directly within Neo4j using the GDS library, operating on an in-memory projection of the trader-level subgraph. This subsection documents the exact queries used, so that the analysis is fully transparent and reproducible. The queries are presented in the order in which they are executed.

The first step prepares the edge weights and creates the in-memory projection. Because weighted Betweenness centrality interprets edge weights as traversal costs, the trade value is inverted ($w_{ij} = 1/\text{usd}_{ij}$) so that high-value relationships are treated as low-resistance, short-distance connections, as motivated in Section 2.3. The projection includes only `Trader` nodes and the directed `SUPPLIES_TO` relationships, carrying both the raw trade value (`usd`) and the inverse weight.

Listing 6.1. Inverse-weight preparation and creation of the in-memory GDS projection of the trader-level subgraph.

```
// Inverse weight: high trade value => low traversal cost
MATCH (:Trader)-[r:SUPPLIES_TO]->(:Trader)
WHERE r.usd IS NOT NULL AND r.usd > 0
SET r.inv_weight = 1.0 / r.usd;

// Project the trader-level subgraph into GDS
CALL gds.graph.project(
  'tungsten_traders',
  'Trader',
  {
    SUPPLIES_TO: {
      orientation: 'NATURAL',
      properties: ['usd', 'inv_weight']
    }
  }
);
```

Weighted Betweenness centrality is then computed over this projection using the inverse-value weight. The query streams each trader together with its sovereign location and betweenness score, ordered from the most to the least central broker.

Listing 6.2. Weighted Betweenness centrality using the inverse-value edge weight.

```
CALL gds.betweenness.stream('tungsten_traders', {
  relationshipWeightProperty: 'inv_weight'
})
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).company_name AS trader,
       gds.util.asNode(nodeId).country_iso AS country,
       score AS betweenness
ORDER BY betweenness DESC;
```

PageRank is computed on the same projection, but weighted by the raw trade value (`usd`) rather than its inverse, since PageRank propagates importance along the direction

and strength of the dependency relationships. The standard damping factor of 0.85 is used, consistent with the formulation in Section 2.3.

Listing 6.3. Value-weighted PageRank centrality with damping factor 0.85.

```
CALL gds.pageRank.stream('tungsten_traders', {
  relationshipWeightProperty: 'usd',
  dampingFactor: 0.85,
  maxIterations: 100,
  tolerance: 0.0000001
})
YIELD nodeId, score
RETURN gds.util.asNode(nodeId).company_name AS trader,
       gds.util.asNode(nodeId).country_iso AS country,
       score AS pagerank
ORDER BY pagerank DESC;
```

Finally, the Geopolitical Risk Exposure Index (GRI) is computed in Cypher by combining each trader’s normalized PageRank with the geopolitical risk score of its sovereign location. The PageRank values, previously written back to the nodes via `gds.pageRank.write`, are min-max normalized across all traders and combined with the country-level risk score `geo_risk_score` using the equal-weight scheme ($w = 0.5$) discussed in Section 6.3.

Listing 6.4. Geopolitical Risk Exposure Index combining normalized PageRank with country-level geopolitical risk.

```
MATCH (t:Trader)
WITH max(t.pagerank) AS pr_max, min(t.pagerank) AS pr_min
MATCH (t:Trader)-[:LOCATED_IN]->(c:Country)
WITH t, c, pr_max, pr_min,
      (t.pagerank - pr_min) / (pr_max - pr_min) AS pr_norm,
      c.geo_risk_score AS country_risk
RETURN t.company_name AS trader,
       c.iso3 AS country,
       0.5 * pr_norm + 0.5 * country_risk AS GRI
ORDER BY GRI DESC;
```

The same in-memory projection also underlies the network-level statistics reported in Section 6.1: the weakly and strongly connected components are obtained through `gds.wcc` and `gds.scc` respectively, and the degree distribution is extracted directly through Cypher aggregation over the `SUPPLIES_TO` relationships.

6.2.2 Weighted Betweenness Centrality

As formalized in Section 2.3, Betweenness centrality measures the extent to which a node acts as a broker by quantifying the proportion of all shortest paths in the network that pass through it. In the supply-chain context, material does not traverse the network at random but follows cost-minimizing procurement routes; a trader with high betweenness therefore occupies a structural bottleneck through which a large share of these optimal routes must pass. To align the topological notion of a shortest path with economic reality, edge weights are defined as the inverse of trade value, $w_{ij} = 1/\text{value}_{ij}$, so that high-value relationships are treated as low-resistance, short-distance connections. Under this

formulation, a trader exhibiting high weighted betweenness is one whose removal would force downstream economies to redirect procurement through fragmented, higher-resistance channels.

Figure 6.4 reports the twenty traders with the highest weighted betweenness scores. The most striking feature of the result is the extreme concentration of brokerage. Of the 192 synthetic traders, only 42 exhibit a non-zero betweenness score; the remaining 149 lie on no shortest path between any other pair of nodes and therefore play no brokerage role whatsoever. Furthermore, the five highest-ranked traders alone account for approximately 68% of the total betweenness mass in the network. The leading brokers are DEU_Trader_01 (betweenness $\approx 8,732$), USA_Trader_01 ($\approx 7,545$), Xiamen Tungsten Co., Ltd. ($\approx 6,581$), JPN_Trader_05 ($\approx 4,670$), and CAN_Trader_03 ($\approx 3,736$). These entities function as the critical structural intermediaries that connect otherwise weakly linked regions of the global trade network. Crucially, this ranking is not a simple reflection of trade volume: German and U.S. intermediaries dominate the brokerage hierarchy despite China's overwhelming share of aggregate tungsten production, indicating that structural control over routing is exercised primarily through brokerage position rather than through raw export tonnage.

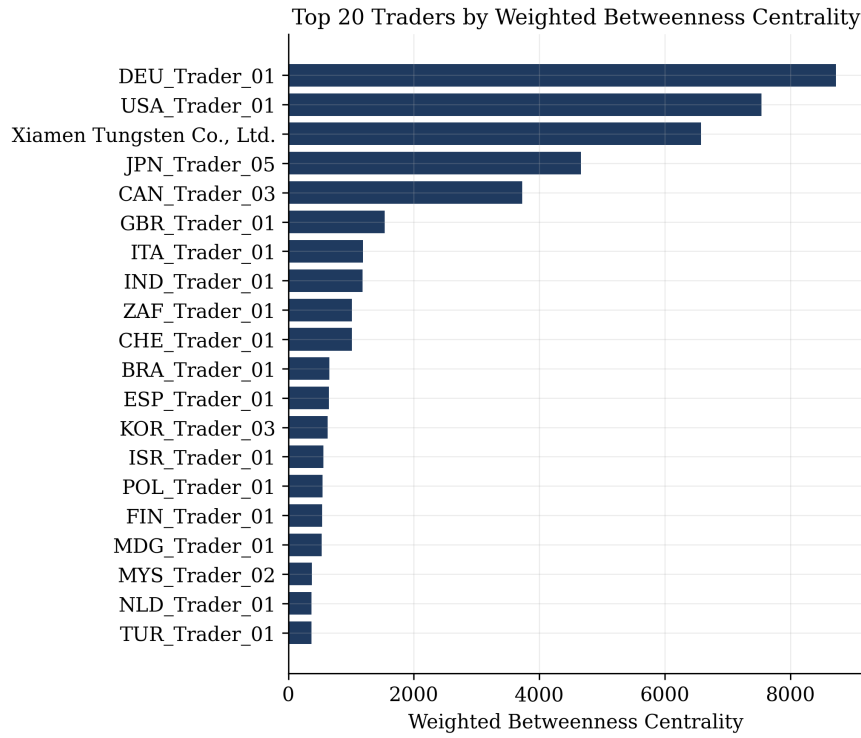


Figure 6.4. *Top twenty traders by weighted Betweenness centrality. Only 42 of the 192 traders record a non-zero score, and the five leading brokers account for roughly 68% of total betweenness.*

6.2.3 PageRank Centrality

Whereas betweenness captures control over routing, PageRank measures recursive systemic importance. As described in Section 2.3, a node’s PageRank score is determined not merely by how many partners point to it, but by how central those partners themselves are: a trader is systemically important when it supplies entities that are themselves embedded deeply within the network. The algorithm was executed to convergence on the value-weighted graph with the standard damping factor of 0.85.

Figure 6.5 presents the twenty highest-ranked traders by PageRank. The leading entities are USA_Trader_01 (PageRank ≈ 3.42), CAN_Trader_03 (≈ 2.58), DEU_Trader_01 (≈ 2.50), MEX_Trader_02 (≈ 2.34), and USA_Trader_08 (≈ 2.25). The PageRank distribution is considerably less skewed than the betweenness distribution — scores range continuously from about 0.15 to 3.42, with a mean of roughly 0.42 — reflecting the fact that recursive importance is distributed more broadly across the dense trading core than brokerage control, which is confined to a handful of bottleneck nodes. United States, Canadian, German, and selected Asian traders are particularly prominent, owing to their embedding within highly interconnected trading ecosystems.

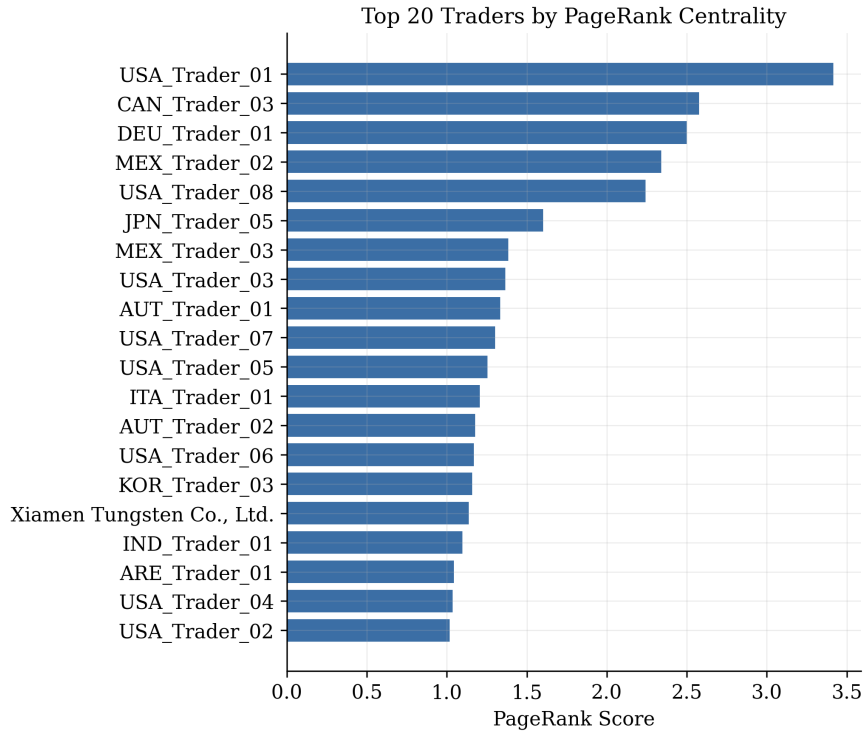


Figure 6.5. *Top twenty traders by PageRank centrality. The distribution is less concentrated than betweenness, reflecting the broader diffusion of recursive systemic importance across the dense trading core.*

6.2.4 Joint Analysis of Betweenness and PageRank

Examining the two metrics in isolation conceals their most analytically valuable property: they capture structurally distinct, only partially overlapping forms of importance. Figure 6.6 plots each trader in the betweenness–PageRank plane, with the dashed lines indicating the median of each metric. The scatter reveals three qualitatively different categories of systemically important traders.

The first category, located in the upper-right region, comprises traders that score highly on *both* metrics — the structural «super-brokers». `DEU_Trader_01`, `USA_Trader_01`, `CAN_Trader_03`, and `JPN_Trader_05` simultaneously control critical routing paths and occupy recursively central positions. These entities represent the most acute single-points-of-failure in the network: their disruption would degrade both the efficiency of material routing and the recursive stability of the downstream dependency structure.

The second category, exemplified by `Xiamen Tungsten Co., Ltd.` (betweenness $\approx 6,581$, PageRank ≈ 1.14), consists of pure brokers: traders with very high betweenness but comparatively moderate PageRank. Such entities act as indispensable gatekeepers on specific high-value routes without being recursively embedded across the wider network. Their criticality is structural and route-specific rather than systemic in the recursive sense.

The third and most instructive category occupies the upper-left region: traders with high PageRank but *zero* betweenness. The clearest examples are `MEX_Trader_02` (PageRank ≈ 2.34 , betweenness = 0) and `USA_Trader_08` (PageRank ≈ 2.25 , betweenness = 0). These entities are recursively important — they are embedded within a dense cluster of mutually reinforcing trade relationships — yet they lie on no shortest path and therefore exercise no brokerage control. They are systemically significant by virtue of *where* they sit in the dependency structure, not because they bridge otherwise disconnected regions. A monitoring framework relying on brokerage metrics alone would entirely overlook this class of node, which is precisely why the two measures must be read together.

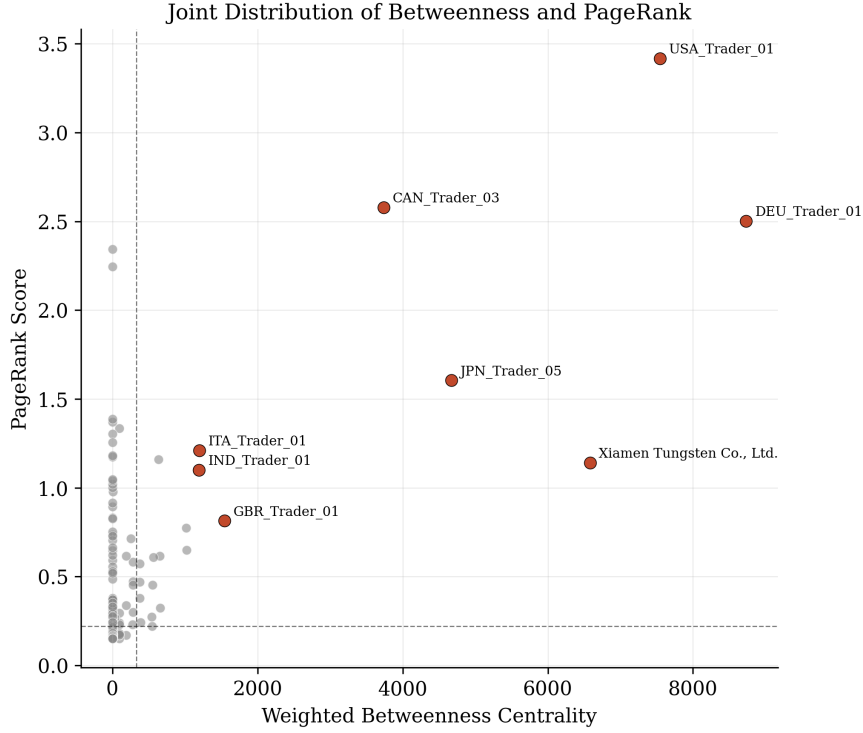


Figure 6.6. *Joint distribution of weighted Betweenness and PageRank. Dashed lines mark the median of each metric. Upper-right: structural super-brokers (high on both). Lower-right / pure brokers: high betweenness, moderate PageRank (e.g. Xiamen Tungsten). Upper-left: recursively central but zero-brokerage traders (e.g. MEX_Trader_02, USA_Trader_08).*

Taken together, the two metrics reveal a dual concentration mechanism: structural control over the tungsten network is exercised both through brokerage dominance over a small number of critical routes and through recursive reinforcement within a densely interconnected core. Neither dimension is reducible to the other, and neither is captured by the aggregate country-level indicators that conventional monitoring frameworks rely upon.

6.3 Geopolitical Risk Exposure

The centrality metrics analyzed above describe a trader's position within the network but say nothing about the external political environment in which it operates. A structurally central node located in a stable jurisdiction poses a very different risk profile from an equally central node located in a volatile one. To capture this interaction, we employ the composite Geopolitical Risk Exposure Index (GRI) introduced in Section 2.3, which blends a trader's recursive structural prestige (its normalized PageRank) with the geopolitical risk

score of its sovereign location.²

Rather than collapsing this two-dimensional construct into a single ranking, Figure 6.7 presents it as a decomposition, plotting each trader’s normalized PageRank (the structural axis) against its country geopolitical risk score (the instability axis), with point color encoding the resulting GRI. This representation makes the qualitatively different sources of exposure immediately legible. Three regions are of particular interest.

In the upper-left region lie traders located in highly unstable jurisdictions but with limited structural prestige: COD_Trader_01 (country risk ≈ 0.91), MMR_Trader_01 (≈ 0.89), and RUS_Trader_01 (≈ 0.78). Their exposure is *risk-driven*: they import volatility into the network through their location rather than through deep structural integration. In the lower-right region lies USA_Trader_01, which combines the maximal normalized PageRank in the entire network with a very low country risk (≈ 0.09). This node exemplifies a *structure-driven* but geopolitically decoupled profile: it is systemically central yet operates in a stable jurisdiction, and its elevated GRI score reflects structural prestige rather than political fragility. The genuinely high-exposure cases are those that combine both dimensions, exemplified by MEX_Trader_02 (normalized PageRank ≈ 0.67 , country risk ≈ 0.55), which records the highest overall GRI in the network.

²Following the construction of [26], the index combines country-level geopolitical risk with PageRank centrality. In this thesis the two components are combined as a weighted average, $GRI = w \cdot \widehat{PR} + (1 - w) \cdot R_{\text{country}}$, with $w = 0.5$; the weight is a transparent modeling parameter that can be adjusted to emphasize structural prestige or geopolitical instability.

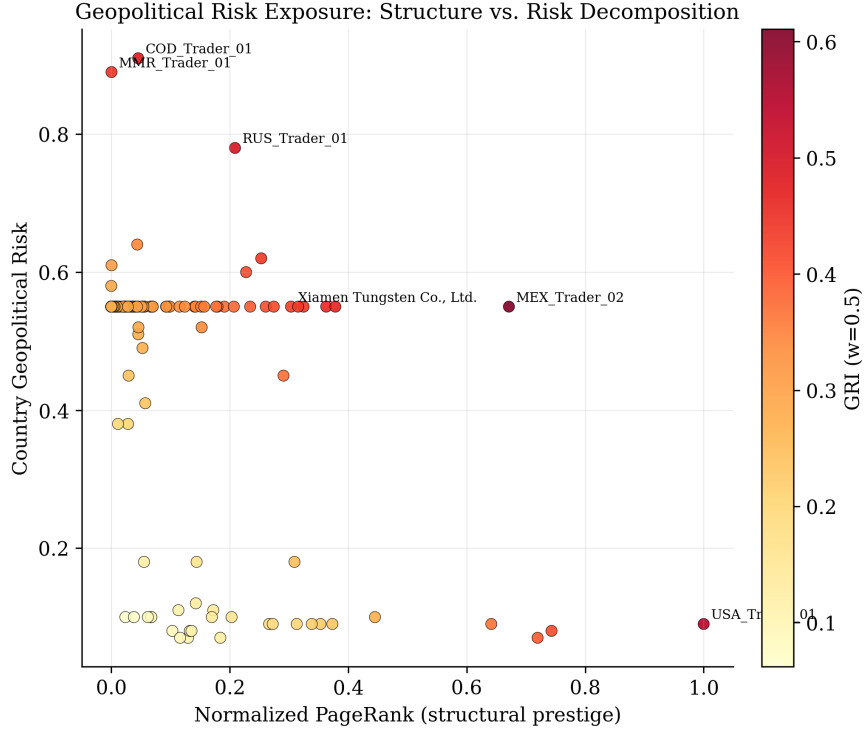


Figure 6.7. *Geopolitical Risk Exposure decomposed into its structural (normalized PageRank) and geopolitical (country risk) components, with point color encoding the composite GRI. Risk-driven nodes (e.g. COD_Trader_01, RUS_Trader_01) occupy the upper-left; the structurally dominant but geopolitically stable USA_Trader_01 occupies the lower-right.*

For completeness, Figure 6.8 reports the corresponding ranked GRI scores under the equal-weight ($w = 0.5$) configuration. It should be noted that a substantial number of countries share an identical geopolitical risk value of 0.55, arising from an approximate governance-indicator assignment for jurisdictions with incomplete data coverage; the exposure scores of traders located in these countries should therefore be interpreted as structurally rather than precisely calibrated, and the decomposition in Figure 6.7 provides a more robust reading than the absolute ranking alone.

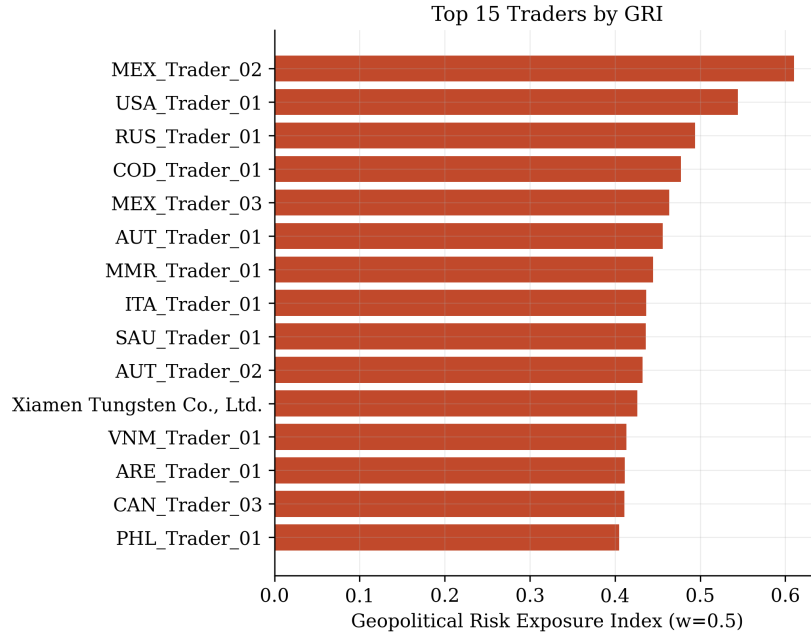


Figure 6.8. *Top fifteen traders by Geopolitical Risk Exposure Index under the equal-weight ($w = 0.5$) configuration.*

6.4 Cross-Metric Synthesis and Country-Level Aggregation

The preceding sections established that betweenness, PageRank, and geopolitical risk exposure each capture a distinct dimension of systemic importance. To quantify the degree to which these dimensions overlap, Figure 6.9 reports the matrix of Spearman rank correlations between the three centrality metrics and the total degree of each trader. The correlations confirm that the metrics are far from redundant. Betweenness and PageRank correlate only moderately ($\rho \approx 0.40$), reinforcing the conclusion of Section 6.2.4 that brokerage control and recursive importance are structurally distinct. PageRank correlates strongly with total degree ($\rho \approx 0.76$), indicating that recursive importance largely tracks weighted connectivity, whereas betweenness correlates only weakly with degree ($\rho \approx 0.36$) — a high-degree trader is not necessarily a broker. The geopolitical risk exposure index correlates weakly with betweenness ($\rho \approx 0.19$) and moderately with PageRank ($\rho \approx 0.38$), which is expected given that its structural component is derived from PageRank. The overall picture is one of complementary metrics: relying on any single measure would yield a partial and potentially misleading map of systemic vulnerability.

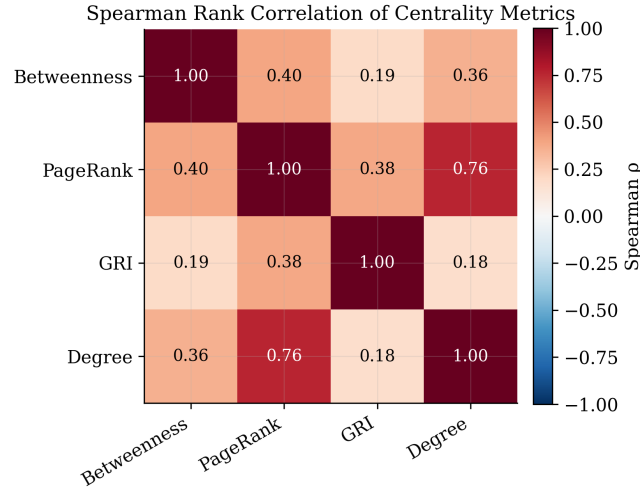


Figure 6.9. *Spearman rank-correlation matrix of the centrality metrics and total degree. The moderate betweenness–PageRank correlation and the strong PageRank–degree correlation confirm that the metrics capture distinct structural dimensions.*

Finally, to connect the firm-level results back to the macroeconomic level at which conventional monitoring operates, the trader-level centrality scores are aggregated to the country level in Table 6.4. The aggregation exposes a structural pattern that aggregate trader counts alone would conceal. Germany, represented by only three synthetic traders, records the highest aggregate betweenness in the entire network, driven almost entirely by the single dominant broker `DEU_Trader_01`. The United States, with eight traders, records the highest aggregate PageRank, reflecting the breadth of its recursively embedded trading entities. China, despite hosting the largest number of synthetic traders (twelve) and dominating physical production, ranks only third in aggregate betweenness. This decoupling between the number of national trading entities, physical production dominance, and structural network control is the central empirical finding of the chapter: systemic importance in the tungsten supply chain is a property of network position, not of trader count or aggregate volume.

Country	Traders	Total Betweenness	Total PageRank
DEU	3	8,732	3.78
USA	8	7,545	12.83
CHN	12	6,581	6.31
JPN	5	4,670	3.31
CAN	3	3,736	3.66
GBR	2	1,541	1.18
ITA	1	1,196	1.21
IND	2	1,192	1.35

Table 6.4. *Country-level aggregation of trader centrality. Aggregate structural control (betweenness) does not track the number of national traders: Germany leads with only three traders, while China, with the largest trader count and dominant physical production, ranks third.*

The quantitative analysis presented in this chapter has established the structural anatomy of the tungsten supply network: a dense, single weakly connected component organized around a strongly connected core of roughly half the traders; a brokerage hierarchy concentrated in fewer than a quarter of the entities; a broader diffusion of recursive importance; and a clear separation between structural centrality and geopolitical exposure. These structural properties describe *where* systemic importance resides under normal operating conditions. They do not, however, quantify how a disruption originating at a given node would actually propagate through the network and affect a target economy. That dynamic question — the short-run sensitivity of the Italian industrial system to localized supply shocks — is the subject of Chapter 7, where the structural metrics established here are confronted with the Nodal Elasticity stress-testing framework.

Chapter 7

Trader-Level Elasticity and Stress-Testing Results

The structural analysis of Chapter 6 mapped *where* systemic importance resides within the tungsten network under normal operating conditions. This chapter turns to the dynamic question: how does a localized supply disruption actually propagate through the reconstructed network, and how sensitive is the Italian industrial system to shocks originating at individual traders? The analysis applies the Nodal Elasticity framework formalized in Section 4.4, in which a shock to a node k is propagated through the column-normalized dependency matrix until convergence, and the resulting loss in incoming trade value to the target economy is measured. Two complementary impact measures are reported throughout: the *nodal elasticity* E_k , defined as the marginal loss per unit of shock under a small (1%) perturbation, and the *closure loss*, defined as the percentage reduction in the target economy's imports under a full (100%) removal of the shocked node. The propagation intensity is treated as a scenario parameter λ ; results are reported for the intermediate ($\lambda = 0.7$) and high ($\lambda = 0.9$) propagation regimes.¹

The chapter is organized as follows. Section 7.1 establishes the direct exposure profile of the Italian target economy, which motivates the distinction between direct and indirect supply risk. Section 7.2 reports the firm-level stress test of Chinese suppliers across propagation regimes. Section 7.3 extends the analysis to a global elasticity ranking across all traders, and Section 7.4 confronts the elasticity results with the structural centrality metrics of Chapter 6, revealing the indirect amplification channels that purely structural indicators fail to detect.

¹The limiting case $\lambda = 1.0$ is excluded from the reported results. With $\lambda = 1.0$ the damping term $(1 - \lambda)$ vanishes, the propagation becomes undamped, and the fixed-point iteration fails to converge within the iteration budget, producing runaway cascade values rather than a stable equilibrium. The analysis therefore restricts attention to the regime $\lambda < 1$ in which the propagation operator is contractive and a unique equilibrium exists.

7.1 The Direct Exposure Profile of the Italian Economy

Before simulating disruptions, it is instructive to characterize the baseline exposure of the target economy. Within the reconstructed network, the Italian industrial system is represented by a single aggregate trading entity importing a baseline value of approximately \$14.33 million from sixty distinct direct (Tier-1) suppliers. Figure 7.1 reports the composition of these direct imports by supplier country.

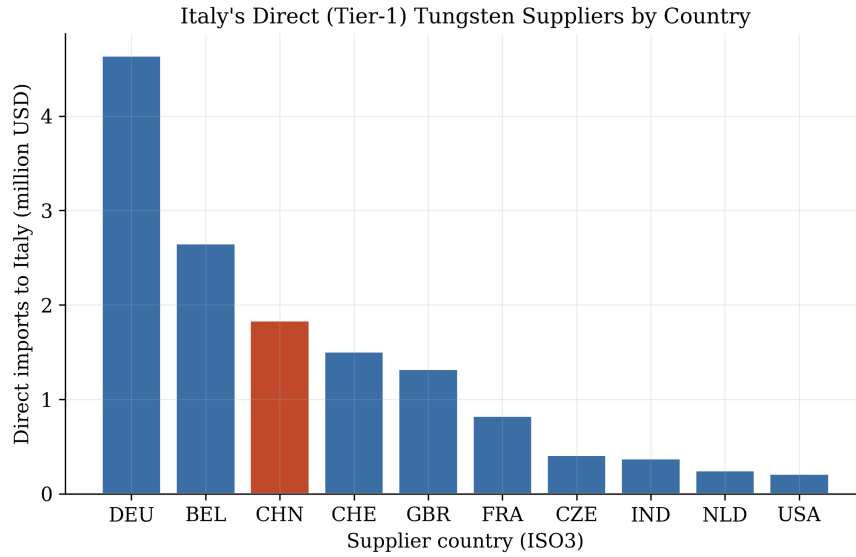


Figure 7.1. *Composition of Italy's direct (Tier-1) tungsten imports by supplier country. European suppliers dominate the direct import structure, while China (highlighted) accounts for only about 12.8% of direct imports.*

The composition reveals a structurally important fact. Italy's direct tungsten supply is dominated not by China but by European intermediaries: Germany is the single largest direct supplier (approximately \$4.64 million across its traders, led by `DEU_Trader_01`), followed by Belgium, and only then by China, which accounts for roughly 12.8% of Italy's direct imports. Switzerland, the United Kingdom, and France follow. Taken at face value, this profile would suggest that Italy's exposure to Chinese supply concentration is modest and heavily mediated by stable European partners. This impression, however, is precisely the kind of reassurance that direct, first-tier indicators provide and that the propagation analysis below shows to be incomplete: the European intermediaries on which Italy directly relies are themselves embedded in dependency chains that trace back to the concentrated upstream tungsten market. The relevant question is therefore not who supplies Italy directly, but how disruptions propagate *indirectly* through the multi-hop dependency structure.

7.2 Firm-Level Stress Testing of Chinese Suppliers

Given the dominance of Chinese groups in upstream tungsten production (Chapter 5), the first stress-testing scenario isolates the effect of disruptions to individual Chinese firms. Each of the twelve synthetic Chinese traders is shocked separately, and the resulting nodal elasticity of Italian imports is recorded across the propagation regimes $\lambda \in \{0.7, 0.8, 0.9\}$. Figure 7.2 reports the results for the six largest Chinese entities.

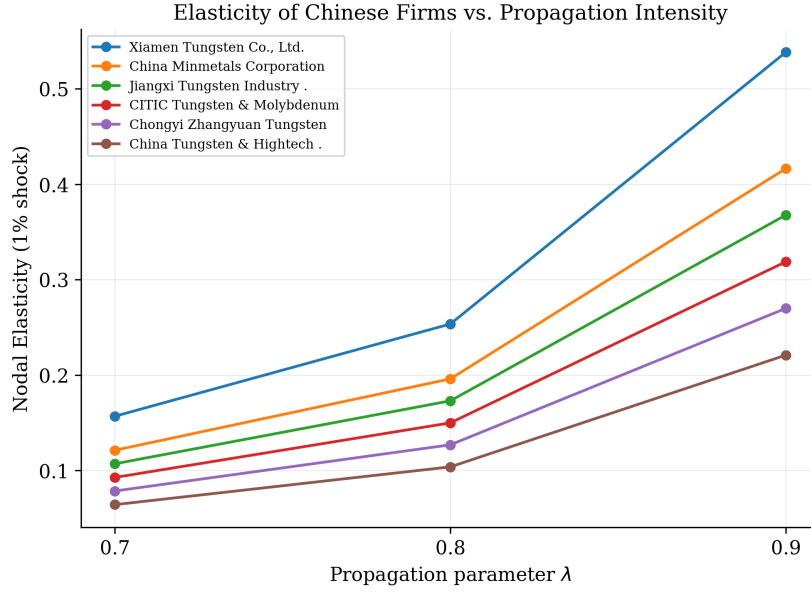


Figure 7.2. *Nodal elasticity of Italian imports to disruptions at the six largest Chinese traders, as a function of the propagation parameter λ . Elasticity rises sharply and convexly with λ , indicating that the systemic importance of these suppliers is driven predominantly by indirect, multi-hop propagation rather than direct bilateral exposure.*

Two observations emerge. First, the elasticity ordering among Chinese firms is stable across propagation regimes and tracks their relative market size: **Xiamen Tungsten Co., Ltd.** consistently records the highest elasticity (rising from 0.157 at $\lambda = 0.7$ to 0.538 at $\lambda = 0.9$), followed by China Minmetals, Jiangxi Tungsten, and CITIC. Second, and more importantly, the elasticity of every Chinese firm increases sharply and convexly with λ . Because λ governs the strength of indirect, multi-step transmission, this steep dependence demonstrates that the systemic impact of Chinese suppliers on Italy operates overwhelmingly through indirect propagation channels rather than through direct bilateral trade — consistent with the finding of Section 7.1 that China supplies only a small fraction of Italy’s imports directly. Under weak propagation, even the largest Chinese supplier produces only a marginal effect on Italian imports; under strong propagation, the same disruption is amplified several-fold as it cascades through the intermediary network.

7.3 Global Elasticity Ranking and Indirect Amplifiers

To identify the most systemically important traders across the entire network, the stress test is extended to every trader: each node is shocked in turn and ranked by the resulting elasticity of Italian imports. Table 7.1 reports the ten highest-ranked traders together with their elasticity and closure-loss values under both propagation regimes, and Figure 7.3 visualizes the elasticity ranking at $\lambda = 0.9$.

Rank	Firm	$\lambda = 0.9$		$\lambda = 0.7$	
		Elasticity	Closure (%)	Elasticity	Closure (%)
1	MYS_Trader_02	0.969	41.35	0.164	12.09
2	JPN_Trader_05	0.853	43.76	0.201	16.02
3	BRN_Trader_01	0.748	31.86	0.104	7.37
4	DEU_Trader_01	0.549	43.94	0.356	33.27
5	Xiamen Tungsten Co., Ltd.	0.538	35.93	0.157	13.99
6	China Minmetals Corporation	0.417	30.06	0.121	11.08
7	MYS_Trader_01	0.393	25.43	0.066	5.77
8	Jiangxi Tungsten Industry Group	0.368	27.43	0.107	9.88
9	JPN_Trader_01	0.327	23.98	0.077	6.99
10	CITIC Tungsten & Molybdenum	0.319	24.62	0.093	8.65

Table 7.1. *Top ten traders by nodal elasticity of Italian imports, with elasticity and closure-loss measures under the high ($\lambda = 0.9$) and intermediate ($\lambda = 0.7$) propagation regimes.*

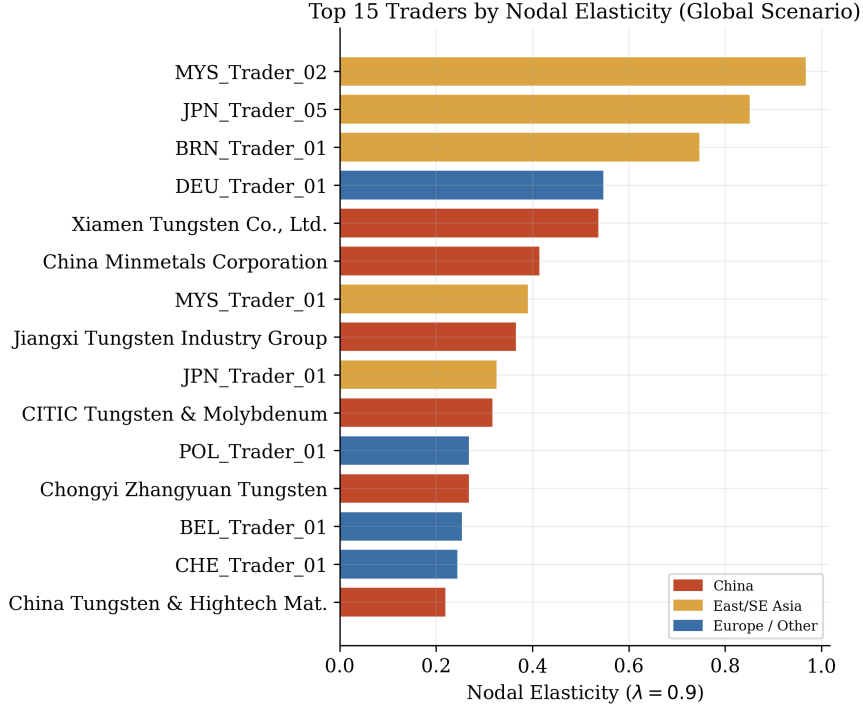


Figure 7.3. *Top fifteen traders by nodal elasticity at $\lambda = 0.9$, colored by region. The highest-elasticity nodes are East and Southeast Asian intermediaries rather than the dominant Chinese producers.*

The global ranking yields the central empirical result of the thesis. The traders that generate the largest downstream effect on the Italian economy under strong propagation are not the dominant Chinese producers, but East and Southeast Asian intermediaries: **MYS_Trader_02** (elasticity 0.969), **JPN_Trader_05** (0.853), and **BRN_Trader_01** (0.748) occupy the top three positions, ahead of the largest German broker and the dominant Chinese group Xiamen Tungsten. These intermediaries do not supply Italy in large direct volumes; rather, they sit at structurally pivotal positions within dense regional production ecosystems through which disruptions are relayed and amplified. They function as *indirect amplification channels* rather than direct exposure points — a role that is entirely invisible to the country-level concentration metrics on which conventional monitoring frameworks rely.

A second, more subtle pattern emerges from the contrast between the elasticity and closure-loss columns. The two measures do not produce the same ranking. Under marginal perturbation ($\lambda = 0.9$ elasticity), **MYS_Trader_02** is the single most sensitive node; yet under full closure, the largest absolute loss to Italy is inflicted by **DEU_Trader_01** (43.94%) and **JPN_Trader_05** (43.76%), with **MYS_Trader_02** producing a somewhat smaller 41.35%. This reordering reflects a genuine non-linearity in the propagation dynamics. The marginal elasticity captures the steepness of the initial response and is dominated by indirect amplification, whereas the closure loss captures the total displacement under complete removal and is influenced by the magnitude of the direct supply relationship. **DEU_Trader_01**,

which Section 7.1 identified as Italy’s single largest direct supplier, naturally produces the greatest absolute loss when fully removed, even though its marginal elasticity is lower than that of the Asian amplifiers. The two measures are therefore complementary: marginal elasticity identifies the firms whose disruptions are most strongly amplified by network position, while closure loss identifies those whose complete failure would cause the greatest absolute damage.

7.4 Nodal Elasticity versus Structural Centrality

The defining contribution of the elasticity framework is its ability to capture forms of systemic importance that structural centrality metrics cannot detect. To make this explicit, Figure 7.4 plots each trader’s nodal elasticity (at $\lambda = 0.9$) against its weighted betweenness centrality from Chapter 6. The two measures correlate only weakly (Spearman $\rho \approx 0.44$), and the structure of their disagreement is highly informative.

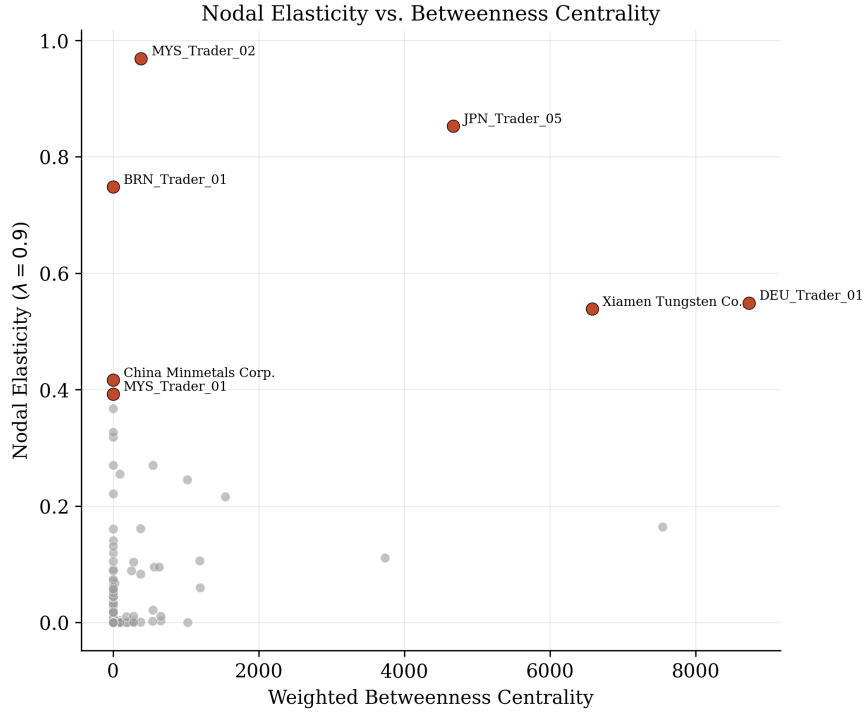


Figure 7.4. Nodal elasticity ($\lambda = 0.9$) versus weighted Betweenness centrality. The highest-elasticity amplifiers (MYS_Trader_02, BRN_Trader_01) occupy the upper-left region — maximal propagation impact with negligible or zero brokerage — making them entirely invisible to structural metrics.

The scatter separates the systemically important traders into two distinct classes. The first, occupying the upper-right region, comprises nodes that are important on both dimensions: JPN_Trader_05 (elasticity 0.853, betweenness $\approx 4,670$) and, to a lesser elasticity extent, DEU_Trader_01 and Xiamen Tungsten combine high propagation impact with

high brokerage control. These nodes would be flagged by any structural analysis. The second class, occupying the upper-left region, is the critical one: **MYS_Trader_02** records the highest elasticity in the entire network (0.969) while carrying a near-negligible betweenness (≈ 385), and **BRN_Trader_01** records the third-highest elasticity (0.748) with a betweenness of *exactly zero*. These traders lie on no shortest path and exercise no brokerage control whatsoever; a monitoring system built on centrality metrics would dismiss them as peripheral. Yet they are among the most dangerous amplifiers of supply shocks to the Italian economy.

This dissociation is the empirical core of the thesis’s argument. Structural centrality identifies the brokers that control the network’s routing under static conditions, but it cannot anticipate which nodes will amplify a disruption once it begins to propagate through the multi-hop dependency structure. Nodal elasticity captures precisely this dynamic, propagation-based dimension of vulnerability. The fact that the highest-elasticity node in the network has almost no betweenness, while the highest-betweenness node has only moderate elasticity, demonstrates conclusively that the two perspectives are complementary and that neither can substitute for the other. For a regulator implementing the stress-testing mandates of the EU CRMA, the implication is direct: resilience assessments based solely on structural concentration or brokerage indicators will systematically overlook the indirect amplifiers — frequently located in politically stable, low-profile intermediary economies — that pose the greatest dynamic risk to downstream industrial systems.

These trader-level results, taken together with the structural analysis of Chapter 6, provide the empirical foundation for the conclusions and policy implications discussed in Chapter 8.

Chapter 8

Conclusions, Limitations and Future Work

8.1 Conclusions

This thesis set out to test whether a firm-level, graph-based reconstruction of a critical raw material supply chain can reveal vulnerabilities that conventional country-level indicators cannot detect. Using the global tungsten trade network as an empirical case study, the results consistently confirm that it can.

The first result concerns market concentration itself. The observed supplier-side export HHI for tungsten is 0.157, describing what appears to be a moderately concentrated market. Yet the firm-level network, reconstructed and calibrated on this same figure, shows that structural control over high-value downstream trade flows is concentrated in a small number of pivotal trading entities — a pattern entirely invisible in the aggregate statistic alone.

The second result concerns network structure. Centrality analysis identifies two distinct types of critical firms: high-betweenness intermediaries that control key routing paths, most notably German and U.S. traders, and high-PageRank firms embedded in densely interconnected regions of the network. Neither type correlates strongly with raw trade volume, confirming that structural position, not size, determines systemic importance.

The third and most significant result concerns disruption propagation. The Nodal Elasticity simulations show that the firms posing the greatest risk to the Italian industrial system are not the dominant Chinese producers, but Japanese and Southeast Asian intermediaries — notably `MYS_Trader_02`, `JPN_Trader_05`, and `BRN_Trader_01` — which act as indirect amplification channels under strong propagation regimes. German intermediaries, by contrast, expose Italy mainly through stable, direct trade relationships. This distinction between direct and amplified exposure is only visible through firm-level, propagation-based analysis and remains undetectable in country-level stress tests.

Beyond these empirical findings, the thesis also proposes a generalized multi-tier, multi-commodity conceptual framework, formalizing the supply chain as a multilayer directed graph aligned with the EU CRMA taxonomy and introducing a Composite Structural Fragility Index to capture cross-commodity risk at the firm level. While its full implementation is left to future work, it provides a regulatory-aligned blueprint that existing

macro-level indicators, such as the HHI, cannot offer.

Taken together, these results show that systemic vulnerability in critical raw material supply chains is not determined solely by aggregate import dependence, but by the position of individual firms within indirect propagation networks. Resilience strategies based only on country-level diversification therefore risk overlooking the intermediaries that matter most.

8.2 Limitations

Despite these contributions, several limitations remain. First, the model relies on synthetic firm generation to compensate for limited data transparency in CRM markets. Although the probabilistic reconstruction is calibrated using empirical trade distributions and macroeconomic benchmarks (e.g., UN Comtrade data), the resulting firms do not correspond to directly observable entities. The objective is therefore not predictive reconstruction of exact firm identities, but rather stress-testing of plausible concentration structures consistent with observed macro-level trade regularities.

Second, nodal elasticity is computed under fixed trade dependencies and a static network structure, excluding adaptive mechanisms such as supplier substitution, inventory buffering, price adjustments, and endogenous market responses. While this simplification isolates pure topological propagation effects, future work could incorporate agent-based and behavioural dynamics to capture adjustment processes under sustained shocks.

Third, the current Knowledge Graph focuses primarily on trading entities and country-level relationships, omitting upstream and downstream actors such as miners, refiners, recyclers, manufacturers, and logistics providers. Consequently, a significant limitation of the empirical framework presented in this thesis is that the intricate multi-tier and multi-commodity graph modeling formalized in Section 4.5 remains purely conceptual. Due to acute data scarcity regarding verified corporate procurement trails and the compressed execution window governed by the academic graduation session, the numerical simulations and derived Nodal Elasticity profiles are bound to the structural limits of the baseline model (*Case 1*, as defined in Section 4.5). The mass-balance tracking is mathematically mapped for a singular material topology, omitting the simultaneous processing loops of parallel critical commodities. Extending the ontology to include these entities would enable a more complete representation of vertical integration, processing bottlenecks, and circular economy dynamics.

Finally, trade flow allocation among synthetic firms is based on probabilistic assumptions consistent with observed concentration patterns. Although this approach enables scalable reconstruction under data scarcity, future research will integrate customs micro-data, corporate disclosures, shipping records, and ownership networks to improve empirical fidelity.

8.3 Policy Implications and Replicability

Beyond its academic contributions, the framework developed in this thesis carries direct implications for regulatory bodies and government agencies responsible for implementing the EU Critical Raw Materials Act. A defining feature of the proposed methodology is that

it relies exclusively on publicly available data sources: UN Comtrade trade flow statistics, World Bank Worldwide Governance Indicators, and the EU JRC criticality assessments. No proprietary datasets, commercial data subscriptions, or firm-level customs records are required to deploy the baseline system. This design choice was deliberate: a monitoring framework that depends on data sources unavailable to public institutions cannot serve as a policy tool.

The full pipeline, from raw Comtrade data extraction to Knowledge Graph population and stress-testing simulation, can be replicated by any government agency or regulatory body with access to a standard computing environment and an open-source Neo4j Desktop installation. Switching the framework from tungsten to any other critical raw material listed in the EU CRMA — such as lithium, cobalt, rare earth elements, or gallium — requires only two changes: replacing the input Comtrade HS codes in the configuration file and updating the EU JRC criticality parameters for the target material. The core algorithmic pipeline, including the Pareto calibration procedure, the Knowledge Graph schema, and the Nodal Elasticity simulation, remains entirely unchanged.

This replicability is not merely a technical convenience. It is a necessary condition for the framework to function as a policy instrument. Regulatory stress-testing methodologies must be transparent, auditable, and independently verifiable. The explicit parameterization of the synthetic firm generation procedure — where every modeling assumption is expressed through a small number of economically interpretable parameters — ensures that policy analysts can understand, challenge, and adjust the underlying assumptions without requiring deep technical expertise in graph databases or network science.

Furthermore, the multi-tier conceptual framework proposed in Section 4.5 establishes a regulatory-aligned blueprint that directly maps onto the traceability obligations defined in Article 24 of the EU CRMA. As firm-level customs microdata and corporate disclosure records progressively become available through regulatory reporting mandates, the framework can be upgraded from the current synthetic baseline to an empirically grounded monitoring system without any structural changes to the underlying Knowledge Graph architecture. The graph schema is designed to accommodate this transition: synthetic Trader nodes can be replaced one by one with empirically verified entities as data becomes available, preserving the full analytical and stress-testing infrastructure already in place.

8.4 Future Work

The approach, therefore, suggests the importance of complementing aggregate trade monitoring with firm-level dependency mapping and network-oriented stress-testing exercises. More broadly, the graph-native representation enables future integration with semantic reasoning, ontology-based querying, and explainable dependency tracing across multi-tier supply networks. From a policy perspective, this could support more targeted interventions, including identifying bottleneck intermediaries, monitoring highly central upstream actors, strategically stockpiling critical downstream components, and diversifying policies that account for indirect propagation channels rather than direct import shares alone.

Building directly upon the conceptual frameworks developed in this research, the primary trajectory for future work entails the operationalization of the multilayer graph formalization. Future research cycles will focus on expanding the software implementation

from the current single-tier transactional platform to a full multi-stage graph system in Neo4j, actively splitting nodes into specialized functional classes such as Miners, Refiners, and Manufacturers. Furthermore, by incorporating advanced algorithmic path-traversal mechanics, future iterations can programmatically ingest corporate Bill of Materials (BOM) instances, allowing automatic risk calculations across multi-commodity chains. This will fully unlock the predictive and policy-monitoring potential of the proposed Strategic Digital Twin architecture.

From a policy deployment perspective, a concrete next step would be the formal integration of this framework into the monitoring infrastructure being developed under the EU Critical Raw Materials Act. The European Commission’s Joint Research Centre and the newly established Critical Raw Materials Board represent natural institutional partners for piloting the system at the EU level. A phased deployment could begin with the materials for which Comtrade data coverage is most complete — tungsten, lithium, and cobalt — and progressively expand as regulatory reporting obligations generate richer firm-level data streams. At the national level, the Italian industrial system analyzed in this thesis represents a particularly suitable starting point for a government-sponsored pilot, given its deep integration into European advanced manufacturing supply chains and its structural exposure to the upstream concentration risks identified in the stress-testing results.

Bibliography

- [1] International Energy Agency (IEA). Strait of Hormuz. IEA, Oil Security and Emergency Response. Available at: <https://www.iea.org/about/oil-security-and-emergency-response/strait-of-hormuz>.
- [2] European Parliament and Council of the European Union. Regulation (EU) 2024/1252 of the European Parliament and of the Council of 11 April 2024 establishing a framework for ensuring a secure and sustainable supply of critical raw materials. Official Journal of the European Union. 2024. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1252/oj>.
- [3] B. Dumont and X.-L. Varela-Irimia, Supply Chain Stress Tests for Critical Inputs: A Proof-of-Concept, (No. WP2024/27), Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, European Commission, 2024.
- [4] U.S. Geological Survey. Mineral Commodity Summaries 2024: Tungsten. U.S. Department of the Interior. 2024. Available at: <https://pubs.usgs.gov/periodicals/mcs2024/mcs2024-tungsten.pdf>.
- [5] European Commission, Joint Research Centre, Raw Materials Scoreboard 2023. Publications Office of the European Union, 2023.
- [6] D. Ivanov, Intelligent Digital Twin (iDT) for Supply Chain Stress-Testing, Resilience, and Viability, International Journal of Production Economics, vol. 263, p. 108938, 2023.
- [7] D. Ivanov, Simulation-based ripple effect modelling in the supply chain, International Journal of Production Research, vol. 55, no. 7, pp. 2083–2101, 2017.
- [8] D. Ivanov, Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case, Transportation Research Part E, vol. 136, p. 101922, 2020.
- [9] A. Ghadge, S. Wurtmann, and S. Seuring, Managing climate change risks in global supply chains: a review and research agenda, International Journal of Production Research, vol. 58, no. 1 (2020): 44-64., Available at SSRN: <https://ssrn.com/abstract=3399073>
- [10] Y. Li, C.W. Zobel, O. Seref, and D. Chatfield, Network characteristics and supply chain resilience under conditions of risk propagation, International Journal of Production Economics, vol. 223, p. 107529, 2020.
- [11] Y. Li and C. W. Zobel, Exploring supply chain network resilience in the presence of the ripple effect, International Journal of Production Economics, vol. 228, p. 107693, 2020.
- [12] E. E. Kosasih and A. Brintrup, A machine learning approach for predicting hidden

- links in supply chain with graph neural networks, *International Journal of Production Research*, vol. 60, no. 17, pp. 5380–5393, 2022.
- [13] T. Sawik, A stochastic optimization approach to maintain supply chain viability under the ripple effect, *International Journal of Production Research*, vol. 61, no. 8, pp. 2452–2469, 2023.
- [14] T. Sawik, Supply Chain Viability: Risk-Averse Decision-Making. In *Stochastic Programming in Supply Chain Risk Management: Resilience, Viability, and Cybersecurity* (pp. 187–217). Springer Nature, 2024.
- [15] Liu, Ming & Ding, Yueyu & Chu, Feng & Dolgui, Alexandre & Zheng, Feifeng, 2024. "Robust actions for improving supply chain resilience and viability," *Omega*, Elsevier, vol. 123(C).
- [16] S. Hosseini and D. Ivanov, Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review, *Expert Systems with Applications*, vol. 161, p. 113649, 2020.
- [17] M. Liu, Z. Liu, F. Chu, F. Zheng, C. Chu, A new robust dynamic Bayesian network approach for disruption risk assessment under the supply chain ripple effect, *International Journal of Production Research*, 59(1), 265–285, 2021.
- [18] A. Pavlov, D. Ivanov, A. Dolgui, and B. Sokolov, Hybrid fuzzy-probabilistic approach to supply chain resilience assessment, *IEEE Transactions on Engineering Management*, vol. 65, no. 2, pp. 65(2), 303–315, 2017.
- [19] A. Brintrup et al., Predicting hidden links in supply networks, *Complexity*, 2018(1), pages 1–12.
- [20] F. Bodendorf, M. Sauter, and J. Franke, A mixed methods approach to analyze and predict supply disruptions by combining causal inference and deep learning." *International Journal of Production Economics* 256 (2023): 108708.
- [21] H. Ren, et al. *Vulnerability to geopolitical disruptions of the global electric vehicle lithium-ion battery supply chain network*, *Computers & Industrial Engineering*, 188 (2024): 109919.
- [22] X. Gabaix, The Granular Origins of Aggregate Fluctuations, *Econometrica*, vol. 79, no. 3, pp. 733–772, 2011.
- [23] C. Freund and M. D. Pierola, The Origins and Dynamics of Export Superstars, *The World Bank Economic Review*, vol. 34, no. 1, pp. 28–47, 2020.
- [24] J. Eaton, S. Kortum, and F. Kramarz, An Anatomy of International Trade: Evidence From French Firms, *Econometrica*, vol. 79, no. 5, pp. 1453–1498, 2011.
- [25] R. L. Axtell, Zipf distribution of US firm sizes, *Science*, vol. 293, no. 5536, pp. 1818–1820, 2001.
- [26] RE4IT, CASCADE Project: Criticality Analysis for Supply Chain Availability and Dependency Evaluation, Confindustria, 2026.
- [27] V. Borisov, K. Seßler, T. Leemann, M. Pawelczyk, and G. Kasneci, "Language Models are Realistic Tabular Data Generators," in *Proc. ICLR*, 2023.
- [28] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular data using Conditional GAN," in *NeurIPS*, 2019.
- [29] A. Kotelnikov et al., TabDDPM: Modelling Tabular Data with Diffusion Models, *arXiv preprint arXiv:2209.15421*, 2022.
- [30] A. Marshall, The elasticity of wants, in *Principles of Economics*, pp. 86–97, Macmillan, 1890.

- [31] C. E. Boehm, A. A. Levchenko, and N. Pandalai-Nayar, The long and short (run) of trade elasticities, *American Economic Review*, vol. 113, no. 4, pp. 861–905, 2023.
- [32] J. di Giovanni and A. A. Levchenko, Country size, international trade, and aggregate fluctuations in granular economies, *Journal of Political Economy*, vol. 120, no. 6, pp. 1083–1132, 2012.
- [33] Ministry of Commerce of the People’s Republic of China (MOFCOM). Reports on Export License Administration for Critical Minerals. 2024, Available at: <https://english.mofcom.gov.cn/>.
- [34] A. Abeltino, T. Bacaloni, A. Bernardini, F. Giancaterini, and A. Pannone, Power and Control in Complex Networks: A Taxonomy and Critical Review, arXiv preprint arXiv:2601.10218, 2026.
- [35] B. Kitchenham, Procedures for performing systematic reviews, Keele University Technical Report SE-0401, 2004.
- [36] Selenium Authors, Selenium WebDriver, 2024. Available at: <https://www.selenium.dev>.
- [37] Yang, An, et al. "Qwen3 technical report." arXiv preprint arXiv:2505.09388 (2025).
- [38] L. Mungo, A. Brintrup, D. Garlaschelli, and F. Lafond, Reconstructing supply networks, *Journal of Physics: Complexity*, vol. 5, no. 1, p. 012001, 2024.
- [39] N. Ramzy, S. Auer, H. Ehm, and J. Chamanara, SENS: Semantic synthetic benchmarking model for integrated supply chain simulation and analysis, arXiv preprint arXiv:2205.06484, 2022.
- [40] P. Li, Q. Zhao, Y. Liu, C. Zhong, J. Wang, and Z. Lyu, Survey and Prospect for Applying Knowledge Graph in Enterprise Risk Management, *Computers, Materials & Continua*, vol. 78, no. 3, 2024.
- [41] M. Shahsavari, O. K. Hussain, P. Sharma, and M. Saberi, Modelling supply chain risk events by considering their contributing events: a systematic literature review, *Enterprise Information Systems*, vol. 19, no. 5–6, p. 2472303, 2025.
- [42] Y. Liu et al., A knowledge graph perspective on supply chain resilience, arXiv preprint arXiv:2305.08506, 2023.
- [43] N. Ramzy, S. Auer, H. Ehm, and J. Chamanara, MARE: Semantic supply chain disruption management and resilience evaluation framework, arXiv preprint arXiv:2205.06499, 2022.
- [44] J. Wen, B. Gabrys, and K. Musial, Toward digital twin oriented modeling of complex networked systems and their dynamics: A comprehensive survey, *IEEE Access*, vol. 10, pp. 66886–66923, 2022.
- [45] V. KEK, J. Kandasmay, V. Rajasekharan, B. George, and R. Joe Mathew, Digital Twin Model of Semiconductor Supply Chain for Managing Disruption and Resilience Through Data Driven Experiments, available at SSRN 4530357, 2023.
- [46] Kadam, Akash Abaji. "Resilient Supply Chains under Tariff Volatility: A Digital Twin-Enabled FLEX Model for Strategic Sourcing and Risk Mitigation." *Journal of Business and Management Studies* 7.4 (2025): 247-257.
- [47] R. Angles, The Property Graph Database Model, in *Proceedings of the 6th International Workshop on Graph Data-management Experiences & Systems (GRADES)*, 2018.
- [48] I. Robinson, J. Webber, and E. Eifrem, *Graph Databases*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2015.

- [49] Neo4j, Inc., Neo4j Graph Data Science Library Manual. Available at: <https://neo4j.com/docs/graph-data-science/current/>.
- [50] World Wide Web Consortium (W3C), Resource Description Framework (RDF). Available at: <https://www.w3.org/RDF/>.
- [51] World Wide Web Consortium (W3C), OWL 2 Web Ontology Language. Available at: <https://www.w3.org/OWL/>.
- [52] World Wide Web Consortium (W3C), SPARQL 1.1 Query Language. Available at: <https://www.w3.org/TR/sparql11-query/>.
- [53] European Commission, Joint Research Centre, Raw Materials Information System (RMIS). Available at: <https://rmis.jrc.ec.europa.eu/>.
- [54] S. P. Borgatti, Centrality and Network Flow, *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [55] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford InfoLab, 1999.
- [56] Neo4j, Inc., Neo4j Desktop. Available at: <https://neo4j.com/download/>.
- [57] Python Software Foundation, Python Programming Language. Available at: <https://www.python.org/>.
- [58] Neo4j, Inc., Neo4j Python Driver Manual. Available at: <https://neo4j.com/docs/api/python-driver/current/>.
- [59] W. McKinney, Data Structures for Statistical Computing in Python, in *Proceedings of the 9th Python in Science Conference*, pp. 56–61, 2010.
- [60] C. R. Harris et al., Array Programming with NumPy, *Nature*, vol. 585, pp. 357–362, 2020.
- [61] P. Virtanen et al., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [62] LM Studio, LM Studio — Discover, Download, and Run Local LLMs. Available at: <https://lmstudio.ai/>.
- [63] United Nations Statistics Division, UN Comtrade Database. Available at: <https://comtrade.un.org/>.
- [64] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, Survey of Hallucination in Natural Language Generation, *ACM Computing Surveys*, vol. 55, no. 12, Article 248, 2023.
- [65] Mordor Intelligence, Tungsten Market Size & Share Analysis — Growth Trends & Forecasts, 2026–2031. Available at: <https://www.mordorintelligence.com/industry-reports/tungsten-market>. Last Accessed on: July 10, 2026.