Supplementary Information for: How Network Structure Shapes Human Rights Risk Propagation and Due Diligence Efficiency

Jan Hurt, Katharina Ledebur, Birgit Meyer, Markus Gerschberger, Stefan Thurner, Klaus Friesenbichler, and Peter Klimek

Supplementary Methods

Background and Limitations

We implement a statistical approach to reconstruct firm networks using Input-Output (IO) tables and trade data as mesoscale ground truths. Note that there are subtle but important differences between IO tables and production networks due to differences in accounting. The most relevant differences for our model are related to investments and financial services (for a more thorough discussion, see the Appendix of [1]). Investments are recorded as links between firms in production networks, whereas in IO tables they are recorded under final demand. Consequently, links to sectors from which investments are often sourced might be underrepresented in IO tables compared to firm-level networks. The measurement, of financial services may differ between national accounts and different types of firm-level networks. However, as we lack detailed financial accounts that would allow us to adjust the IO tables for these differences in accounting, we acknowledge these limitations and proceed with caution when interpreting results for the financial sectors in our model.

Furthermore, we construct the firm-network based on the fact that empirical characteristic distances in firm-level production and supply networks are statistically identical to their expectations from configuration models. These properties can be conveniently and simultaneously captured in a class of random network models known as degree-corrected stochastic block models [2]. Stochastic block models dissect the node set into disjunct blocks and specify link densities between nodes from different blocks. In our case, these blocks correspond to industries from specific regions and the link densities are proportional to entries in the IO table (within EU links) or trade data (links between EU and non-EU companies). Degree-corrected stochastic block models additionally constrain the (in- and out-) degree distributions, essentially leading to a configuration model with a block structure. Recent research has demonstrated that the configuration model and variations that employ node strengths instead of node degrees generally provide a realistic generative model [3].

Data sources and harmonization

We utilize three pillars of quantitative input: (1) Eurostat Structural Business Statistics (SBS) to parameterize the firm population by country, sector, and size (2) OECD Inter-Country Input–Output tables (release used in the study is 2021) to obtain sector-by-sector flow intensities across countries as meso-level link propensities; and (3) UN Comtrade merchandise trade to allocate non-EU import origins by sector. Sector classifications are harmonized to the OECD IO sectoral aggregation. Harmonization follows a two-step crosswalk: HS→NACE (for trade) and NACE→OECD IO sectors. All sector-level results presented in the manuscript are aggregated to the OECD IO sectors to ensure consistency between the meso-ground truth and the generated micro-links.

Reconstructing missing values in Eurostat's Structural Business Statistics

At the NACE 2-digit × size-class level, Eurostat's Structural Business Statistics (SBS) contain suppressions due to confidentiality. Roughly $\sim 20\%$ of cells are missing, though these omissions represent only $\sim 2\%$ of aggregate turnover. We address this with a multi-step, aggregate-consistent imputation procedure.

Guiding principles. We use a hierarchical procedure: (i) preservation of aggregates, meaning finer-level totals must match the next-higher reported totals (country × NACE L1 and, where available, country × NACE L2); (ii) proportional filling, meaning missing cells are imputed using observed share patterns from countries × sectors × size classes where these shares are available, and when detail is missing at a given tier we borrow the corresponding share pattern from the next broader tier (country-level to EU-level) and re-normalize to maintain aggregate consistency.

(1) Imputing revenue at NACE L2 within L1 aggregates. Let $L2(g^{(L1)})$ denote the set of NACE L2 (two–digit) sectors that belong to a given NACE L1 (one–digit) group $g1$. For each country c and L1 sector $g1$, let $R^{(L1)}_{c,g1}$ be the reported L1 total and $R^{(L2)}_{c,h}$ the (possibly missing) L2 totals for $h \in L2(g^{(L1)})$. Define the L1–L2 gap

$$\Delta^{(L1\text{–}L2)} R_{c,g1} \;=\; R_{c,g1} \;-\; \sum_{h2 \,\in\, L2(g^{(L1)})} R_{c,h2}. \tag{1}$$

For any country c with complete L2 coverage inside $g1$, define the within-$g1$ revenue share of L2 code $h2$ as

$$s_{c,h2|g1} \;=\; \frac{R_{c,h2}}{\sum_{h2' \in L^{(2)}(g1)} R_{c,h2'}}\,. \tag{2}$$

Using the set of countries with complete L2 coverage in group $g^{(L1)}$, we compute reference shares $\bar{s}_{h|g1}$, and then allocate the gap proportionally to missing L2 codes:

$$\widehat{R}_{c,h2} \;=\; R^{\text{obs}}_{c,h2} \;+\; \mathbf{1}\{h2 \text{ missing}\}\, \bar{s}_{h2|g1}\, \Delta^{(L1\text{–}L2)} R_{c,g1}\,. \tag{3}$$

(2) Disaggregating L2 revenue to size classes. Within each L2 sector, we apply the same aggregate-consistent proportional scheme used for L1→L2 to disaggregate total L2 revenue to size classes and to impute any missing size-class revenues.

(3) Imputing *rpe*, employees per firm, employment, and firm counts. For missing revenue-per-employee $rpe_{c,h2,b}$ use country–sector–size averages where available, otherwise substitute the sector–size EU average. Missing average employees per firm $\bar{e}_{c,h2,b}$ are imputed analogously from EU sector–size averages. Then obtain employment

$$E_{c,h2,b} \;=\; \frac{\widehat{R}_{c,h2,b}}{rpe_{c,h2,b}}\,, \tag{4}$$

and derive the number of firms

$$N_{c,h2,b} \;=\; \frac{E_{c,h2,b}}{\bar{e}_{c,h2,b}}\,. \tag{5}$$

This ordering preserves additivity across size classes and matches known aggregates by construction.

Additional Information on the Sampling Size Distribution

The first step in reconstructing the firm-level production network is to create a set of synthetic companies whose size distribution, in terms of the number of employees, turnover, and in-/out-degree, matches the data described above. We assume that the distribution of company sizes, measured by the number of employees per company, follows a heavy-tailed distribution [4, 5]. We parameterize this distribution as a Pareto Type II distribution (see equation (1) in the main text). This reduces to Zipf's law when $\alpha = 1$, which is commonly observed in firm size distributions [6]. For each firm size group, we calculate $\lambda^{(c,s,b)}$ such that:

$$E^b(f_{(1,\lambda)}) := \frac{\int_{b_{lower}}^{b_{upper}} x \cdot f_{(1,\lambda)}(x)dx}{\int_{b_{lower}}^{b_{upper}} f_{(1,\lambda)}(x)dx} = epc^{(c,s,b)}$$

where b_{lower} and b_{upper} are the lower and upper bounds for the number of employees of firms in that size bucket: $(0, 9)$, $(10, 19)$, $(20, 50)$, $(50, 150)$, $(150, 250)$, $(250, 1,000,000)$.

For some size groups, $epc^{(c,s,b)}$ is too small, and varying α alone is insufficient to fit the distribution. In such cases, we optimize both α and λ such that:

$$\min_{\alpha^{(c,s,b)}\lambda^{(c,s,b)}} \left(E^b(f_{(\alpha,\lambda)}) - epc^{(c,s,b)}\right)^2.$$

This procedure yields an $(\alpha^{(c,s,b)}, \lambda^{(c,s,b)})$ pair for each size bucket. For each size bucket, we then sample the number of employees of each company as

$$empl_1^{(c,s,b)}, ..., empl_{nc}^{(c,s,b)} \sim \text{Pareto II}\left(\alpha^{(c,s,b)}, \lambda^{(c,s,b)}\right).$$

Additional Information on In- and Out-degrees

For each firm-network, the in- and out-degrees are sampled according to equations (2) and (3) in the main text with $\alpha^{\text{network}}out \sim \text{Normal}(0.335, 0.025)$ and $\alpha^{\text{network}}in \sim \text{Normal}(0.7, 0.1)$ following [4]. We then calculate the out-degree for each company as

$$(k_{in})_i^{(c,s,b)} \sim \left((k_{out})_i^{(c,s,b)}\right)^{\alpha_{in}^{\text{network}}}$$

$$(k_{out})_i^{(c,s,b)} = \bar{k}_{out} \cdot N \cdot \left[\sum_{i,(c,s,b)} \left(rev_i^{(c,s,b)}\right)^{\alpha_{out}^{\text{network}}}\right]^{-1} \cdot \left(rev_i^{(c,s,b)}\right)^{\alpha_{out}^{\text{network}}},$$

and the in-degree subsequently as

$$(k_{in})_i^{(c,s,b)} \sim \left((k_{out})_i^{(c,s,b)}\right)^{\alpha_{in}^{\text{network}}}$$

$$(k_{in})_i^{(c,s,b)} = \bar{k}_{in} \cdot N \cdot \left[\sum_{i,(c,s,b)} \left(rev_i^{(c,s,b)}\right)^{\alpha_{in}^{\text{network}}}\right]^{-1} \cdot \left((k_{out})_i^{(c,s,b)}\right)^{\alpha_{in}^{\text{network}}},$$

where $N = \sum_{(c,s,b)} nc^{(c,s,b)}$ is the total number of companies. The average in- and out-degree are set as $\bar{k}_{out} = 30$ and $\bar{k}_{in} = 33$ according to [4]. The in-degree is set 10% higher than the out-degree to account for imports.

Reconstruction Algorithm

The network reconstruction algorithm iteratively builds a synthetic firm-level network by probabilistically sampling pairs of supplier and buyer sectors and countries based on OECD Input-

Output tables. Individual firms within these sectors and countries are selected proportionally to their predetermined in- and out-degrees, establishing links that represent supplier–customer relationships.[1] The structure of this network is represented by its adjacency matrix A, where $A_{ij} = 1$ if there is a link (supplier–customer relationship) from j to firm i, and $A_{ij} = 0$ otherwise. This process repeats until achieving a target network density, resulting in a detailed network of approximately 30 million companies and 890 million links.

Algorithm 1 Network reconstruction

Require: $PZ((c_1, s_1) \overset{Link}{\leftarrow} (c_2, s_2))$, $\{(k_{in})_i^{(c,s,b)}\}_{i=1,\dots,N}$, $\{(k_{out})_j^{(c,s,b)}\}_{j=1,\dots,N}$

while $\sum_{i,j} A_{ij}/N < 30$ **do**

  Sample $((s_{in}, c_{in}), (s_{out}, c_{out})) \sim PZ_{(c_{in},s_{in}) \leftarrow (c_{out},s_{out})}$

  Sample $i_{in} \sim (k_{in})_{i_{in}}^{(c_{in},s_{in})} / \sum_j (k_{in})_j^{(c_{in},s_{in})}$

  Sample $j_{out} \sim (k_{out})_{j_{out}}^{(c_{out},s_{out})} / \sum_i (k_{out})_i^{(c_{out},s_{out})}$

  if $A_{i_{in},j_{out}} = 0$ **then**

    Set $A_{i_{in},j_{out}} = 1$

    Set $k_{\text{in}}^{i_{in}} -\!= 1$ and $k_{\text{out}}^{j_{out}} -\!= 1$

  end if

end while

return A

The algorithm reconstructs a full inner-EU supply network. Since the EU CSDDD focuses solely on suppliers, we ignore exports from the EU to the rest of world. Imports are captured by using dummy companies identified only by their sector and origin, i.e., each sectoral import link from a non-EU country is represented by exactly one company. These dummy companies are assigned a specific country of origin probabilistically, using sectoral trade flows for the year 2021 from United Nations Comtrade database [7], which were converted from the Harmonized System (HS) codes to NACE codes and OECD sectoral aggregates [8]. Specifically, each dummy company's origin country c_{origin} is assigned according to the importing EU country and sector-based probabilities

$$P(c_{\text{origin}} = c \mid c_{\text{importing}}, s_{\text{origin}}) = \frac{Imp_{c_{\text{importing}}}^{(c,s_{\text{origin}})}}{\sum_{c' \in C} Imp_{c_{\text{importing}}}^{(c',s_{\text{origin}})}} \tag{6}$$

with $Imp_{c_{\text{importing}}}^{(c,s_{\text{origin}})}$ denoting the import value from country c_{origin} in sector s_{origin} to importing country $c_{\text{importing}}$, ensuring that country assignments accurately reflect trade distributions.

Distance in Production Networks

The small-world effect, which is crucial for regulatory design, refers to the empirical observation that even large networks, consisting of millions of nodes, can have a surprisingly short characteristic path length. This is defined as the average length of the shortest path (network distance) between any two firms in the production network [4].[2]

[1] If the sampled connection does not already exist, the algorithm adds it to the network and decreases the respective in- and out-degrees by 1.

[2] When considering networks with complete visibility of all links, as derived from VAT data, the characteristic path length is found to be between three and four. This is shorter than in human contact networks, which exhibit a characteristic distance of around six—the famous "six degrees of separation," where each person on the planet is connected to any other through no more than six contacts. [9]

We demonstrate that the distribution of shortest path lengths in the reconstructed networks is similar to that of real firm-level production networks. To analyze this, we convert each reconstructed network to an undirected version and calculate the fraction of nodes that can be reached from a randomly selected node. This process is repeated for 50 random nodes from each reconstructed network, and we calculate the average, as illustrated in Supplementary Fig. 2.

$$
\begin{aligned}
B_{ij} &= (A_{ij} + A_{ji} + \delta_{ij}) > 0 \qquad (\delta_{ij} = \{1 \text{ for } i = j; 0 \text{ else}\}) \\
\text{Frequency}_k(v) &= \sum_{i=1}^{N} \frac{(B^k v)_i}{N} \\
\text{Average Pathlength} &= \sum_k \text{Frequency}_k \cdot k
\end{aligned}
$$

The shortest path lengths in the reconstructed networks display a similar pattern to those in real-world networks: approximately 60% of nodes are at a minimum distance of 3 from each other, and the average path length is calculated to be 3.31 with a standard deviation of 0.31, consistent with the findings in the literature [1].

Human-rights risk mapping

Table 1: **Supplementary Table 1.** List of OECD sectors in non-European Union countries identified as risk-flagged, based on the List of Goods Produced by Child and Forced Labor from the United States Department of Labor and the HR allegation lawsuits database from the Business and Human Rights Resource Centre.

Country	Child and Forced Labor	Business and Human Rights
AFG	D01T02, D05T06, D07T08, D13T15, D23	D51
AGO	D23	
ARG	D01T02, D10T12, D13T15, D23	D05T06, D19, D20, D29
AUS		D05T06, D07T08, D09, D45T47
AZE	D13T15	
BEN	D07T08, D13T15, D23	
BFA	D07T08, D13T15, D23, D24	
BGD	D01T02, D07T08, D10T12, D13T15, D20, D23, D31T33	D05T06, D13T15, D19, D45T47
BLZ	D01T02	
BOL	D01T02, D07T08, D10T12, D23, D24	
BRA	D01T02, D10T12, D13T15, D16, D23	D01T02, D03, D05T06, D07T08, D09, D10T12, D19, D21, D29, D35, D41T43, D68, D69T75
CAF	D23	
CAN		D05T06, D07T08, D09
CHL		D05T06, D07T08, D09, D10T12
CHN	D01T02, D05T06, D10T12, D13T15, D20, D23, D25, D26, D31T33	D01T02, D03, D05T06, D07T08, D09, D10T12, D19, D20, D26, D31T33, D58T60, D64T66
CIV	D01T02, D10T12	D01T02, D03, D05T06, D10T12, D19, D35, D53
CMR	D01T02, D07T08, D10T12, D24	D05T06, D07T08, D09, D25
COD	D07T08, D23, D24, D25, D27	D01T02, D05T06, D07T08, D09, D25, D26
COL	D01T02, D05T06, D07T08, D10T12, D23, D24, D31T33, D58T60	D01T02, D03, D05T06, D10T12, D19, D20, D25, D77T82
CRI	D01T02, D10T12	
CUB		D30, D51
DOM	D01T02, D10T12	
ECU	D01T02, D07T08, D10T12, D23, D24	D05T06, D19
EGY	D07T08, D13T15, D23, D31T33	D05T06, D19, D41T43, D51
ERI		D05T06, D07T08, D09
ETH	D01T02, D07T08, D10T12, D13T15, D24	
FRA		D01T02, D05T06, D19, D35, D49
GBR		D01T02, D03
GHA	D01T02, D03, D07T08, D10T12, D13T15, D24	
GIN	D01T02, D07T08, D10T12, D23, D24	D05T06, D07T08, D09
GTM	D01T02, D07T08, D10T12, D20, D23	D05T06, D07T08, D09, D35
GUY		D05T06, D07T08, D09
HND	D01T02, D03, D10T12	
IDN	D01T02, D07T08, D10T12, D13T15, D22, D24	D03, D05T06, D07T08, D09, D13T15, D19
IND	D01T02, D07T08, D10T12, D13T15, D20, D23, D25, D31T33	D05T06, D07T08, D09, D10T12, D20, D21, D64T66
IRN	D13T15, D23	
IRQ		D77T82
ISR		D26, D49, D62T63, D64T66, D77T82
ITA		D13T15, D23, D41T43
JPN		D24, D25, D41T43
KAZ	D13T15	

Continued on next page

Country	Child and Forced Labor	Business and Human Rights
KEN	D01T02, D07T08, D10T12, D13T15, D23, D24	D01T02, D05T06, D10T12, D19, D25, D26
KGZ	D01T02, D10T12, D13T15	
KHM	D01T02, D07T08, D10T12, D13T15, D16, D22, D23	D01T02, D10T12, D22
LBN	D01T02, D10T12	
LBR	D22, D23	D22
LBY		D25, D26, D77T82
LSO	D01T02, D10T12	
MAR		D51
MDG	D01T02, D07T08, D10T12, D23, D31T33	
MEX	D01T02, D10T12, D13T15, D58T60	D05T06, D07T08, D09, D35
MLI	D01T02, D07T08, D10T12, D13T15, D24	
MMR	D01T02, D07T08, D10T12, D13T15, D16, D22, D23, D31T33	D05T06, D07T08, D09, D19
MNG	D05T06, D07T08, D24	
MOZ	D01T02, D10T12	
MRT	D01T02, D10T12	
MWI	D01T02, D10T12	D05T06, D07T08, D09
MYS	D01T02, D10T12, D13T15, D22, D26	
NAM		D64T66
NER	D01T02, D07T08, D10T12, D24	D05T06, D07T08, D09
NGA	D01T02, D07T08, D10T12, D23, D24	D03, D05T06, D07T08, D09, D19, D21, D69T75
NIC	D01T02, D03, D07T08, D10T12, D23, D24	
NOR		D05T06, D19
NPL	D13T15, D23	
PAK	D01T02, D05T06, D10T12, D13T15, D23, D26, D31T33	D13T15
PAN	D01T02, D10T12	
PER	D01T02, D07T08, D10T12, D16, D20, D23, D24	D05T06, D07T08, D09, D19, D35
PHL	D01T02, D07T08, D10T12, D13T15, D20, D22, D24, D58T60	
PNG		D05T06, D07T08, D09, D19
POL		D27
PRK	D05T06, D07T08, D13T15, D16, D23, D24, D25, D26	
PRY	D01T02, D07T08, D10T12, D23, D31T33, D58T60	
PSE		D41T43
QAT		D41T43, D45T47, D90T93
RUS	D16, D23, D58T60	D24, D25, D45T47, D49
RWA	D01T02, D10T12	
SDN	D07T08, D24	D05T06, D07T08, D09, D19, D25, D35, D69T75
SEN	D07T08, D24	
SLE	D01T02, D07T08, D10T12, D23	D05T06, D07T08, D09
SLV	D01T02, D03, D10T12, D20	
SSD	D01T02, D10T12	D05T06, D19, D35
SUR	D07T08, D24	
SWZ	D01T02	
SYR		D23, D26, D41T43
TCD	D01T02, D10T12	
THA	D01T02, D10T12, D13T15, D58T60	D01T02, D03, D10T12, D35, D45T47
TJK	D13T15	
TKM	D13T15	
TUR	D01T02, D10T12, D13T15, D31T33	
TWN	D01T02, D10T12	D26

Continued on next page

Country	Child and Forced Labor	Business and Human Rights
TZA	D01T02, D03, D07T08, D10T12, D13T15, D24, D31T33	D05T06, D07T08, D09, D55T56, D77T82
UGA	D01T02, D07T08, D10T12, D23, D24	D01T02, D05T06, D10T12, D19, D35, D41T43, D58T60
UKR	D05T06, D07T08, D20, D23, D58T60	
USA		D01T02, D05T06, D07T08, D09, D10T12, D13T15, D19, D20, D23, D25, D26, D29, D30, D31T33, D41T43, D45T47, D49, D62T63, D64T66, D69T75, D77T82, D84, D90T93
UZB	D13T15	
VEN	D07T08, D24	
VNM	D01T02, D10T12, D13T15, D16, D22, D23, D31T33	D03, D13T15, D20, D24, D25
YEM	D01T02, D10T12	
ZAF		D05T06, D07T08, D09, D23, D29, D51
ZMB	D01T02, D07T08, D10T12, D13T15, D31T33	D05T06, D07T08, D09, D19, D84
ZWE	D01T02, D07T08, D10T12, D24	D23

Supplementary Figures

Properties of the Reconstructed Network

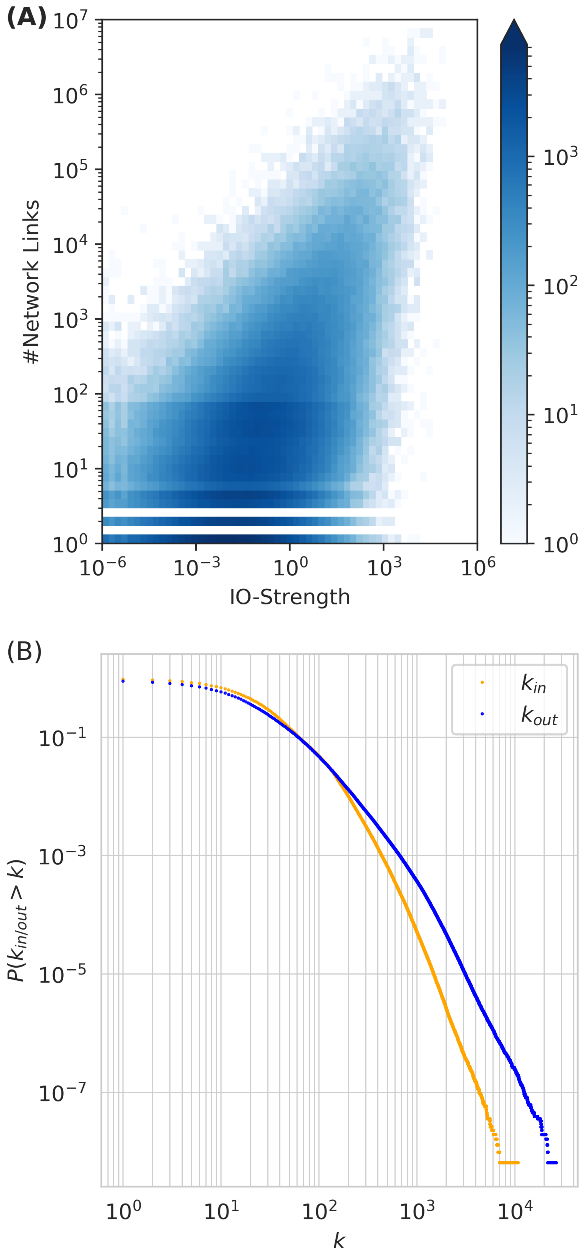

Figure 1: **Supplementary Figure 1.** Characteristics of the reconstructed network. The figure shows the strength between the (sector,country) pairs in the OECD IO-table against the number of links in the reconstructed networks (Panel A) and the cumulative probability density function of the in-/out-degree of the reconstructed networks (Panel B).

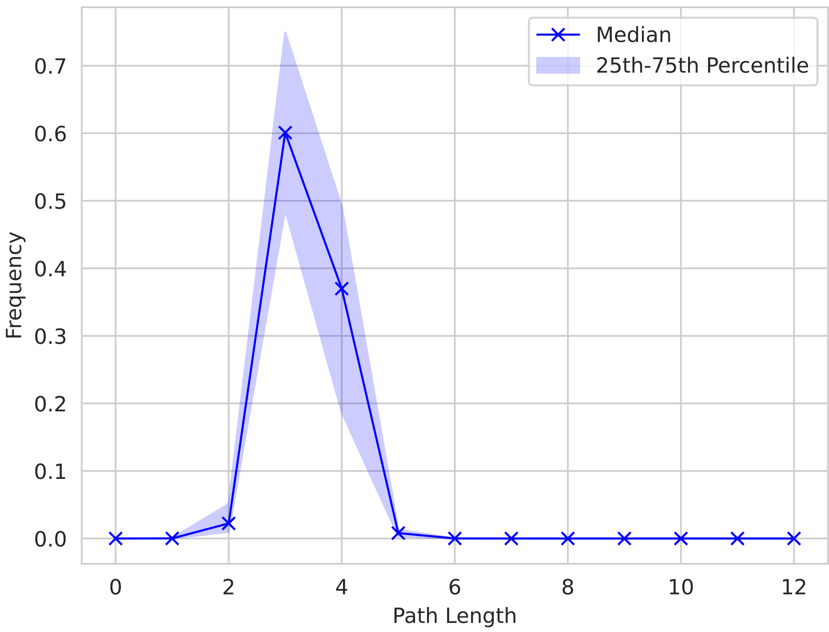

Figure 2: **Supplementary Figure 2.** Distribution of the shortest path lengths between pairs of random nodes in the reconstructed networks.

We converted the network to an undirected format and, for each of the 10 network realizations, selected 50 random nodes. For each selected node, we calculated the number of nodes reached as a function of path length.

Risk Indicators

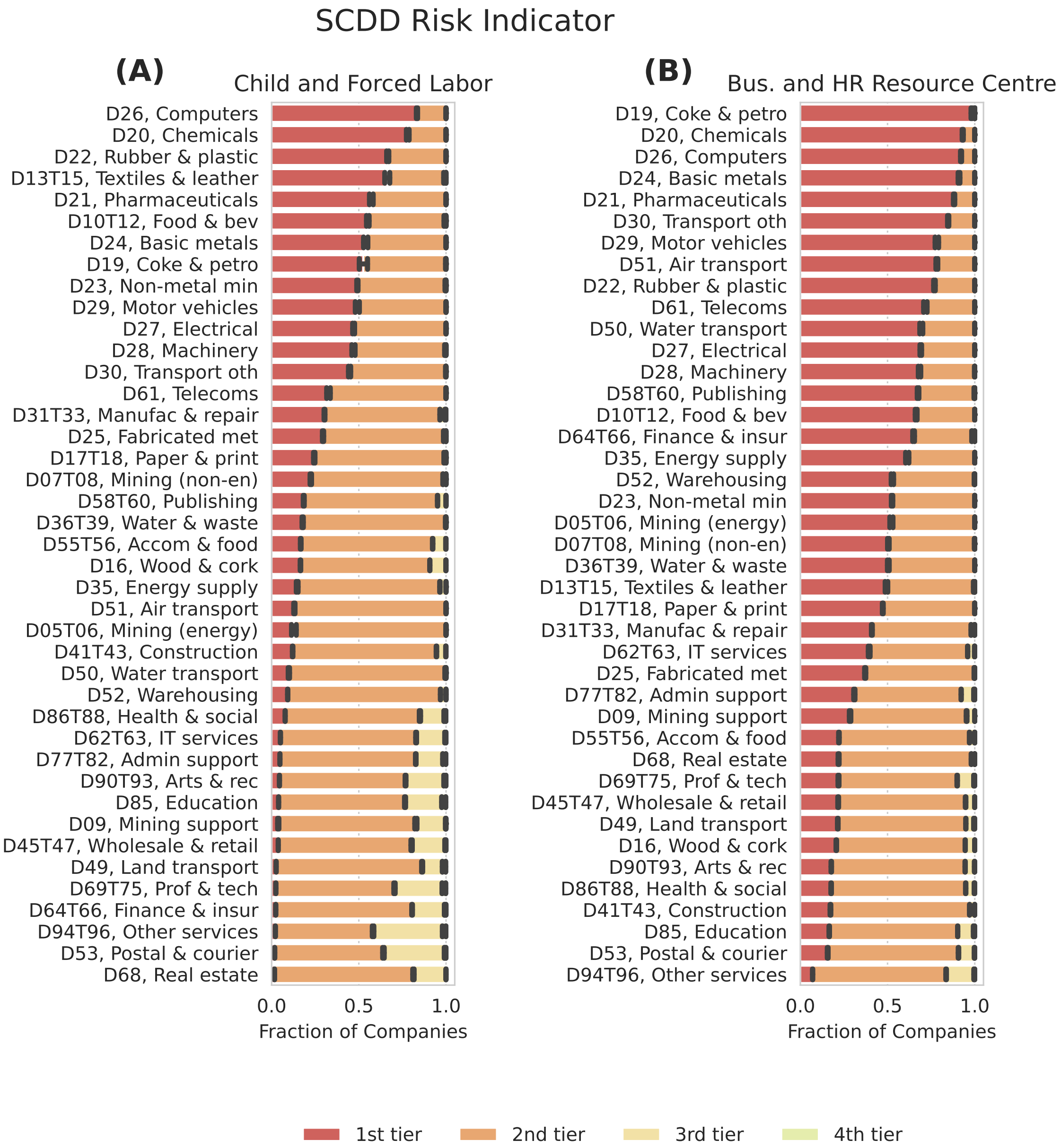

Figure 3: **Supplementary Figure 3.** Results for the SCDD Risk Indicator on a sector level. Results are shown for indicators on tier one to four (red to yellow colors) for (A) allegations of child and forced labor and (B) lawsuits. Rankings of sectors by tier one indicators are similar across both datasets, with highest values for computers, basic metals, chemicals, pharmaceuticals, motor vehicles and textiles. Considering the 4th or higher tiers, all indicators are close to 100%.

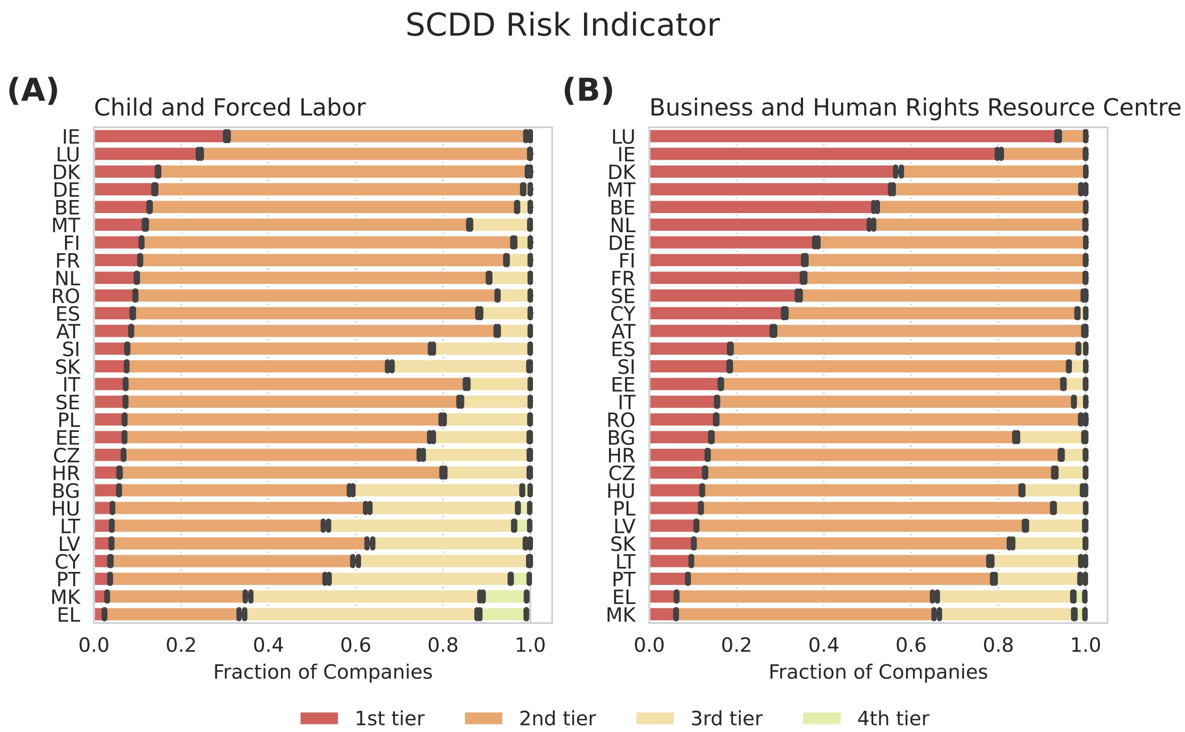

Figure 4: **Supplementary Figure 4.** Results for the SCDD Risk Indicator at country level. Results are shown for indicators on tier one to four (red to yellow colors) for (A) allegations of child and forced labor as well as (B) lawsuits collected by the Business and Human Rights Resource Centre. The indicator is shown only up to the 4th tier, where it is approaching 100% for most countries.

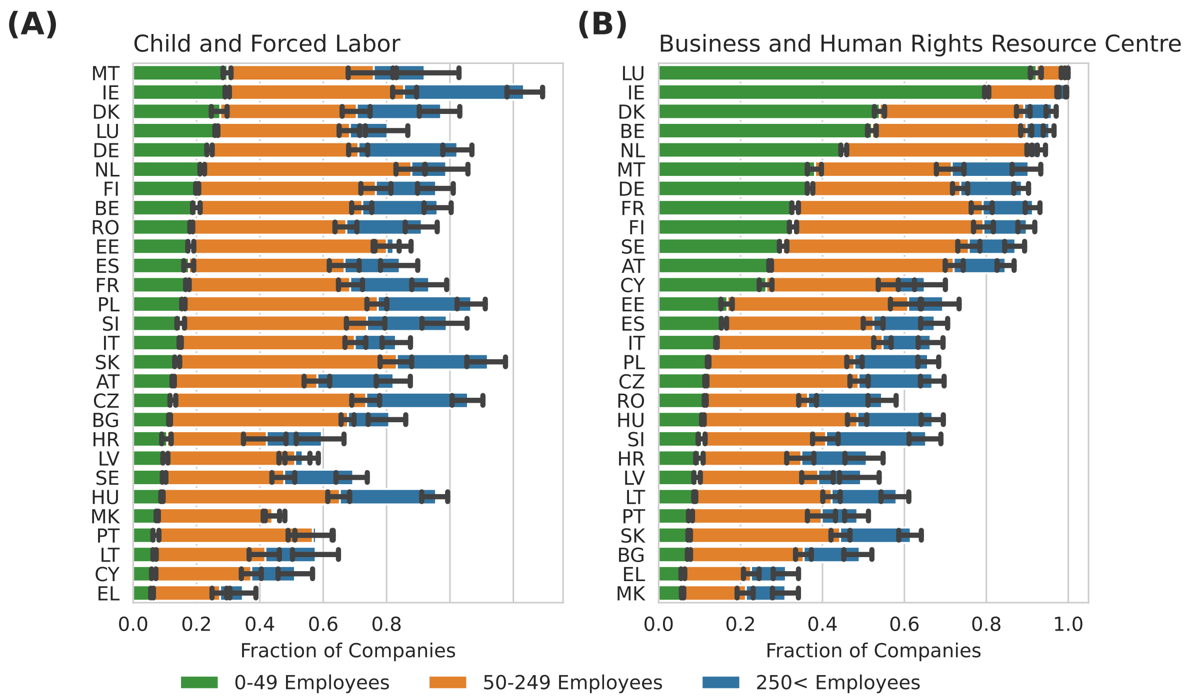

Figure 5: **Supplementary Figure 5.** Results for the SCDD Risk Indicator by country and firm size.

Results for the SCDD Risk Indicator at the first tier by countries on tier one, stratified by the number of employees per company. Companies with 1–50 employees show substantially lower risks than larger companies for (a) child and forced labour allegations and (b) HR-related lawsuits.

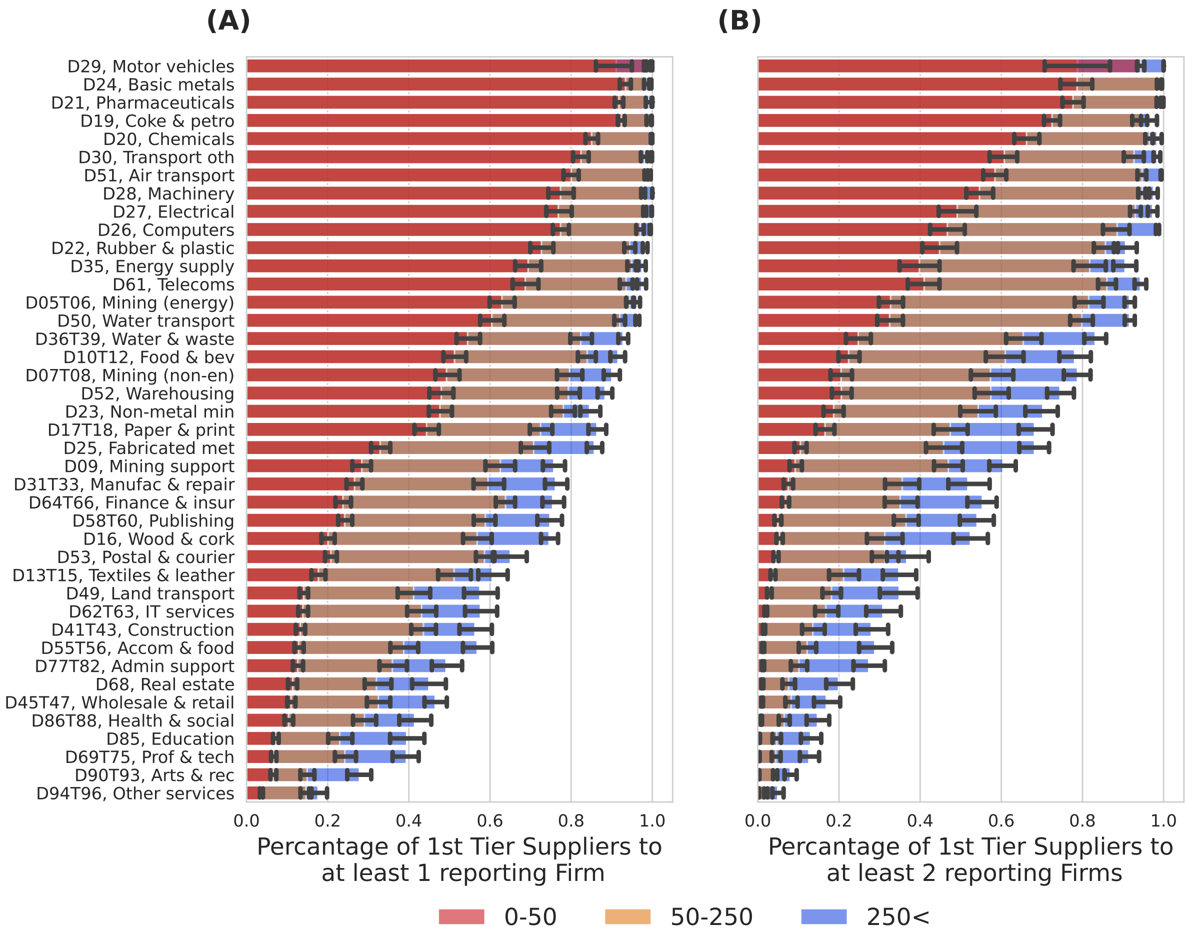

Figure 6: **Supplementary Figure 6.** Percentage of Tier 1 suppliers connected to reporting firms, stratified by supplier company size and sector. Panel (A) shows suppliers connected to at least one reporting firm. Panel (B) shows suppliers connected to at least two reporting firms. Company size categories are 0-50, 50-250, and >250 employees.

References

1. Bacilieri, A. & Austudillo-Estevez, P. Reconstructing firm-level input-output networks from partial information. *arXiv preprint arXiv:2304.00081* (2023).
2. Karrer, B. & Newman, M. E. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **83,** 016107 (2011).
3. Ialongo, L. N. *et al.* Reconstructing Firm-Level Interactions in the Dutch Input–Output Network from Production Constraints. *Scientific Reports* **12,** 11847. ISSN: 2045-2322. (2023) (2022).
4. Bacilieri, A., Borsos, A., Astudillo-Estevez, P. & Lafond, F. Firm-level production networks: what do we (really) know? *INET Oxford Working Paper* **May 2023.** https://www.inet.ox.ac.uk/files/Firm-level-production-networks-what-do-we-really-know.pdf (2023).
5. Borsos, A. & Stancsics, M. *Unfolding the hidden structure of the Hungarian multi-layer firm network* tech. rep. (MNB Occasional Papers, 2020).
6. Axtell, R. L. Zipf distribution of US firm sizes. *science* **293,** 1818–1820 (2001).

7. United Nations. *UN Comtrade Database: Exports* 2021. https://comtradeplus.un.org.
8. United Nations. *Classifications on economic statistics* 2021. https://unstats.un.org/unsd/classifications/Econ.
9. Leskovec, J. & Horvitz, E. *Planetary-scale views on a large instant-messaging network* in *Proceedings of the 17th international conference on World Wide Web* (2008), 915–924.