Fetzer, Thiemo; Lambert, Peter John; Feld, Bennet; Garg, Prashant

Working Paper

AI-Generated Production Networks: Measurement and Applications to Global Trade

CESifo Working Paper, No. 11497

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Fetzer, Thiemo; Lambert, Peter John; Feld, Bennet; Garg, Prashant (2024) : AI-Generated Production Networks: Measurement and Applications to Global Trade, CESifo Working Paper, No. 11497, CESifo GmbH, Munich

This Version is available at:
https://hdl.handle.net/10419/308393

CESifo WORKING PAPERS

11497
2024

November 2024

AI-Generated Production Networks: Measurement and Applications to Global Trade

Thiemo Fetzer, Peter John Lambert, Bennet Feld, Prashant Garg

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11497*

AI-Generated Production Networks: Measurement and Applications to Global Trade

Abstract

This paper leverages generative AI to build a network structure over 5,000 product nodes, where directed edges represent input-output relationships in production. We layout a two-step 'build-prune' approach using an ensemble of prompt-tuned generative AI classifications. The 'build' step provides an initial distribution of edge-predictions, the 'prune' step then re-evaluates all edges. With our AI-generated Production Network (AIPNET) in toe, we document a host of shifts in the network position of products and countries during the 21st century. Finally, we study production network spillovers using the natural experiment presented by the 2017 blockade of Qatar. We find strong evidence of such spill-overs, suggestive of on-shoring of critical production. This descriptive and causal evidence demonstrates some of the many research possibilities opened up by our granular measurement of product linkages, including studies of on-shoring, industrial policy, and other recent shifts in global trade.

JEL-Codes: F140, F230, L160, F520, O250, N740, C810.

Keywords: supply-chain network analysis, large language models, on-shoring, industrial policy, trade wars, econometrics-of-LLMs.

Thiemo Fetzer
Warwick University / United Kingdom
t.fetzer@warwick.ac.uk

Peter John Lambert
London School of Economics and Political Science (LSE), London / United Kingdom
p.j.lambert@lse.ac.uk

Bennet Feld
London School of Economics and Political Science (LSE), London / United Kingdom
b.l.feld@lse.ac.uk

Prashant Garg
Imperial College London / United Kingdom
prashant.garg@imperial.ac.uk

November 17, 2024
Click here for the latest version:
https://drive.google.com/file/d/1M-fbFOax98MlF5H_RHmNmknYbHpL3pi4/view?usp=drive_link
Fetzer acknowledges support through the Leverhulme Trust Prize in Economics, the European Research Council Grant (ERC, MEGEO, 101042703), and Deutsche Forschungsgemeinschaft (EconTribute, DFG, EXC 2126/1-390838866).

1 Introduction

Global economic integration—a defining feature of the late 20th century—faces mounting structural challenges as nations navigate rising protectionism, pandemic-induced disruptions, proliferating industrial policies, and conflict-related supply chain shocks. Evidence of these challenges includes growing trade restrictions, the paralysis of WTO dispute mechanisms, industrial policy announcements, and persistent supply chain disruptions (see Figure 1). Against this backdrop of fragmentation, technological change continues to deepen the complexity of production processes, necessitating ever-greater economic interdependence.

This paper develops and introduces the *AI-generated Production Network* (AIPNET) to help unpack these contrasting shifts in trade and production. Our measurement approach utilizes frontier generative AI to recover granular network structures spanning more than 5,000 product categories.[1] The network maps each product's position in the production process by identifying its input-output relationships to other products. This paper uses these product inter-dependencies to analyze recent shifts in global trade and production but also contributes this network dataset to facilitate wider interdisciplinary analysis.[2]

The construction of this granular production network harnesses recent advances in generative AI. We propose a two-step 'build-prune' methodology to implement generative AI towards the construction of graphical network data. Both steps employ an ensemble of prompt-tuned generative AI classifications. The initial 'build' step generates a distribution of edge-predictions, identifying potential connections at the possible expense of spurious links. The subsequent 'prune' step rigorously evaluates these edges, ensuring fidelity to the underlying latent structures of interest. While we develop this

[1] Our categories are based on the 'harmonized system', a nomenclature of product codes maintained by the World Customs Organization (WCO) and used universally by national trade authorities. These 5,000+ classifications are highly granular, e.g. 'Milk And Cream Of A Fat Content, By Weight, Exceeding 10%, Not Concentrated Nor Containing Added Sugar Or Other Sweetening Matter' (HS Code 0401.50).

[2] Researchers can download and utilize AIPNET, available at aipnet.io.

methodology specifically for production networks, it offers broader applicability in settings where nodes are known but edges require discovery through information embedded in language models.

Our application of this measurement framework to recover production linkages across products yields important insights into global trade patterns. We find a sizable shift over the last decade towards global trade of more 'central' upstream products, both in terms of their overall position in the network and based on the number of downstream connections of each good. We also highlight divergent trends across countries, finding that China has increased the relative import intensity of more upstream products. In contrast, the US has shifted to a more downstream-intensive import mix. Globally, we see a rise in more upstream goods which sit atop the value chain, especially in high-tech products like chips and electronic circuitry, as well as critical minerals for modern technologies such as lithium.

To more fully leverage our network structure, we examine patterns of import substitution across vertically linked products. Using unit price indices constructed for each country-product pair, we identify persistent supply shocks that manifest through sustained price increases. These structural breaks in unit prices—ranging from trade disputes and shortages to geopolitical events—emerge as key drivers of changes in import patterns, particularly in upstream goods linked to affected products.

This systematic analysis of global trade reveals clear evidence of 'onshoring' or production localization. When countries face significant disruptions to downstream products, they respond by increasing imports of vertically connected upstream goods.

To establish clean causal identification of this 'onshoring' mechanism, we lastly analyze an unanticipated supply shock which came about during the 2017 blockade of Qatar. Exploiting the shock's exogenous timing and substantial variation in product-specific exposure, we implement a dyadic-difference-in-difference (DyDiD) specification using each connection in our network as a single unit of observation. This allows for a granular design to measure substitution patterns from vertically linked upstream capital

and intermediaries and away from downstream consumer-facing goods. We find that products which were sourced wholly from blockading countries prior to this event saw a 44 percent decrease in imports in the five years post-blockade. More interestingly, for the same exposed products, the vertically connected upstream products saw imports *increase* by 18 percent. This evidence suggests the Qatar blockade fundamentally altered the "make-or-buy" calculus, resulting in shifts up the import-value-chain to facilitate domestic production.

This work contributes to several strands of literature. First, we advance the measurement of production and trade networks. As Johnson (2018) discusses in a survey of global value chains measurement, the field has relied on two main approaches: country-level Input-Output tables and firm-to-firm transaction records. While Input-Output tables offer comprehensive economic overviews and broad cross-country coverage, they lack granularity. Transaction-level data provides granular firm-level insights but has limited availability and coverage. Moreover, firm boundaries may not accurately capture production steps, and intermediaries in international trade can lead to underestimating firms' trade exposure (Dhyne et al., 2020). In contrast to these standard approaches, the AIPNET product this paper introduces provides a publicly available and highly granular measure of product-to-product connections, focused on the underlying technologies and production processes rather than economic transactions.

Several other authors have also sought more novel approaches to bridge the gap between granular transaction data and IO tables. Karbevska and Hidalgo (2023) and Andersen et al. (2022), study geographical and product category expansions. Frésard, Hoberg, and Phillips (2020) measure firm vertical relatedness using Input-Output tables and product text from SEC filings.

An important contribution to measuring production-network features focus on locating countries and industries in terms of their 'upstreamness' i.e. distance to final consumption by leveraging variation in input-output tables (Fally, 2012; Antràs et al.,

2012; Antràs & Chor, 2018).[3] These studies complement our more granular and methodologically distinct efforts to characterize production networks and measure centrality of products using AIPNET.

Second, we contribute to the emerging literature on LLM and AI applications in economic measurement. For comprehensive introductions to LLMs in economics, see Ash and Hansen (2023), Dell (2024), Giesecke (2024), and Korinek (2023). Most applications use LLMs to structure and classify data, such as analyzing remote work in job postings (Hansen et al., 2023), measuring housing regulation (Bartik, Gupta, & Milo, 2023), extracting information from loan documents (Schindler & Lambert, 2024) or extracting causal claims from scientific publications (Garg & Fetzer, 2024). Our approach extends beyond classification to leverage LLMs' reasoning capabilities for knowledge graph construction.[4]

Third, we contribute to research on production networks' responses to shocks. This literature builds on theoretical work modeling cascade effects, substitutability, and supply dependencies.[5] Related theoretical foundations come from work on export decisions (Melitz, 2003; Krugman, 1980)[6] and sourcing decisions.[7] Studies of systemic shocks document widespread trade flow rewiring.[8]

Our focus on upstream responses complements work on downstream effects by Boehm et al. (2019) and Acemoglu et al. (2015). This connects to literature examining major trade disruptions including COVID-19 (Freeman & Baldwin, 2020; Antràs, 2020), Brexit,[9] and the 2008 financial crisis (Behrens et al., 2013; Bricongne et al., 2012).

[3]While our measurement is based on AI/text, Feenstra and Jensen (2012) highlights that the proportionality assumption embedded in the construction of IO-tables can drive mechanical correlations between 'upstreamness' and 'downstreamness' at the county-level.

[4]For broader discussion of LLMs in knowledge graph construction and applications to supply chains, see (Peifeng et al., 2024), (Pan et al., 2024), (Jewson et al., 2022), (Buehler, 2024) and (Kosasih et al., 2024).

[5]For foundational work on production network modeling, see Elliott, Golub, and Leduc (2022), Baqaee and Farhi (2019), Oberfield (2018), (Carvalho & Gabaix, 2013), and Acemoglu et al. (2012).

[6]For a comprehensive survey of export theory and evidence, see Bernard et al. (2012).

[7]Key contributions on sourcing and offshoring include Huang et al. (2024), Grossman, Helpman, and Redding (2024), Bernard et al. (2018), Antràs, Fort, and Tintelnot (2017), and Garetto (2013).

[8]For analysis of systematic trade network responses to shocks, see Elliott and Jackson (2024), Baldwin and Freeman (2022), Caselli et al. (2020), Bachmann et al. (2022), and Baqaee and Farhi (2021).

[9]For comprehensive analysis of Brexit's trade impacts, see Freeman et al. (2022), Alabrese, Fetzer, and

Finally, our Qatar blockade analysis contributes to research on geopolitically motivated trade decoupling. Antràs (2020) discusses recent trade slowdowns as responses to uncertainty rather than deglobalization. Others document reduced East-West trade and FDI flows owing to global conflicts (Gopinath et al., 2024; Blanga-Gubbay & Rubínová, 2023). A growing literature examines US-China decoupling.[10]

The remainder of this paper is organized as follows: Section 2 lays out a framework for measuring graphical network data using generative AI, Section 3 introduces AIPNET and discusses our methodology for construction and validation. Section 4 presents global trends and tests the micro-foundations of onshoring. Section 5 leverages a dyadic difference-in-difference (DyDiD) econometric design to provide a causal analysis of onshoring following the Qatar blockade. Section 6 concludes.

2 Building Graphs with Generative AI

We begin with a high level outline of this papers approach to building network datasets (i.e., graphical data) using generative AI. This section lays out a framework that can be applied in cases where the set of nodes is known, but the edges connecting them are unknown and must be estimated. Those readers who prefer a more applied articulation of our approach may wish to proceed to the subsequent section 3, where we directly discuss the construction of AIPNET and outline the practical steps necessary to apply the framework introduced below.

Wang (2024), Costa, Dhingra, and Machin (2019), P. D. Fajgelbaum et al. (2020), Hassan et al. (2024), Born et al. (2018), Douch et al. (2018), and Breinlich et al. (2017).

[10]For recent analysis of US-China trade relations and decoupling, see Handley and Limão (2017) on policy uncertainty, and Alfaro and Chor (2023), Crosignani et al. (2023), Freund et al. (2023), and Utar, Torres Ruiz, and Zurita (2023) on nearshoring evidence; Fetzer and Schwarz (2021) explore the high-dimensional strategic considerations in engaging in trade wars.

2.1 Estimating Graph Edges

Our goal is to estimate a latent network $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the pre-existing set of n nodes, and E is the set of unobserved edges connecting these nodes. Each node has at least one observable property, such as a name or label.[11] Our approach to estimating edges E follows two key steps. First, we 'build' the network by asking a generative AI tool *node*-specific questions.[12] Second, we 'prune' the network by asking *edge-level* questions.[13] The remainder of this section lays out this approach more formally.

Step 1 'Build' In this step, we use a Large Language Model (LLM) to answer questions q about nodes v. The questions are designed such that their answers identify the set of potential connections. Let f_{LLM} be a function which represents this question- answering capability:

$$f_{\text{LLM}} : V \times Q \times \Theta \to \mathcal{P}(V)$$

where V is the set of all nodes, Q is the set of all possible questions or prompts, Θ is the set of possible model parameters, and $\mathcal{P}(V)$ is the power set of V, representing all possible subsets of nodes.[14]

For a given node $v_i \in V$, we compute:

$$C_i = f_{\text{LLM}}(v_i, q, \theta),$$

where $q \in Q$ is a specific question or prompt designed to elicit connected nodes, $\theta \in$

[11] To economize on notation, we use v_i to refer both to the node it's self, and also it's observable properties.

[12] An example of a node-specific query is: *'What kinds of machines do I need to produce Skim Milk?'*.

[13] Examples of an edge-specific question: *'Do I need Dairy Cows to produce Skim Milk?'*

[14] More formally, Equation 2.1 defines a parameterized family of set-valued functions. For fixed $(q, \theta) \in Q \times \Theta$, each $f_{q,\theta} : V \to \mathcal{P}(V)$ is a set-valued function mapping the vertex set to its power set. This formulation allows us to characterize the entire space of possible network configurations induced by varying questions and model parameters, while maintaining the fundamental node-to-subsets mapping structure.

Θ represents the current state of the LLM's parameters[15], $C_i \subseteq V$ is the set of nodes potentially connected to v_i.

The function f_{LLM} can be helpfully represented as the maximizer of a probability conditional on the initial question or 'prompt':

$$f_{\text{LLM}}(v_i, q, \theta) = \arg\max_{C \subseteq V} P(C \mid v_i, q, \theta),$$

where $P(C \mid v_i, q, \theta)$ is the probability distribution over subsets of V, conditioned on the input node, question, and model parameters.

To incorporate uncertainty and create a probabilistic edge measure, we run the LLM multiple times:

$$C_i^{(k)} = f_{\text{LLM}}(v_i, q_k, \theta_k), \quad k = 1, 2, \ldots, K,$$

where K is the number of runs, q_k and θ_k represent potentially varying questions and model parameters for each run[16].

We then define an edge overlap function $p : V \times V \to [0, 1]$:

$$p(v_i, v_j) = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\{v_j \in C_i^{(k)}\}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, equal to 1 if, for a specific kth iteration of the LLM function, the two vertices v_i and v_j are connected, and zero otherwise. This provides a measure of the average number of times across K draws two nodes are deemed to be connected.[17].

[15]We will be more explicit about parameters available in practice in the subsequent section. These parameters govern the behavior of the LLM, for example one must set a 'seed' value to govern the models behaviour. One must also set a 'temperature' value, which defines a penalty applied to more creative responses.

[16]In practice, we want to have many more runs than we have possible combination of parameters q_k and questions q_k

[17]This is the simplest way to aggregate over the K responses. One can also consider giving different draws more/less weight, or combining the results of each draw in other parametric ways, possibly with fine-tuned parameter values based on some known set of edges. This is discussed more in Section 2.2

Finally, we create an initial edge set $\tilde{E}$ using a cutoff $\tau \in [0,1]$:

$$\tilde{E} = \left\{ (v_i, v_j) \mid p(v_i, v_j) \geq \tau, \quad i \neq j \right\}.$$

This set $\tilde{E}$ represents our initial set of discrete edges based on the probabilistic output of multiple distinct calls to an LLM.

Step 2: Prune In this step, we refine the initial edge set $\tilde{E}$ by evaluating the relevance or importance of each edge using another LLM-based function[18]. Let g_{LLM} represent the LLM's edge evaluation capability:

$$g_{\text{LLM}} : V \times V \times [0,1] \times Q' \times \Theta' \to \mathbb{R},$$

where $V \times V$ represents pairs of nodes forming potential edges, $[0,1]$ is the domain for the edge probability from the build step, Q' is the set of all possible questions or prompts for edge evaluation, Θ' is the set of possible model parameters for the pruning LLM. For each edge $(v_i, v_j) \in \tilde{E}$, we compute a relevance score:

$$r_{ij} = g_{\text{LLM}}(v_i, v_j, p(v_i, v_j), q', \theta'),$$

where $p(v_i, v_j)$ is the edge probability from the build step, $q' \in Q'$ is a specific question or prompt designed to evaluate edge relevance, $\theta' \in \Theta'$ represents the current state of the pruning LLM's parameters.

To account for potential variability in the LLM's output, we perform multiple evaluations:

$$r_{ij}^{(l)} = g_{\text{LLM}}(v_i, v_j, p(v_i, v_j), q'_l, \theta'_l), \quad l = 1, 2, \ldots, L,$$

where L is the number of evaluation runs, and q'_l and θ'_l represent potentially varying

[18]For cases where the set of all possible edges is small, one might jump straight to this step and initialize the edge set $\tilde{E}$ by using the complete network

questions and model parameters for each run.

We then calculate the average relevance score:

$$\bar{r}_{ij} = \frac{1}{L} \sum_{l=1}^{L} r_{ij}^{(l)}.$$

We determine the final edge set E using one of two methods: **Threshold-based pruning**:

$$E = \left\{ (v_i, v_j) \in \tilde{E} \mid \bar{r}_{ij} \geq \phi \right\},$$

where $\phi \in \mathbb{R}$ is a predefined threshold.

Top-k pruning:

$$E = \text{Top-}k \left(\left\{ \left((v_i, v_j), \bar{r}_{ij} \right) \mid (v_i, v_j) \in \tilde{E} \right\} \right),$$

where k is the desired number of edges to retain, and Top-k selects the k edges with the highest relevance scores.

The resulting graph $G = (V, E)$ represents our final constructed network dataset, with edges pruned based on their evaluated relevance.

2.2 Higher-Dimensional Extensions and Fine-Tuning

The build-prune method described above can be extended to incorporate higher-dimensional information and fine-tuned for specific applications. By leveraging multiple parameter types and question variations, we can create a richer representation of potential edges in the network.

In the build step, instead of using a single question q, we can employ a set of diverse questions $Q = \{q_1, q_2, \ldots, q_m\}$, each designed to probe different aspects of node relationships. Similarly, we can vary model parameters $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ to capture different model behaviors. The prune step can be similarly extended. This approach generates a multi-dimensional feature space for each potential edge, and can be tailored

for many applications including social networks, professional networks, cross-linkages between economic and other classifications structures, and so on.

The rich data generated from these higher-dimensional build and prune processes can then be used to fine-tune a final edge classifier function. This has at least two desirable properties - it ensures edge classification is robust to variation in the way the LLM is utilized, and also allows recovery of edges to mirror any existing proxies for node connections, allowing for the incorporation of domain specific knowledge. One can also extend this process using agentic AI processes, such that the parameters and questions used to recover linkages can be automatically adjusted based on information that is generated in prior steps.

3 AI-generated Production Networks (AIPNET)

Our main applied objective in this paper is to build, test, and validate an AI-generated Production Network (AIPNET), which is able to recover connections between products based on input/output relationships. We follow the build-prune approach laid out in Section 2 above. This approach uses a number of generative AI tools, which are combined in a style similar to ensemble methods in machine learning.

3.1 Nodes in the Production Network

We begin with the Harmonized System (HS) product codes, which define over 5,000 product nodes.[19] This structure provides a system for categorizing goods. Each category has both a numeric code as well as a free-text label describing the product (see examples in Appendix Table A.1).

Most people know that you need a printing press to print a newspaper. Far fewer people can list the enzymes used in synthesizing insulin, but this information is acces-

[19]The HS is a hierarchical classification for international trade, organized into increasingly specific levels. It begins with 21 broad sections, then narrows to 99 chapters (2-digit codes), about 1,244 headings (4-digit), and approximately 5,224 subheadings (6-digit).

sible in seconds through a simple web search. For most, if not all, of the product nodes contained within AIPNET, the set of input-output relationships is either quite obvious or a simple web search away. The challenge in building the network of input-output linkages across the full universe of HS product codes is not one of information availability, but rather one of efficient information synthesis at scale. This challenge is ideally suited to the latest pre-trained AI language models, whose training over vast corpora of documents, web pages, books, transcripts, and other sources provides sufficient detail on just about every production process imaginable.

To fix ideas, consider some examples: *Full-fat milk and cream* (HS 0401.50) is input to various dairy consumption products, such as *processed cheese* (HS 0406.30) and *Miscellaneous Food Preparations* (HS 2106.90). Figure 2 provides a visualization of this network structure.[20] On the other hand, *Wind generators* (HS 8502.31) require a broad array of capital inputs and intermediary goods, such as *Electrical Control Parts* (HS 8538.90) and *Hot-rolled Stainless Steel* (HS 7219.23). Figure 3 provides a visualization of this network structure. These examples highlight different types of goods: a consumer-facing non-durable good (Full-fat milk and cream) and an essential capital input (Wind generators).

(Figure 2)

For additional examples of solar panels and lithium compounds, see Appendix Figure A.1. For a visualization of the entire AIPNET, see Appendix Figure A.2, where each node represents a Harmonized System (HS) 6-digit product code, and edges depict vertical production relationships.

(Figure 3)

[20]Each figure presents a one-degree network (i.e., direct connections) of a single focal product. Node size is proportional to network centrality of nodes, as detailed in Section 3.5. Nodes are color-coded by their BEC5 end-use classification: pink for final-primary goods, green for final-processed goods, blue for capital goods, yellow for raw materials, and grey for intermediary inputs.

3.2 Using Generative AI to measure Production Linkages

To construct AIPNET, we utilized a structured pipeline that integrates advanced AI techniques with rigorous data processing. This pipeline is visually summarized in Figure 4 and involves several key steps:

Step 1: Large Language Model (LLM) Setup and Prompting We began by adopting GPT-4o, a state-of-the-art foundational AI model. Although GPT-4o's pre-training spans the requisite information to detail input-output structures, we had to further tailor the model to our specific data extraction task through careful selection of parameters and custom prompt-tuning. The model was iteratively queried using each HS code as the focal point, instructing the model to list and describe all products vertically linked to the focal product. This process was repeated multiple times (K draws) to account for the probabilistic nature of the model, capturing a broad range of possible input-output relationships. To manage expectations, we prompted the LLM to list up to 20 goods, prioritizing those with higher importance scores and focusing on internationally traded products.[21]

Step 2: Generating and Parsing Model Output For each focal HS code, the LLM produced free-text descriptions of related products, which were then parsed to extract vertical linkages. These descriptions were aggregated across iterations to form a preliminary network structure, where nodes represent HS product codes and edges denote the input-output relationships inferred by the model. To ensure robustness and consistency across iterations, the output from each iteration was weighted and aggregated, preserving the most consistent connections. This step allowed us to capture a range of possible relationships while ensuring the network's stability against potential LLM-induced variability.

[21]Note, the importance scores do not have any meaning across individual draws but provide a mere rank ordering.

Step 3: Creating Vector Embeddings and Matching HS Codes Given the free-text descriptions generated by the LLM, the next challenge was to accurately map these to the corresponding HS codes. Directly asking the LLM for HS codes posed several risks, including nomenclature inconsistencies and a trade-off between precision and recall.[22] To overcome this, we used text embeddings to match LLM-generated descriptions to official HS codes. Both the LLM-generated descriptions and the official HS code descriptions at the 6-digit level were transformed into vector embeddings. A cosine similarity threshold of 0.75 was applied to balance precision and recall effectively, ensuring that the connections within AIPNET were both accurate and consistent with official HS nomenclature. This approach mitigated the risks associated with direct code retrieval and provided a robust method for extracting and matching the free-text responses to the correct HS codes.[23]

Step 4: Pruning the network After building the initial network, we implemented a pruning stage to improve precision. Additional LLM queries were conducted to assess whether each proposed input-output pair represented a legitimate production linkage. For each pair of upstream and downstream HS code descriptions, the AI was prompted to evaluate if one product could realistically be an input for the other. This step removed incorrect linkages, enhancing the accuracy of the network by ensuring that only valid input-output relationships remained. The pruned edge list was used to construct the final network. This network was bidirectional, representing both upstream and downstream relationships. Further details on how we operationalized the latest AI tools to populate the entire network structure over HS product codes are documented in Ap-

[22]This was a significant issue as the model would often hallucinate numerical codes and incorrectly assign products to codes that did not match. The model had clearly learned the basic idea that codes were numerical and had technical-sounding names, but could not recover specifics.

[23]We experimented with 'Retrieval Augmented Generation' or RAG, which involves producing further context to the LLM. This has shown promise in other applications (examples), but in our context it failed. Our experience was such that the RAG-approach was unable to list multiple codes, perhaps as this technology is geared towards findings a single accurate response rather than searching across a large text-corpus for many valid answers. It is likely that future generations of AI models will overcome this, making RAG a good choice for high-dimensional classification challenges such as this.

pendix B.

(Figure 4)

3.3 Network Descriptive Statistics

Table 2 presents an overview of AIPNET's key structural characteristics at both the HS6 and HS4 levels of product aggregation, such as the number of nodes and edges, average degree, and the size of the largest connected component, across different configurations.[24]

At the HS6 level, the network consists of 5,633 nodes and 980,018 edges, with an average degree of approximately 348. Similarly, at the HS4 level, the network comprises 1,190 nodes and 48,212 edges, with an average degree of approximately 81. Both networks are sparsely connected, with a density of 0.03, yet they exhibit moderate clustering, as indicated by global clustering coefficients of 0.37 and 0.32 at the HS6 and HS4 levels, respectively. This suggests the presence of clusters within the production process. The assortativity is positive at 0.14 for the HS6 network and 0.06 for the HS4 network, indicating a mild tendency for products with similar connectivity to associate with each other, potentially forming specialized production clusters. Both networks are contained within a single connected component, suggestive of cohesive structures.

To gain a better intuitive understanding of the relationship between different nodes and the number of incoming edges, we leveraged the Broad Economic Categories (BEC) classification to explore to what extent our network is predictive of the typical classification of a specific node in terms of input/output labels or its relative position in a supply chain. For each HS6 good, we label this good as per the BEC classification as

[24]Network terms used in this section include: (1) **Network density**—the ratio of actual edges to possible edges, reflecting how interconnected the network is; (2) **Global clustering coefficient**—a measure of the degree to which nodes in a network tend to cluster together; (3) **Assortativity**—the preference for nodes to attach to others that are similar in some way, such as having a similar degree; (4) **Average degree**—the average number of connections per node, indicating typical connectivity; (5) **Connected components**—distinct sub-networks within the overall network. These terms are commonly used in network analysis to describe a network's structural properties.

either an end use being capital, intermediary, or final consumption. This categorisation is mutually exclusive and collectively exhaustive.

Figure 6 characterises each node's classification in terms of the BEC and compares this with a key parameter of interest: the number of output edges or output uses. Not surprisingly, intermediary processed, capital goods, and primary intermediary goods, on average, have many more output uses, suggesting that they are inputs in many other economic processes before reaching end consumption.

(Figure 6)

Nodes that are classified as (end) consumption, on average, and not surprisingly, have few other output uses. This implies that, in empirical exercises, where we are contrasting end consumption goods with other types, intermediary goods are characterised by relatively few input edges and many potential output uses. This highlights that, on average, when we explore variation that arises from the number of input edges, for example, such goods would typically be more likely to be capital goods, while goods with many output edges are unlikely to serve the function of final goods consumption but are much more likely to be considered intermediary goods.

The distinction between different types of goods is particularly relevant, as interventions such as industrial policy or persistent economic shocks are much more likely to result in a growth in trade of capital goods to build local capabilities, while temporary supply disruptions or shocks may more likely yield an increase in imports of intermediary goods as a mechanism of temporary substitution.

3.4 Network Validation

One commonly used source of information on the structure of production networks is the Input-Output Tables (IOTs) and Supply-Use Tables (SUTs) published by many national statistics authorities. The IOTs record transactions across industries, whereas the

SUTs show the products and services being used and produced across each industry.[25] Conceptually, the IOTs/SUTs are based on a domestic economy's transactions covering both goods and services. This is vastly different from the way AIPNET was built, which instead utilizes the vastness of the textual pre-training in recent LLMs to recover relationships between traded physical products. Nonetheless, we would expect an overlap in particular among the goods-producing sectors with the input/output structure that AIPNET implies. We proceed with such an empirical comparison and validation.

For this comparison, we collected data from the IOTs published by the United States' Bureau of Economic Analysis (BEA) and the Mexican national statistics agency, INEGI. The BEA periodically produces industry-level Input-Output tables derived from the economic census and complementary sources at the Department of Commerce.[26] We employed the 2017 version obtained from Bureau of Economic Analysis, U.S. Department of Commerce (2022). Similarly, we used the 2023 release of the Mexican Input-Output tables published by INEGI (INEGI, 2023). These datasets include the Use table and the Direct Requirements table for both countries.

The *Use table* is a matrix of approximately 402 commodity-producing and 402 receiving industries for the U.S. data. The Mexican matrix is a product-by-product matrix with 263 product categories. Thus, a square in the matrix reports the value of commodities produced by industry i consumed by industry j. The US BEA-industry classification closely corresponds to the NAICS, while Mexican 4-digit SCIAN is identical to the 4-digit NAICS.

Requirement tables are derivatives of use tables. The agencies infer the share of inputs from industry i used to produce a dollar output of industry j. The *Direct Requirement table* accounts for immediate inputs. Thus, direct requirement tables are normalized by

[25] This is often referred to by statistical authorities as 'balancing the SUTs'. This involves compiling the unbalanced SUTs from raw data inputs, which span the three different views of total GDP, namely the production, expenditure, and income view. In theory, these three views of GDP should equate, and so the balancing procedure seeks to impose this accounting restriction by apportioning data across different products and industries in the economy.

[26] Refer to Horowitz and Planting (2006) for detailed accounts of the BEA methodology.

the industry's size and report which upstream industries produce inputs, i.e. components, required for a downstream industry's production. It is the empirical measurement closest to AIPNET.

To bring our network to the BEA and INEGI industry levels, we used correspondence tables for 6-digit HS codes and industry codes as provided by Antràs et al. (2012) for the U.S. and by Pierce and Schott (2012) for the Mexican concordance. Our 6-digit code network is generally more granular with more than 5,000 product codes, assigning a binary variable that takes the value 1 for connected dyads. As multiple product dyads often map into a single industry dyad, we averaged the binary connection values in our network. We then standardized both the I-O table values and the network scores. A drawback of aggregating our network using these correspondence tables is the considerable loss of data granularity. We remained with approximately 205 BEA industries for the U.S. and 263 for Mexico. These industries broadly correspond to the goods-producing sectors of the economy.

Our comparison of AIPNET to the US official economic structure is performed using the below regression model:

$$Y_{i,j} = \beta N_{i,j} + U_i + D_j + \epsilon_{j,i} \tag{1}$$

where $Y_{i,j}$ denotes the inputs produced in industry i by industry j according to the BEA Input-Output tables. $N_{i,j}$ is an aggregated industry-level score and indicates whether, according to our network, industry i's output is a required input for industry j's production. U_i and D_j are either sector-specific intercepts for the sectors of industries i and j or industry-specific intercepts for the industries i and j.[27]

The binscatter plots in Figure 5 illustrate the strong positive relationship between the empirical networks derived from the Input-Output tables and our network. As anticipated, the Network score shows the strongest explanatory power in the Requirement

[27]The BEA categorizes industries into 23 sectors.

tables, particularly the Direct Requirements table. The slope coefficients from the regression analysis indicate that a one standard deviation increase in the Network score is associated with an increase of approximately 0.035 standard deviations in the U.S. Use Value, and 0.208 standard deviations in U.S. Direct Requirements. Similarly, the results for Mexico show comparable patterns, with the Network score explaining a significant proportion of the variation in Mexico's Input-Output relationships. This strong correlation supports the credibility of our network model in capturing production relationships that are reflected in empirical inter-industry trade data across different national contexts.

Appendix Table A.3 reports the regression outputs of equation 1. We regress the standardized Input-Output values on the standardized network scores without fixed effects in columns (1), (4), and (7), with upstream and downstream sector-fixed effects in columns (2), (5), and (8), as well as with industry-fixed effects in the remaining columns. The coefficients are highly significant across all models and can be interpreted as the standard deviation increases in the Input-Output value as our Network score increases by one standard deviation. The consistency of these results across both U.S. and Mexico Input-Output tables reinforces the robustness of AIPNET.

(Figure 5)

Validation of AIPNET with Perturbed Networks There may be natural concerns about the extent to which an AI-retrieved production network may be subject to noise due to the probabilistic nature of the inference that is carried out by LLM in the information retrieval step. We evaluate to what extent making the AIPNET stochastically noisier, dampens the correlation between AIPNET and the SUT/IO tables. We do so by introducing perturbations on AIPNET, probabilistically rewiring the edges in the network varying from 1% to 100%. The rewiring was conducted using the Erdos-Renyi random graph model, a method that introduces controlled randomness while preserving the network's structural properties (Erdos, Rényi, et al., 1960). Specifically, edges between nodes were rewired with a probability proportional to the perturbation intensity, re-

sulting in a randomized network that maintains the original number of edges but with varying levels of structure degradation.

We then compared these perturbed networks against the Input-Output (I/O) tables by calculating the difference in R^2 between the original, unperturbed network and each perturbed version. The findings, presented in Appendix Figure B.7, demonstrate a clear trend: as the intensity of perturbation increases, the R^2 difference also increases monotonically, signifying a decrease in the network's predictive accuracy as it deviates from its original structure. This pattern holds consistently across both U.S. and Mexico I/O Tables—Direct and Use—highlighting that AIPNET's structure encapsulates meaningful economic relationships, which are progressively lost as noise is introduced. Throughout the paper we will document that the positive results are robust to working with perturbed networks.

3.5 Integrated Global Product Centrality (IGPC) and AIPNET

AIPNET provides linkages between different production pairs. For global trade analysis it may be desirable to combine binary production links with actual data on the volume of trade. We propose an Integrated Global Product Centrality (IGPC) measure to assess the importance of products in global trade. The IGPC measure is formulated as an extension of the PageRank algorithm, incorporating both the network structure and product- and country-specific trade data. It is measured as:

$$\mathbf{X} = (1 - d)\mathbf{B} + d\mathbf{AWX}$$

where $\mathbf{X}$ is the vector of IGPC scores, d is a damping factor, $\mathbf{B}$ is a base importance vector, $\mathbf{A}$ is the adjacency matrix derived from AIPNET, and $\mathbf{W}$ is a diagonal matrix of weight adjustments. This formulation allows the importance of a product to be influenced by both its intrinsic trade characteristics and its position in the global production network.

The base importance $\mathbf{B}$ and weight matrix $\mathbf{W}$ incorporate two key trade metrics:

Global Trade Share (GTS) and Trade Concentration (TC). For a product i, these are combined as:

$$B_i = TC_i^{\alpha} \times GTS_i^{\beta}$$

$$W_{ii} = GTS_i^{\gamma}$$

The parameters α, β, and γ allow flexible weighting of trade concentration, global trade share, and their interaction in the network structure.

The AIPNET structure, encoded in $\mathbf{A}$, determines how importance propagates through the production network. A product used as input in many other products will have more non-zero entries in its corresponding row of $\mathbf{A}$, allowing it to accumulate importance from a wider range of downstream products. This captures that products central to many production processes are inherently more important to global trade.[28]

The iterative nature of the IGPC calculation allows importance to flow through multiple levels of the production chain. This feature captures the significance of products that might not have high direct trade volumes but are crucial inputs in complex production processes. For instance, a specialized component used in the production of advanced electronics might have a higher IGPC score than its direct trade volume would suggest due to its position in the production network.

By integrating the structural information from AIPNET with empirical trade data, IGPC provides a nuanced measure of product importance. It balances the intrinsic trade characteristics of products and their roles in global supply chains, offering insights that go beyond simple trade volume statistics. This makes IGPC a valuable tool for analyzing global trade patterns, identifying key products in supply chains, and informing trade policies in an increasingly interconnected global economy.

We next turn to studying these phenomena using global trade data.

[28] We provide more formal details on the IGPC construction in Appendix E.

4 Global Trade Patterns

Next, let us showcase the value of using a product-level production network by leveraging AIPNET to document patterns in global trade data. Specifically, we merge AIPNET with international trade data to measure the Integrated Global Product Centrality (IGPC) of goods. This allows us to shed light on global trends in trade, focusing on the criticality and centrality of goods in the global production network.

4.1 Global trends in goods trade centrality

IGPC provides a comprehensive measure of product centrality that reflects both the complex structure of global production networks and the empirical realities of international trade regarding trade volume and concentration (see section 3.5). We begin by constructing import-volume indices at constant prices weighted by a time-invariant IGPC measure of product centrality.

Aggregated data Figure 7 Panel A compares countries' import centrality in global production networks. China's index has risen sharply while the US's has declined, with the EU remaining stable until a marked shift in 2022. These trends suggest two interpretations: China's increasing sophistication may indicate growing domestic production capabilities concentrated around importing critical goods, while declining US import centrality could reflect reduced global dependence on domestic supply of critical inputs[29].

Panel B compares trends between a range of products classified by their baseline IGPC index in terms of their importance across quartiles. On average, we observe that goods with a higher IGPC importance score have trended upward, suggesting that global trade of these critical products increased during this period.

(Figure 7)

[29] A notable driver of the reduction in the centrality of US import bundle over this period is the reduced reliance on imported hydrocarbon productsdue to the shale booms (Arezki, Fetzer, & Pisch, 2017).

Figure 7 Panel C and D residualises the raw data using country-by-year and time-invariant country-by-product fixed effects allowing us to observe the compositional shifts underlying the raw trends in the aforementioned panels. We find similar patterns that highlight the divergence in both the composition of imports by production network centrality across China/US (panel C), as well as the shift in global trade away from lower-importance goods towards higher-importance goods (panel D). These patterns suggest a structural shift in the composition of imports after 2016, when trade barriers and other geopolitical events became more salient.

4.2 Time-varying Product Importance

Alternatively, we can study the evolution of specific products' importance over time, leveraging the time-varying aspect of IGPC. Figure 8 illustrates shifts in the relative importance of different products within global supply chains between 2010 and 2022, as measured by the Input Global Product Centrality (IGPC) index. Panel A highlights trends in the importance of traditional energy products in the production network, showing a decline in the trade network centrality of crude oils and nuclear energy. It also highlights the significant spike of electrical and natural gas energy in 2022, likely driven by the dislocations of global energy trade following the Russian invasion of Ukraine. Panel B shows a selection of products whose importance has decreased notably. This includes products like CDs/DVDs, photographic film, and typewriter parts. Panel C looks at a selection of emerging products whose importance has risen dramatically (note the log scale in this exhibit). These emerging technologies include digital integrated circuits (e.g. chips) and lithium compounds (e.g. used in battery technology).

(Figure 8)

These structural changes are further reflected in the broader economic categories (Panel D of Figure 8), where capital goods and industrial supplies have shown sustained increases in production network importance, rising by approximately 30-50% relative to

their 2010 levels. In contrast, fuel and lubricants have experienced a notable decline, particularly after 2014, aligning with the reduced importance of traditional energy products shown in Panel A. This transformation suggests a fundamental shift in global supply chains toward high-technology products and away from conventional industrial and energy inputs, potentially reflecting technological change and evolving environmental priorities.

4.3 Trends for Climate Action and Defense Products

We next present some summary evidence focused on trade data along value chains for two classes of goods that are particularly salient in light of geopolitical developments: goods used for climate action compared to goods for primarily military use.[30]

(Figure 9)

In Figure 9 we show a simple index of import volume at constant global average prices, aggregating global imports of both the focal set of products (Panel A) as well as the set of directly related inputs used to produce these goods according to AIPNET (Panel B). Panel A shows results for the importation of the focal set of products. We see a sharp increase in imports of climate action-related goods from 2020 onward.Panel B turns to products directly upstream of arms and green transition products. For green transition products, we see a more gradual rise in imports of inputs. We also see a steady rise in global trade for products used to produce defense-related products. These patterns may suggest increasing localization or 'onshoring' of the production of the end products.

[30]We focus on Climate Action Goods, 841861/841581 - Heat Pumps, 850231 - Wind Turbines, 854140 - Solar Panels, 840110 - Nuclear Reactors, 841011 - Hydroelectric Turbines, 870380 - Electric Vehicles, 850760 - Electric Storage; for the Military goods we consider 930190 - Military weapons, 930200 - Revolvers and pistols, 1930320 - Shotguns, 930120 - Artillery weapons 930621/930630 - Cartridges, 930690 - Munitions, 871000 - Armored vehicles, 930700 - Swords and bayonets, 880220/880230 - Helicopters, 880240 - Unmanned aircraft, 880250 - Spacecraft 880220/880230/880212 - Military drones, 880240 - Military aircraft, 880330 - Aircraft parts.

4.4 Sanctioned goods

Another use case of the study of import patterns in focal goods and their inputs are sanctions established against Russia in response to the invasion of Ukraine. These are presented in Panel C and D of figure 9 respectively. In both figures, we consider sanctions on 'critical', 'military', and other 'sanctioned' products.[31]

Panel C shows the import index (by value) for goods identified as 'critical' and other 'sanctioned' dual-use products.[32] Despite the war and sanctions, import values of these goods have not declined significantly and display fluctuating but resilient import values, suggesting widespread evasion of sanctions possibly via third countries. We see a permanent drop in military imports, as measured by official publicly available trade data[33]

Panel D presents the import index for inputs used to produce sanctioned goods. Remarkably, the import of inputs, especially those related to military goods, has experienced the strongest contraction since the onset of the conflict. This steeper decline indicates a tighter restriction on the components that are essential for manufacturing military-related products, likely reflecting more vigorous enforcement or greater disruption in supply chains for these input goods. This is perhaps surprising, given the incentive to evade sanctions through localization of production. The sharp decline is, however, followed by a rebound and perhaps foreshadows a structural transformation in years to come. This could also reflect well-designed sanctions that foresaw these potential localization efforts.

[31] The list of sanctioned products were taken from EU lists, respectively at https://finance.ec.europa.eu/publications/list-economically-critical-goods_en ('critical' goods) and https://finance.ec.europa.eu/publications/list-common-high-priority-items_en (other 'sanctioned' products).

[32] An immediate corollary implication of AIPNET is to leverage it to identify or map out the capability space of countries.

[33] The data underlying these indices comes from monthly trade values reported by partner countries trading with Russia, as available from the UN Comtrade database. Due to limited quantity data, the analysis relies on trade value trends, which may even understate the actual decline in volume, as wartime inflationary pressures likely elevate prices. Thus, the observed drop in trade values likely corresponds to an even greater reduction in the physical volume of imports. This may well be masked. (Egorov, Korovkin, Makarin, & Nigmatulina, 2024) present even richer patterns using more reliable trade data from Russian customs authorities and study sanctions and their effectiveness with high precision.

The capacity to study rich patterns of substitution across the production network using AIPNET opens up many avenues of further analysis of sanction. Beyond the study of outright evasion of sanctions, the network can be used to study the role of dual-use products, which embed a range of productive capabilities and facilitate smoother structural transformation of the domestic economy.

4.5 Trends in Supply Shocks along Value Chains

We next present evidence of persistent supply shocks, as measured by persistent increases in unit prices. We measure these shocks using reduced-form analysis of time-series patterns, identifying structural breaks in the log-unit prices in a parsimonious manner, discussed in Appendix F. In this subsection, we document patterns of these supply shocks across space, time and along value chains. We will use these price shocks in our analysis of spill-overs in Section 5, where we consider how such shocks drive import substitution across the supply chain.

In figure 1, panel D, we documented the large increase in supply shocks since 2016. In Figure 10, we further break this down to characterize these shocks along input/output linkages. Specifically, we characterize both the spatial distribution of supply shocks post-2016 (Panel A), the product composition of changes to supply shocks pre- and post-2016 (Panel B), and the evolution of supply shocks across the production network centrality distribution (Panel C).

In Panel A, we see that Canada, China, the US, Russia, Saudi Arabia, Argentina and Switzerland have been most exposed to said shocks. These are broadly aligned with priors on a host of trade frictions and other global events, including the escalation of trade frictions, and other relevant events, some of which we alluded to in Figure 1.

Panel B characterizes the distribution of the *change* in the count of these supply shock events during the pre- vs post-2016 periods. We see that the level change in the count of supply shocks is concentrated among intermediary processed goods and (final) consumption goods. This highlights that typically large persistent shifts in the unit price

of traded goods since 2016 have skewed towards products closest to the bottom of the production network, i.e. closer to final consumption.

Lastly, figure 1 Panel C documents the level change in supply shocks according to quartiles of production network centrality, again using our IGPC measure. We observe that the increase in shocks is most pronounced for goods that have relatively low criticality or importance. This may square well with a narrative of localization suggested in Panel B and D in Figure 7, where we noted that imports as of goods with relatively low IGPC importance, in relative terms, saw a decline in imports – possibly owing to the consequence of these goods experiencing many more price shocks in the post-2016 period relative to those with higher network centrality.

Overall, the analysis of price shocks (as measured by structural breaks in unit prices) reveals that such shocks are unevenly spread across geography, increased since 2016, and that this post-2016 increase skews towards more consumer-facing goods with lower criticality in production networks that may be easier to onshore.

(Figure 10)

The rich connections between granular products and our production network facilitate many new ways to look at global trade patterns. We demonstrate this with a handful of key results but stress that the potential for more focused analysis using this product extends well beyond these findings. In the next section, we demonstrate a more direct way to exploit the network representation, wherein we consider how import demand spills over (i.e., substitutes) along the production network.

5 Onshoring as Evidence of Production Network Spillovers

The previous section highlighted a broad set of use cases and documented some patterns in trade data that are informed by AIPNET. Some of these patterns may be consistent with import demand substitution. When demand moved from more downstream to

more upstream goods, this is evidence of 'onshoring' i.e. the localization of production processes.

To study this kind of onshoring, we return to global trade data and utilise the production network linkages to construct product-pairs or dyads. We carry out two sets of exercises using this structure: First, a cross-country exercise exploiting variation of a range of 'supply shocks' (i.e. structural breaks in the the unit-price of traded goods). Second, we leverage a natural experiment which imposed a sizable supply shock to specific products following an economic blockade. This latter exercise allows for a cleaner causal analysis, whereas the first exercise allows us to compare countries according to their propensity to substitute across the production network in the face of price shocks. In both cases we leverage dyadic regression designs.

We first layout a general overview of the structure of the data for the analysis, layout our production network dyadic empirical design, and lastly present the results of our analysis.

5.1 Data Structure and Product-Dyads

To study import demand substitution along the production network, one would ideally have knowledge of the intended use of a given import product. That is, ideally, we would observe trade data that was *use*-specific.

Such data is hard to come by for at least three reasons: First, it requires highly granular information on importer-firm production processes.[34] Second, the ex ante stated uses often belies the ex-post use, especially when goods are multi-purpose and can be

[34]Data that may capture the intended end-use of an upstream good is available with tax and custom authorities, in particular considering capital goods or goods that could have dual uses in production. As such, these goods would be subject to specific forms of export handling protocols that would produce such a data trace: export control licenses; end-user certificates; trade finance data; international commercial terms or contracts. Each of these would typically state the intended end use. With increasing digital payments, or through the use of DLT technology, solutions for supply chain transparency could arise producing anonymized data with similar effect; alternatively, as countries tax systems are better digitally integrated with customs declaration and invoice-level VAT, the tax system could support a low cost solution. This is particularly salient when it comes to ESG goals, but may reduce the economic role of a range of – typically licensed – traditional service sector occupations.

deployed, resold, or repurposed (a feature especially of capital equipment). Third, there are often a host of 'middle men' who facilitate trade. Such entities may identify in a certain industry, but very likely resell imports without transformation - making the industry code of imports a poor proxy for the end use of capital and intermediary inputs.[35] These challenges not withstanding, we recognize that for many goods our network identifies ex ante *possible usages*, not actual ex post production.

Being limited to our fixed production network and global trade data at the product, but not the input/output dyad level poses some challenges. Namely, we have to reconcile that the driver of variation in the observable import demand for a specific good, $\text{Import}_{u,c,t}$, is influenced by a host of factors. This includes the direct usage of the good for final consumption, or the usage of the good for a potentially large set of production applications. If we make a strong assumption of independence, we can express the variance of the import demand time-series of one good u in country c during time period t as:[36]

$$\text{Var}(\text{Import}_{u,c,t}) = \overbrace{\text{Var}(\text{Import}_{u,f,c,t})}^{\text{Final consumption}} + \underbrace{\overbrace{\text{Var}(\text{Import}_{u,k,c,t})}^{\text{Capital use}} + \overbrace{\sum_{j=1}^{\kappa} \text{Var}(\text{Import}_{u,d_j,c,t})}^{\text{Intermediary}}}_{\text{Production Network Linkages}} \quad (2)$$

where $\text{Import}_{u,c,t}$ is the *observed* import demand, $\text{Import}_{u,f,c,t}$ is the latent unobserved demand for this product for final consumption, and where $\text{Import}_{u,f,c,t}$ and $\text{Import}_{u,f,c,t}$ represent the latent demand for product u in country c during period t as capital or various intermediate uses.[37]

[35]The case of de-minims trade and direct sale to consumers reduces the middle men involvement, but threatens to hollow out traditional service sector business models designed around bulk purchase, storage, distribution and after sale services. The environmental costs associated with direct shipping, often involving airfreight needs to be considered in cost-benefit analysis (P. Fajgelbaum & Khandelwal, 2024).

[36]In the econometric specification, we will account for country c by HS2 digit level u' time fixed effects, which will absorb variation in demand shifts towards close substitutes that would be typically found within the same harmonised system nomenclature branch.

[37]This decomposition has be proposed by (Bems, Johnson, & Yi, 2011) and (di Giovanni, Levchenko, & Méjean, 2014), who highlighting the amplification effect of intermediate goods trade. This and other empirical work finds that, in particular, the variation in capital goods trade tends to be sizable, allowing

It is clear from the above that final consumption may be a confounder, which can vary due to local preferences, macroeconomic shocks, and the like.[38] To address this, we utilise a highly saturated set of controls in our product-dyad regression design, which– along with evidence that final demand is far less volatile that capital and intermediary demand,[39] we proceed with some measure of comfort that our analysis is informed by the relevant variation in product-country-time import demand.

5.2 Cross country evidence

We begin by document trends in global trade within this fully dyadic empirical framework that exploits variation in the topology of the production network graph. That is, we explore to what extent upstream goods that have many downstream uses – which, we can characterize as intermediate or capital goods – see, on average, a differential evolution of imports over time.

We estimate on a fully dyadic panel dataset the following specification:

$$log(\text{Import}_{u,c,t}) = \sum_{\tau=2012}^{2022} \beta_\tau \times \mathbb{1}(t = \tau) \times \omega_{u,d} + \alpha_{u,d,c} + \eta_{u_{hs2},c,t} + \epsilon_{u',d'} \tag{3}$$

Here, the dependent variable is the import of a 4-digit upstream goods u, into country c in year t. The $\omega_{u,d}$ is our binary linkage indicator that captures whether a specific upstream good u serves as an input to downstream good d:

$$\omega_{u,d} = \begin{cases} 1 & \text{if input } u \text{ is a an AIPNET input used to produce } d \\ 0 & \text{otherwise} \end{cases}$$

for reasonable assumptions about this latent demand component being a useful source of variation, as our application will require.

[38] While final demand is a confounder, there is evidence that consumer tastes and preferences evolve quite slowly, limiting the variation attributable to this latent component (see e.g. (Bronnenberg, Dube, & Gentzkow, 2012) ,(Handbury & Weinstein, 2014),(Bar-Gill & Fershtman, 2021),(Griffith, O'Connell, & Smith, 2019))

[39] See (Bronnenberg et al., 2012) ,(Handbury & Weinstein, 2014),(Bar-Gill & Fershtman, 2021),(Griffith et al., 2019)

That is, with this data structure and empirical framework, we allow each good, potentially, to serve as an input in all other HS4 goods.

The baseline specification adjusts for dyad by country specific fixed effects, $\alpha_{u,f,c}$, which accounts for idiosyncratically higher levels of imports for each potential downstream d use of an upstream good u. For example, a country with a relatively low level of economic development may exhibit a notably higher level of imports of certain upstream goods u that are not feeding as an input into the production of marketable goods, but rather, enter directly final goods consumption that may involve a process of home production. We also account for country-specific HS2-digit level time fixed effects for each upstream good. This removes country-specific unobservable idiosyncratic shocks to import demand within each HS2 good. This accounts for demand changes potentially induced by price shocks that may trigger substitution to close substitutes, which, typically, would be classified within the same two digit class of goods. Standard errors in this specification are blocked by upstream HS2 x downstream HS2 two digit pairs, allowing for arbitrary correlation between countries and within countries error terms at a rather coarse two digit by two digit block structure.

Note, that this specification is broadly equivalent to estimating

$$log(\text{Import}_{u,c,t}) = \sum_{\tau=2012}^{2022} \beta_\tau \times \mathbb{1}(t=\tau) \times C_{k_u} + \alpha_{u,c} + \eta_{u_{hs2},c,t} + \epsilon_{(i,t)}$$

where

$$C_{\kappa_u} = \frac{1}{J} \sum_{d \in \kappa, d \neq u} \omega_{u,d}$$

measures the number of downstream uses, or outdegree links of a good u.

That is, the empirical interpretation of patterns in specification 3 would indicate that the β_τ capture the differential evolution of imports of upstream goods that have relatively more downstream uses in production processes, which is generally the case for intermediary goods and capital goods (see Figure 6).

Results The results from this analysis are presented in Figure 11. The plot visually presents the estimated coefficients β_τ with 2016 being the omitted year. We note that, over time, and relative to 2016, there has been a notable increase in imports of upstream goods with, on average, more downstream uses as per AIPNET, compared to goods that have fewer downstream uses.[40] Given the characterisation in Figure 6, this suggests that, globally, there has been a trend towards importing capital and intermediary goods – that have many downstream uses – and less so, importing goods that are primarily serving final consumption demand.

(Figure 11)

If anything, the dynamic prior to 2016 suggested a gradual decline, while from 2016 onwards, imports of goods with many downstream uses have increased with 2018 seeing a sharp increase, potentially related to the starting escalation of trade disputes during President Trump's presidency, followed by yet another sharp increase in 2021, and then, again a further increase in 2022, that marked the start of the war in Ukraine. This may map into the characterisation of the supply shocks that we detected in Section 4.5, which suggested an increasing number of supply shocks for finished consumption goods and processed intermediary goods. The above analysis suggests that countries, on average, increased imports of intermediary and capital goods, which, on average, have more outdegree links.

We next explore this pattern country-by-country and study the onshoring hypothesis following price shocks more explicitly.

[40] The dependent variable here is the log of imports, capturing the intensive margin. Appendix Figure A.3 presents the results where the dependent variable is a binary indicator indicating any imports.

5.3 Cross country evidence on supply shocks and onshoring

We next estimate the above specification split-sample country-by-country, replacing the flexible time-dummies with a simple before- and after 2016 dummy. That is, we estimate:

$$log(\text{Import}_{u,c,t}) = \beta_c \times \mathbb{1}(t > 2016) \times \omega_{u,d} + \alpha_{u,d,c} + \eta_{u_{hs2},c,t} + \epsilon_{u',d'} \tag{4}$$

Note that here, β_c captures a separate coefficient, one for each country c. This allows us to explore variation in the extent to which the trend that was detected globally varies between countries. This reduced form measure, β_c, itself has no substantive economic meaning apart from indicating that countries that exhibit a positive coefficient would have seen an increase in imports of upstream goods with many downstream uses, which, we characterized as being, more likely to be considered intermediary or capital goods based on their BEC classification.

Given the observation in Panel D of Figure 1 and in Figure 10, we can think of the post 2016 dummy as being a crude proxy for the sharp increase in supply shocks that started hitting intermediary and consumption goods from 2016 onwards. We also estimate a parametric specification where we replace post 2016 dummy variable with the filtered price shocks. That is, we estimate:

$$\begin{aligned}\log(\text{Import}_{u,c,t}) = [\beta_{d,c} \times \text{Shock}_{d,c,t} + \beta_{u,c} \times \text{Shock}_{u,c,t}] \times \omega_{u,d} + \xi \times \text{Shock}_{d,c,t} \\ + \eta \times \text{Shock}_{d,c,t} + \alpha_{u,d,c} + \eta_{u_{hs2},c,t} + \epsilon_{u',d'}\end{aligned} \tag{5}$$

Here, the specification and the parameters of focus are $\beta_{d,c}$ and $\beta_{u,c}$. Specifically, a positive sign, $\beta_{d,c} > 0$ would be indicative of onshoring. That is, imports of upstream unfinished goods used in the production to make a finished downstream good d, are increasing following a supply shock that leaves the downstream finished goods persistently more expensive. Naturally, we would only expect such a dynamic to play out between price shocks hitting downstream goods d, if said upstream good u is an AIP-NET input, $\omega_{u,d} = 1$, in the production of downstream good d.

Further, we would, on average, expect that the net direct effect of price shocks to upstream goods to be negative, $\beta_{u,c} < 0$. That is: imports of unfinished upstream goods contract, if the upstream goods become more expensive, on average.

To understand the relationship between the two specifications, consider that the first model uses a post-2016 indicator interacted with the upstream-downstream linkage to capture changes in import patterns after 2016. In this specification, the coefficient β_c measures the average change in imports of upstream good u into country c after 2016, specifically for goods linked via $\omega_{u,d}$ to downstream uses d. Essentially, β_c captures the combined effect of all factors that changed after 2016 and affected imports through these linkages. We would expect that the emergence of industrial policy and explicit localisation policies that Panel C of Figure 1 suggested to leave marked direct effect, on average, irrespective of the role of market signals or supply shocks could play, which would be picked up by the $\beta_{d,c}$ in specification 5. That is, we would expect the coefficients β_c to be a catch all term that may pick up industrial policy induced localisation, as well as price-shock induced localisation. This should be particularly the case for countries that have been subject to a broad range of supply shocks, such as China, the US, the Middle Eastern countries (Saudi Arabia, Qatar and the United Arab Emirates) and Canada.

We present the results from estimating specification 4 and 5 visually in Figure 13. Each dot that is plotted on this canvas represents a pair of point estimates, $(\widehat{\beta_c}, \widehat{\beta_{c,d}})$. The x-axis plots the non-parametric post 2016 coefficient, which captures, to what extent a country has seen a differential increase in its imports of upstream goods that have many AIPNET input-output linkages – typically capital or intermediary goods. The vertical axis captures the extent to which imports of upstream goods are increasing following supply shocks to downstream goods.

(Figure 13)

The figure suggests that the two phenomena that we are measuring here are not statistically independent. Along the y-axis providing the estimate of the price-shock

induced onshoring coefficient, for 75 out of the 120 countries, we observe a positive coefficient $\widehat{\beta_{c,d}}$ indicating an onshoring dynamic following a price shock. Considering the x-axis, for 77 countries out of 120, we notice a marked post 2016 increase of imports of upstream goods with more AIPNET outdegree links – typically capital and intermediary goods. The top right quadrant captures the overlap of the two estimated relationships. That is, a set of countries where we noticed a marked increase in imports after 2016 of mostly upstream capital or intermediary goods, coincides with the set of countries that have experienced notable increases in upstream imports following downstream supply shocks that mostly persistently increased the cost of finished consumption goods and processed intermediate goods. We can interpret this as capturing the extent to which economies have sufficiently well developed market-based institutions.

Among the set of countries that appears in the top right, we explicitly want to highlight Qatar. We will present a natural experiment pertaining to Qatar that highlights the onshoring effect following a large trade blockade by its neighbouring countries.

Robustness Table 3 presents pooled regression results from the global exercise on the fully dyadic dataset. Column (1) presents the average post 2016 increase in imports of goods with, on average, more AIPNET outdegree links that is visually presented in Figure 11. Column (2) focuses on the price-shock regresson equivalent. The sample here is slightly smaller owing to the fact that we do not have price measures for all goods. Column (3) and (4) perform the equivalent exercises just focusing on the extensive margin of imports. Note that in column (4) the direct own effect is removed as this variable is perfectly collinear with observing imports as prices are derived from the import data. The pooled regression highlights that globally, an onshoring dynamic seems to be in place. Our subsequent exercises break this dynamic down for each country in turn.

Appendix Figure A.4 presents the results for the perturbation exercise whereby we successively make the network more noisy through rewiring edges (see section B.6). The top panel of the figure presents the distribution of the country-level coefficients for

two values of the perturbation parameter: $p = 0$ and $p = 50\%$. We note that as the production network becomes more noisy, the country-level estimated coefficients shrink towards zero along both axis. The bottom panel presents these observations plotting the average distance of each estimand from the respective origin across different perturbed AIPNET production networks. We observe that, as the perturbation increases to 100%, i.e. when the production network becomes full noise, that the length of the average squared estimated country-level coefficients converge to zero. This highlights that we detect systematic variation that is bespoke to the production network we are leveraging.

An alternative way to estimate specification 5 would be to focus only on edges that are part of the prodution network, i.e. we focus the estimating sample on edges where $\omega_{u,d} = 1$. The estimating equation then becomes

$$\begin{aligned} \log(\text{Import}_{u,c,t}) =& \xi_{c,d} \times \text{Shock}_{d,c,t} \\ &+ \eta_{c,u} \times \text{Shock}_{d,c,t} + \alpha_{u,d,c} + \eta_{u_{hs2},c,t} + \epsilon_{u',d'} \end{aligned} \tag{6}$$

We present the results from this specification in Appendix Figure A.5. We contrast the regression coefficients from the fully dyadic specification with the subsample where $\omega_{u,d} = 1$. We find qualitatively very similar results. Throughout, it is worth to take specific note of the case of Qatar which, as becomes evident from the analysis, exhibits a notable onshoring response in the cross country exercises. We next expand on the Qatar natural experiment that relaxes a range of identification assumptions or concerns that one may have with the cross-country exercise presented here.

5.4 Causal Evidence of Onshoring from a Natural Experiment

A natural experiment allows us to employ AIPNET to causally estimate how a surge in trade barriers prompts an onshoring response. Between 2017 and 2021, Saudi Arabia and its allies ceased all trade with Qatar.[41] In response to the trade shock, Qatar increased

[41] The countries that participated in the blockade were the immediate neighbours Saudi Arabia and UAE, Bahrain along with Egypt and some smaller countries such as Mauritania, Djibouti, and the Maldives.

its previously low industrial and agricultural production capacity by shifting imports towards capital and intermediary inputs of goods affected most by the blockade.

Figure 14 shows anecdotal evidence of the response in Qatar's import patterns in products primarily sourced through Saudi Arabia before the blockade. Panel (a) plots the import value index of dairy products and the import value index of goods upstream of these dairy products according to AIPNET. Panel (b) is the corresponding plot for poultry. With the onset of the blockade, imports of dairy and poultry slumped. Meanwhile, total import values of their upstream goods surged.

In the following, we provide causal evidence of this onshoring response. In most settings, it is impossible to causally study an onshoring response due to confounding factors muting an onshoring response. For example, we may be concerned about exchange rate movements, general equilibrium effects and complex interactions with existing economic capabilities. The case of the Qatar blockade relaxes many of these concerns. Qatar's GDP is less than 1% of the global GDP, making it a relatively small but open economy. The Qatari Riyal is also pegged to the US dollar, eliminating the effects of the nominal exchange rate.

Blockade Exposure and Data Monthly import data at the HS product level for 2012 until 2023, published by Qatar's National Planning Council, allows clear identification of the onset of the blockade and a granular measurement of blockade exposure. We measure blockade exposure for the 4-digit and 6-digit HS product category j as the share of imports from blockading countries B before the blockade.

$$BlockadeExposure_j = \frac{\sum_{s=2013}^{2016} \sum_{c \in B} Imports_{j,c,t}}{\sum_{t=Jan2013}^{June2016} \sum_{c \in C} Imports_{j,c,t}} \tag{7}$$

where B is the set of the countries participating in the blockade, and C denotes the universe of trading partners. $Imports_{j,c,t}$ is the value of imports of product category j from country c in year t. The measurement is bounded by 0 and 1 and varies substantially across product categories as shown in Appendix Table A.2. Figure 15 shows a

stark association between the blockade exposure of goods and the increase in import price increase and volume decrease after the blockade's onset.

Dyadic regression evidence of Onshoring We next turn to studying onshoring in a dyadic regression setting similar to the cross-country framework. Each observation is a pair of a single upstream good u and a single downstream good d. Each product is related in our supply-chain network representation.[42] Rather than using a measure of the price-shock, we use a goods blockade exposure, to proxy for the supply disruptions, to then document the impact that downstream supply disruptions have on the importation of upstream goods. We estimate the following specification:

$$\begin{aligned}\log(\text{import}_{u,d,t}) = \beta \times \text{Post}_t \times \text{Blockade Exposure}_u \\ + \gamma \times \text{Post}_t \times \text{Downstream Blockade Exposure}_d \\ + \alpha_{u,d} + \theta_t + \epsilon_{u',d'}\end{aligned} \tag{8}$$

where u is an upstream product defined at the 4-digit or 6-digit product level, d is an downstream product defined at the 4-digit or 6-digit product level, and t represents each year-quarter. Fixed effects $\alpha_{u,d}$ and θ_t control for dyad-specific and year-quarter unobservables. As vertically related goods might be correlated in their blockade exposure, we control for the direct effects of exposure for upstream good d by including $\text{BlockadeExposure}_d$. Subsequently, γ recovers the causal onshoring response. A positive γ means that an upstream good's imports increase when the downstream good is strongly affected by the blockade – suggesting onshoring. The specification mirrors the cross-country specification 6, except that here we replace the price shock measure with our reduce form blockade exposure measures for both the upstream and the downstream good. Standard errors are clustered at the 2- by 2- digit HS upstream and downstream code pairs, $\epsilon_{u',d'}$. An alternative is to carry out two-way clustering at the upstream and

[42] See Section 3 for an overview of how we construct this network using our Large Language Model approach.

downstream level. The chosen structure allows for arbitrary cross correlation within blocks of related goods. In addition to this form of clustering, we also carry out the perturbation exercise which induces noise in the dyadic relationships, which can be considered as a type of randomisation inference.

Results We present the results from the onshoring regression in the Qatar case study in table 4. We present the analysis at the quarterly temporal resolution and at the HS4 and HS6 digit granularity. We use blockade exposure as our primary causal shock variable but do not differentiate between the blockade exposure of the upstream good (direct exposure) relative to the blockade exposure of the downstream good in each dyad.

Panel A uses quarterly data at the 4-digit HS product level. Columns (1) through (3) estimate the extensive margin effect of the blockade on the imports of upstream goods. The direct impact is negative. In other words, if the upstream good's import supply chain becomes disrupted by the blockade, the likelihood of this good being imported to Qatar will fall. The indirect effect of downstream goods exposure to the blockade shock is positive. Thus, if the supply of downstream goods is disrupted, the importation of upstream goods is more likely.

The extensive margin effects are presented in columns (4) to (6) using the log value of upstream imports as the dependent variable. We note a large and negative direct effect. The coefficient of 0.56 log points suggests that for the product with the median exposure of 15% and positive imports after the blockade, imports fell by approximately 8.4 percent. Yet, this effect was notably muted if these goods have many downstream uses and the downstream goods are also subject to large supply shocks. The indirect, downstream exposure effect is positive and significant.

Taken together, these provide significant evidence of onshoring, namely that the downstream exposure predicts an increase in imports of upstream goods. While the direct effect dominates the average treatment effect of downstream exposure, one additional percentage of downstream exposure increases imports by approximately 16 per-

cent. That this is significant and positive is strong evidence of onshoring, especially given the onerous controls and likely attrition bias present due to sample pooling.

(Table 4)

Panel B carries out this analysis at the HS6 digit level. We find qualitatively very similar results. However, the high volatility in trade data and the noisiness of measurement cause the treatment effects to be subject to be measured less precisely.

Robustness In Figure 16 we present evidence on the timing of the effects. Effectively, we simply estimate a time-varying coefficient capturing the blockade-exposures time-varying impact on the imports of upstream goods. We note throughout, that the onset of the blockade in the second quarter of 2017 is associated with a sharp decline of upstream imports for which the blockading countries were the dominant suppliers. Conversely, though, we see a notable increase of the import of upstream goods that are AIPNET linked as inputs in the production of output downstream goods that is more pronounced if said downstream goods also exhibited a large exposure to the blockade. This can be considered as evidence in support of the underlying common trends assumption that is implicit to this dosage difference-in-differences estimation.

In Appendix Figure A.6 we present the results from the network perturbation exercise whereby we perturb the AIPNET supply chain linkages. The figure plots out the evolution of the pooled estimated coefficients for both the direct- and the onshoring effect. We would expect that, as the AIPNET network linkages become more and more noise infused as the perturbation increases, that the onshoring indirect effect shrinks to zero. On the other hand, we would expect that the direct effect stays more static. This is exactly the pattern that is visible. As the AIPNET network becomes more noisy, the estimated coefficient on the onshoring exercise gradually shrinks towards zero, while the direct effect remains intact.

6 Conclusion

The resurgence of industrial policy, trade restrictions, and onshoring is evident worldwide, driven by recent supply shocks, including the global pandemic and heightened geopolitical tensions. This paper introduces the AI-generated Production Network (AIPNET), mapping input-output relationships across over 5,000 Harmonized System (HS) product groups using large language models (LLMs). Our analysis with AIPNET reveals a decrease in global imports of downstream goods relative to upstream inputs, signaling a shift toward localized production. Persistent supply shocks post-2016 have driven this trend, with many countries exhibiting an onshoring response to these disruptions. The 2017 blockade of Qatar provides causal evidence, demonstrating a shift toward domestic sourcing for critical inputs.

AIPNET serves as a valuable tool for analyzing global trade dynamics, aiding efforts in supply chain transparency, carbon emissions tracking, sanction efficacy, and price shock propagation. Additionally, a promising direction for future development lies in the creation of local AIPNET products—country or region-specific production networks that reflect unique economic structures, resource endowments, and local industrial policies. By tailoring AIPNET to individual countries or regions, we can capture more precise, context-specific production dependencies and trade flows. Such localized versions would be especially beneficial in identifying region-specific vulnerabilities to supply chain disruptions, the potential for targeted onshoring, and the effectiveness of local industrial strategies. Local AIPNETs could also reveal distinctive production network characteristics shaped by national endowments, regulatory environments, or regional trade agreements, offering tailored insights for policymakers working to strengthen economic resilience at local levels.

Future research could also extend AIPNET into other economic domains. Expanding to occupations could reveal skill dependencies and labor shifts, while mapping scientific discovery networks using patent data would enhance our understanding of knowledge

diffusion and innovation. Developing networks of capital flows or investment linkages could shed light on financial interdependencies and resilience to economic shocks, and creating consumer demand networks could highlight patterns in consumption, substitution, and price transmission. These extensions underscore the potential of AI-driven network datasets to advance economic insights and foster interdisciplinary collaboration in policy studies.

While AIPNET currently focuses on trade in physical goods, future iterations could also incorporate the service sector, especially given the role of AI and capital goods in economic development. Expanding AIPNET's scope would provide a more comprehensive framework for analyzing the complexities of modern economies. In an increasingly interconnected yet fragmented world, AIPNET and its many potential adaptations stand as foundational assets for academia, industry, and policymakers, offering a robust and adaptable toolkit for understanding and shaping the future of trade and production.

References

Acemoglu, D., Akcigit, U., & Kerr, W. (2015, July). Networks and the macroeconomy: An empirical exploration. *NBER Working Paper*, *21344*.

Acemoglu, D., Carvalho, V. M., Asuman Ozdaglar, & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, *80*(5), 1977–2016. Retrieved from http://doi.wiley.com/10.3982/ECTA9623 doi: 10.3982/ECTA9623

Alabrese, E., Fetzer, T., & Wang, S. (2024). Leveling up, by pushing down? *mimeo*(486).

Alfaro, L., & Chor, D. (2023, September). Global supply chains: The looming "great reallocation" [Working Paper]. (31661). Retrieved from http://www.nber.org/papers/w31661 doi: 10.3386/w31661

Andersen, A. L., Huber, K., Johannesen, N., Straub, L., & Vestergaard, E. T. (2022). *Disaggregated economic accounts* (Tech. Rep.). National Bureau of Economic Research.

Antràs, P. (2020, November). De-globalisation? global value chains in the post-covid-19 age. *NBER Working Paper Series*(w28115). doi: 10.3386/w28115

Antràs, P., & Chor, D. (2018). On the measurement of upstreamness and downstreamness in global value chains. In L. Y. Ing & M. Yu (Eds.), *World trade evolution: Growth, productivity and employment* (pp. 1126–1194). Taylor & Francis.

Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012, May). Measuring the upstreamness of production and trade flows. *American Economic Review*, *102*(3), 412-16. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.102.3.412 doi: 10.1257/aer.102.3.412

Antràs, P., Fort, T. C., & Tintelnot, F. (2017). The margins of global sourcing. *The American Economic Review*, *107*(9), 2514–2564. Retrieved from https://www.jstor.org/stable/10.2307/26527920

Arezki, R., Fetzer, T., & Pisch, F. (2017, jul). On the comparative advantage of U.S. manufacturing: Evidence from the shale gas revolution. *Journal of International Economics*, *107*, 34–59. Retrieved from https://linkinghub.elsevier.com/retrieve/

pii/S0022199617300284 doi: 10.1016/j.jinteco.2017.03.002

Ash, E., & Hansen, S. (2023, July). Text algorithms in economics. *Annu. Rev. Econom.*.

Bachmann, R., Baqaee, D. R., Bayer, C., Kuhn, M., Löschel, A., Moll, B., ... Schularick, M. (2022, March). *What if? the economic effects for germany of a stop of energy imports from russia* (Tech. Rep.). ECONtribute.

Baldwin, R., & Freeman, R. (2022). Risks and global supply chains: What we know and what we need to know. *Annual Review of Economics*, *14*, 153–180. Retrieved from https://www.annualreviews.org/content/journals/10.1146/annurev-economics-051420-113737 doi: 10.1146/annurev-economics-051420-113737

Baqaee, D. R., & Farhi, E. (2019). The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. *Econometrica*, *87*(4), 1155–1203. doi: 10.3982/ECTA15202

Baqaee, D. R., & Farhi, E. (2021). Keynesian production networks and the covid-19 crisis: A simple benchmark. *AEA Papers and Proceedings*, *111*, 272–276. doi: 10.1257/pandp.20211107

Bar-Gill, O., & Fershtman, C. (2021). Behavioral stability and brand loyalty. *American Economic Journal: Microeconomics*, *13*(3), 279–307.

Bartik, A., Gupta, A., & Milo, D. (2023). The Costs of Housing Regulation: Evidence from Generative Regulatory Measurement. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4627587 (Working Paper)

Bartolucci, S., Caccioli, F., Caravelli, F., & Vivo, P. (2024). *Correlation between upstreamness and downstreamness in random global value chains.* Retrieved from https://arxiv.org/abs/2303.06603

Behrens, K., Coreos, G., & Mion, G. (2013). Trade crisis? what trade crisis? *The Review of Economics and Statistics*, *95*(2), 702–709. Retrieved from https://www.jstor.org/stable/43554414

Bems, R., Johnson, R. C., & Yi, K.-M. (2011). Vertical linkages and the collapse of global trade. *American Economic Review*, *101*(3), 308–312.

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2012). The empirics of firm heterogeneity and international trade. *Annual Review of Economics*, *4*, 283–313. Retrieved from https://www.jstor.org/stable/42949939

Bernard, A. B., Moxnes, A., & Ulltveit-Moe, K. H. (2018). Two-sided heterogeneity and trade. *Review of Economics and Statistics*.

Blanga-Gubbay, M., & Rubínová, S. (2023, November). Trade in goods between hypothetical east and west blocs: Trends and impact of geopolitical tensions. *Staff Working Paper: Research ERSD-2023-10*. (Manuscript date: 30 November 2023)

Boehm, C. E., Flaaen, A., & Pandalai-Nayar, N. (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. *The Review of Economics and Statistics*, *101*(1), 60–75. doi: 10.1162/rest_a_00750

Born, B., Müller, G. J., Schularick, M., & Sedláček, P. (2018). The Costs of Economic Nationalism: Evidence from the Brexit Experiment. *CEPR Discussion Paper No. 12454*. Retrieved from http://users.ox.ac.uk/{~}econ0506/Documents/Working/Brexit{_}new.pdf

Breinlich, H., Leromain, E., Novy, D., & Sampson, T. (2017). The Consequences of the Brexit Vote for UK Inflation and Living Standards: First Evidence. *mimeo*(11). Retrieved from http://cep.lse.ac.uk/pubs/download/brexit11{_}technical{_}paper.pdf

Bricongne, J.-C., Fontagné, L., Gaulier, G., Taglioni, D., & Vicard, V. (2012). Firms and the global crisis: French exports in the turmoil. *Journal of International Economics*, *87*(1), 134–146. doi: 10.1016/j.jinteco.2011.07.002

Bronnenberg, B. J., Dube, J.-P., & Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, *102*(6), 2472–2508.

Buehler, M. J. (2024). Accelerating Scientific Discovery with Generative Knowledge Extraction, Graph-Based Representation, and Multimodal Intelligent Graph Reasoning. *Machine Learning: Science and Technology*, *5*(4), 045005. Retrieved

from https://iopscience.iop.org/article/10.1088/2632-2153/ad7228 doi: 10.1088/2632-2153/ad7228

Bureau of Economic Analysis, U.S. Department of Commerce. (2022). *BEA Data Archive: Gross Domestic Product by Industry and Input-Output Statistics.* Retrieved from https://www.bea.gov/itable/ (Accessed on April 21, 2024)

Carvalho, V., & Gabaix, X. (2013, August). The great diversification and its undoing. *The American Economic Review*, *103*(5), 1697–1727. doi: 10.1257/aer.103.5.1697

Caselli, F., Koren, M., Lisicky, M., & Tenreyro, S. (2020, February). Diversification through trade. *Q. J. Econ.*, *135*(1), 449–502.

Costa, R., Dhingra, S., & Machin, S. (2019, May). *Trade and worker deskilling* (Tech. Rep.). CEP Discussion Paper No. 1622.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th acm conference on recommender systems* (pp. 191–198).

Crosignani, M., Han, L., Macchiavelli, M., & Silva, A. F. (2023). Geopolitical risk and decoupling: Evidence from u.s. export controls [Working Paper]. *Federal Reserve Bank of New York Staff Reports*. Retrieved from https://ssrn.com/abstract=4563485

Dell, M. (2024, August). Deep learning for economists. *National Bureau of Economic Research Working Paper Series*(32768). Retrieved from http://www.nber.org/papers/w32768

Dhyne, E., Kikkawa, A. K., Mogstad, M., & Tintelnot, F. (2020, October). Trade and domestic production networks. *Rev. Econ. Stud.*, *88*(2), 643–668.

di Giovanni, J., Levchenko, A. A., & Méjean, I. (2014). Firms, destinations, and aggregate fluctuations. *Econometrica*, *82*(4), 1303–1340.

Douch, M., Edwards, T. H., & Soegaard, C. (2018, August). *The trade effects of the brexit announcement shock* [Working Paper]. Coventry. Retrieved from http://wrap.warwick.ac.uk/107767/

Egorov, K., Korovkin, V., Makarin, A., & Nigmatulina, D. (2024). Trade sanctions.

(Forthcoming or in-progress publication)

Elliott, M., Golub, B., & Leduc, M. V. (2022). Supply network formation and fragility. *American Economic Review*, *112*(8), 2701–2747.

Elliott, M., & Jackson, M. O. (2024). Supply chain disruptions, the structure of production networks, and the impact of globalization. *Cambridge Working Papers in Economics*, *CWPE2424*. doi: 10.17863/CAM.109525

Erdos, P., Rényi, A., et al. (1960). On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, *5*(1), 17–60.

Fajgelbaum, P., & Khandelwal, A. (2024, October). The value of de minimis imports. (Presented at Yale NBER)

Fajgelbaum, P. D., Goldberg, P. K., Kennedy, P. J., & Khandelwal, A. K. (2020, feb). The Return to Protectionism. *Quarterly Journal of Economics*, *135*(1), 1–55. Retrieved from http://www.nber.org/papers/w25638.pdfhttps://academic.oup.com/qje/article/135/1/1/5626442 doi: 10.1093/qje/qjz036

Fally, T. (2012). Production staging: Measurement and facts. *Working Paper*.

Feenstra, R., & Jensen, J. B. (2012). Evaluating estimates of materials offshoring from us manufacturing. *Economics Letters*, *117*(1), 170–173.

Fetzer, T., & Schwarz, C. (2021). Tariffs and politics: Evidence from trump's trade wars. *The Economic Journal*, *131*(636), 1717–1741. Retrieved from https://doi.org/10.1093/ej/ueaa122 (Published: 09 December 2020) doi: 10.1093/ej/ueaa122

Freeman, R., & Baldwin, R. (2020). Supply chain contagion waves: Thinking ahead on manufacturing contagion and reinfection from the covid concussion. *VoxEU*.

Freeman, R., Manova, K., Prayer, T., & Sampson, T. (2022). Uk trade in the wake of brexit.

Freund, C., Mulabdic, A., & Ruta, M. (2023). *Is us trade policy reshaping global supply chains?* (Tech. Rep. No. 10593). Washington, DC: World Bank. (License: CC BY 3.0 IGO) doi: http://hdl.handle.net/10986/40558

Frésard, L., Hoberg, G., & Phillips, G. M. (2020, July). Innovation activities and integra-

tion through vertical acquisitions. *The Review of Financial Studies*, *33*(7), 2937–2976. Retrieved from https://doi.org/10.1093/rfs/hhz106 doi: 10.1093/rfs/hhz106

Garetto, S. (2013). Input sourcing and multinational production. *American Economic Journal: Macroeconomics*, *5*(2), 118–151. doi: 10.1257/mac.5.2.118

Garg, P., & Fetzer, T. (2024). Causal claims in economics. *CESifo Working Paper*. Retrieved from https://www.cesifo.org/en/publications/2024/working-paper/causal-claims-economics

Giesecke, O. (2024, February 22). Ai at the frontier of economic research. *SSRN*. Retrieved from https://ssrn.com/abstract=4736003 doi: 10.2139/ssrn.4736003

Goldberg, Y., & Levy, O. (2014). word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.

Gopinath, G., Gourinchas, P.-O., Presbitero, A., & Topalova, P. B. (2024). Changing global linkages: A new cold war? *Available at SSRN*.

Griffith, R., O'Connell, M., & Smith, K. (2019). Habit formation and consumer responses to gasoline prices. *American Economic Journal: Economic Policy*, *11*(4), 129–159.

Grossman, G. M., Helpman, E., & Redding, S. J. (2024). When tariffs disrupt global supply chains. *American Economic Review*, *114*(4), 988–1029. doi: 10.1257/aer.20211519

Handbury, J., & Weinstein, D. E. (2014). Goods prices and availability in cities. *Review of Economic Studies*, *81*(3), 1583–1620.

Handley, K., & Limão, N. (2017). Policy uncertainty, trade, and welfare: Theory and evidence for china and the united states. *American Economic Review*, *107*(9), 2731–2783.

Hansen, S., Lambert, P. J., Bloom, N., Davis, S. J., Sadun, R., & Taska, B. (2023, March). *Remote work across jobs, companies, and space* (No. 31007).

Hassan, T. A., Hollander, S., Lent, L. V., & Tahoun, A. (2024, feb). The Global Impact of Brexit Uncertainty. *The Journal of Finance*, *79*(1), 413–458. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/jofi.13293 doi: 10.1111/jofi.13293

Horowitz, K. J., & Planting, M. A. (2006). *Concepts and methods of the u.s.*

input-output accounts. U.S. Bureau of Economic Analysis, U.S. Department of Commerce. Retrieved from https://www.bea.gov/sites/default/files/methodologies/IOmanual_092906.pdf (Updated in April 2009)

Huang, H., Manova, K., Perelló, O., & Pisch, F. (2024, July). Firm heterogeneity and imperfect competition in global production networks. *CEP discussion paper*(2020).

INEGI. (2023). *Subsystem of Economic Information: Input-Output Matrix (IOM). Base year 2018.* Retrieved from https://www.inegi.org.mx/programas/mip/2018/ (Accessed on August 16, 2024)

Jewson, J., Li, L., Battaglia, L., Hansen, S., Rossell, D., & Zwiernik, P. (2022). *Graphical model inference with external network data* (CEPR Discussion Paper No. DP17638). Paris and London: CEPR Press.

Johnson, R. C. (2018). Measuring global value chains. *Annual Review of Economics*, *10*, 207–236. Retrieved from https://www.jstor.org/stable/10.2307/26773849

Juhász, R., Lane, N., & Rodrik, D. (2023). The New Economics of Industrial Policy. *SSRN Electronic Journal*(August), 1–46. doi: 10.2139/ssrn.4542861

Karbevska, L., & Hidalgo, C. A. (2023). Mapping global value chains at the product level. *arXiv.org*.

Korinek, A. (2023). Generative ai for economic research: Use cases and implications for economists. *Journal of Economic Literature*, *61*(4), 1281–1317.

Kosasih, E. E., et al. (2024). Towards knowledge graph reasoning for supply chain risk management using graph neural networks. *International Journal of Production Research*, *62*(15), 5596–5612.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review*, *70*(5), 950–959.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*(6), 1695–1725.

Mikolov, T. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Oberfield, E. (2018). A theory of input-output architecture. *Econometrica*, *86*(2), 559–589.

Pan, S., et al. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*.

Peifeng, L. I. U., et al. (2024). Joint knowledge graph and large language model for fault diagnosis and its application in aviation assembly. *IEEE Transactions on Industrial Informatics*.

Pierce, J. R., & Schott, P. K. (2012). A concordance between ten-digit us harmonized system codes and sic/naics product classes and industries. *Journal of Economic and Social Measurement*, *37*(1-2), 61–96.

Rong, X. (2014). word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.

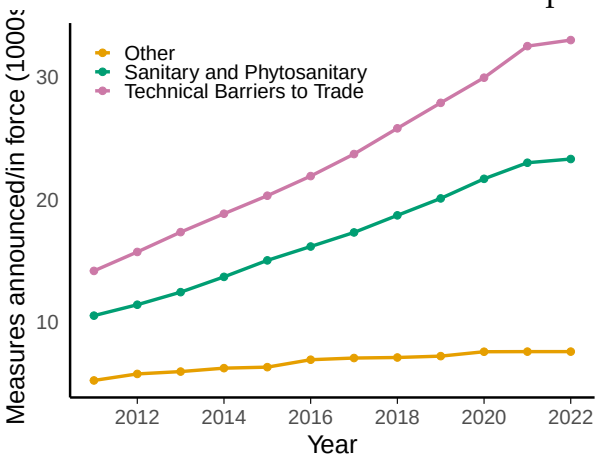
Schindler, Y., & Lambert, P. J. (2024, November). Bad bank, bad luck? Evidence from 1 Million Firm-Bank Relationships. *Working Paper*. Retrieved from https://www.yannickschindler.com/research

Utar, H., Torres Ruiz, L. B., & Zurita, A. C. (2023). *The us-china trade war and the relocation of global value chains to mexico* (Tech. Rep. No. 10638). CESifo. Retrieved from https://ssrn.com/abstract=4568757 doi: http://dx.doi.org/10.2139/ssrn.4568757

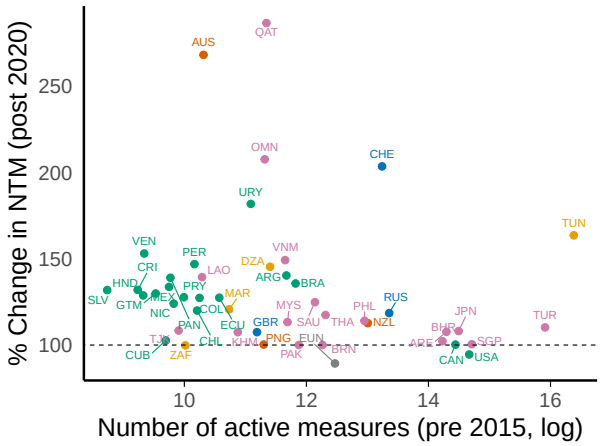
7 Main Figures and Tables

Figure 1: Globalization under stress

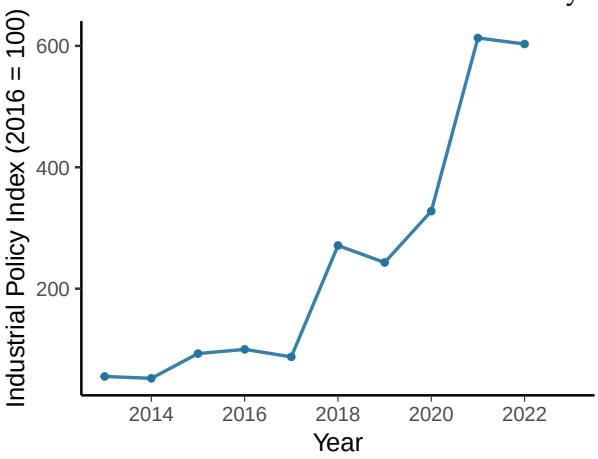
Panel A: Restrictive trade measures in place



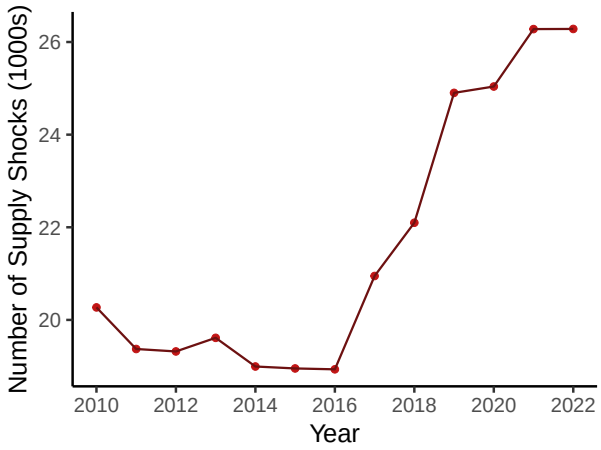
Panel B: Growth in non-tariff measures



Panel C: Proliferation of Industrial Policy

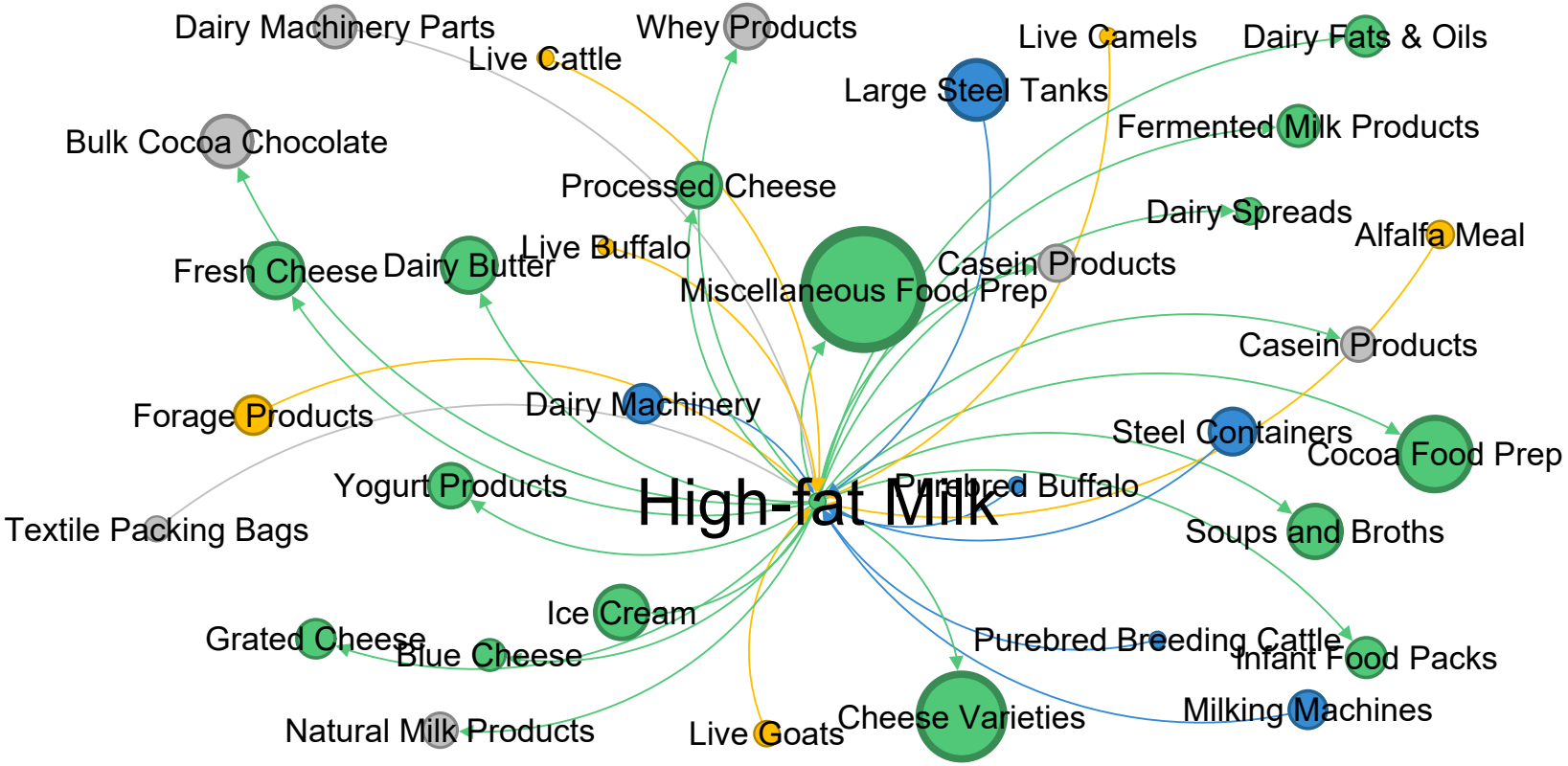


Panel D: Persistent supply shocks



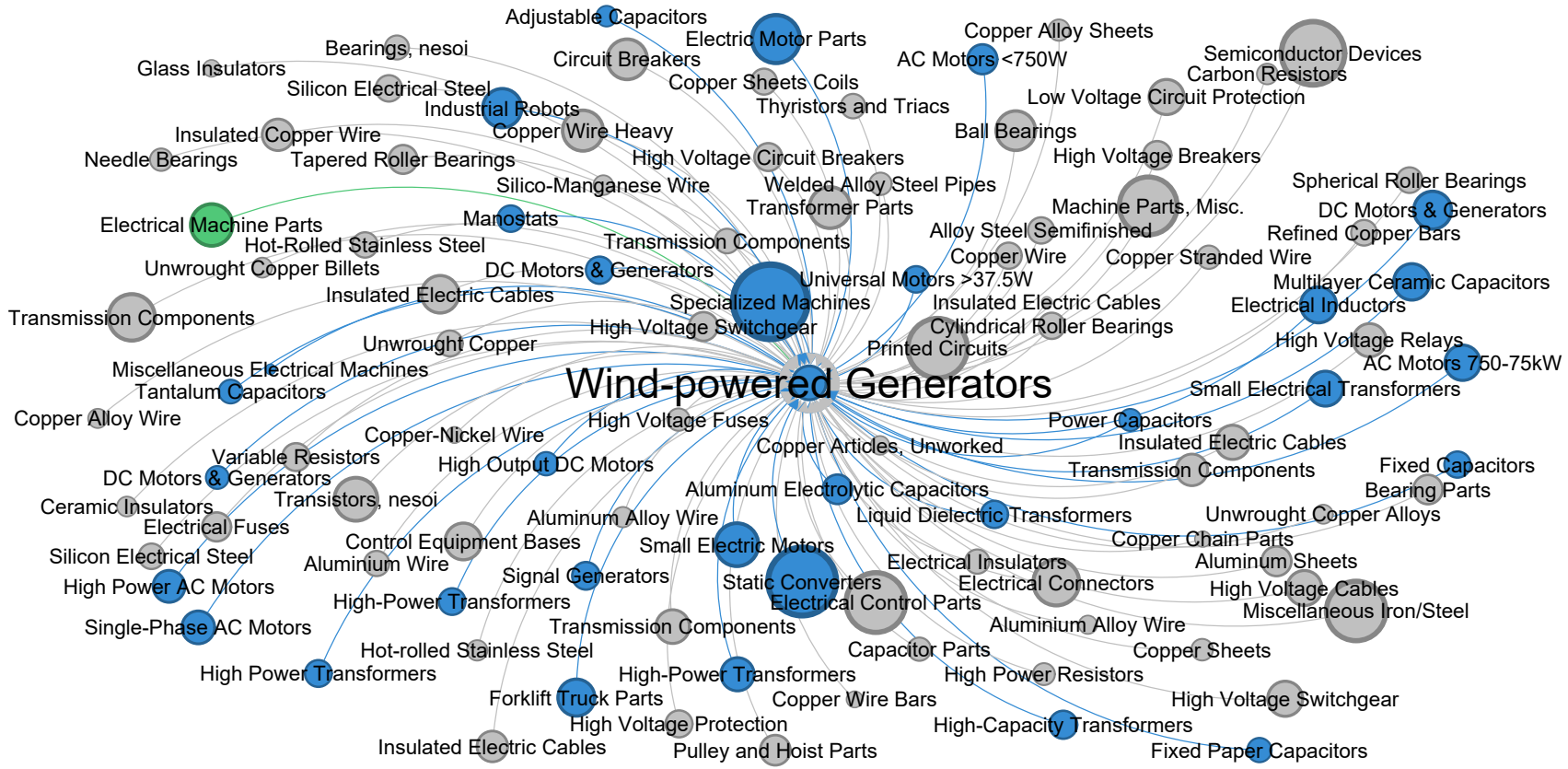
Note: Panel A measures the number of active or announced restrictive trade measures recorded with the WTO in 1000s. Panel B plots the level and change of non-tariff barriers applicable at the HS6 good level from TRAINS UNCTAD. Focus is on sample of countries for which NTM data is available before or in 2015 and in the most recent year. The vertical axis captures the relative change in the NTM relative to the baseline by country. Panel C documents the proliferation of industrial policies using the measure from (Juhász, Lane, & Rodrik, 2023). Panel D presents a measure of the number of country-by-product pairs that have had semi-persistent import price increases. We describe this measure in more detail in Section 5.

Figure 2: One-degree Network for High-Fat Milk (HS: 0401.50)



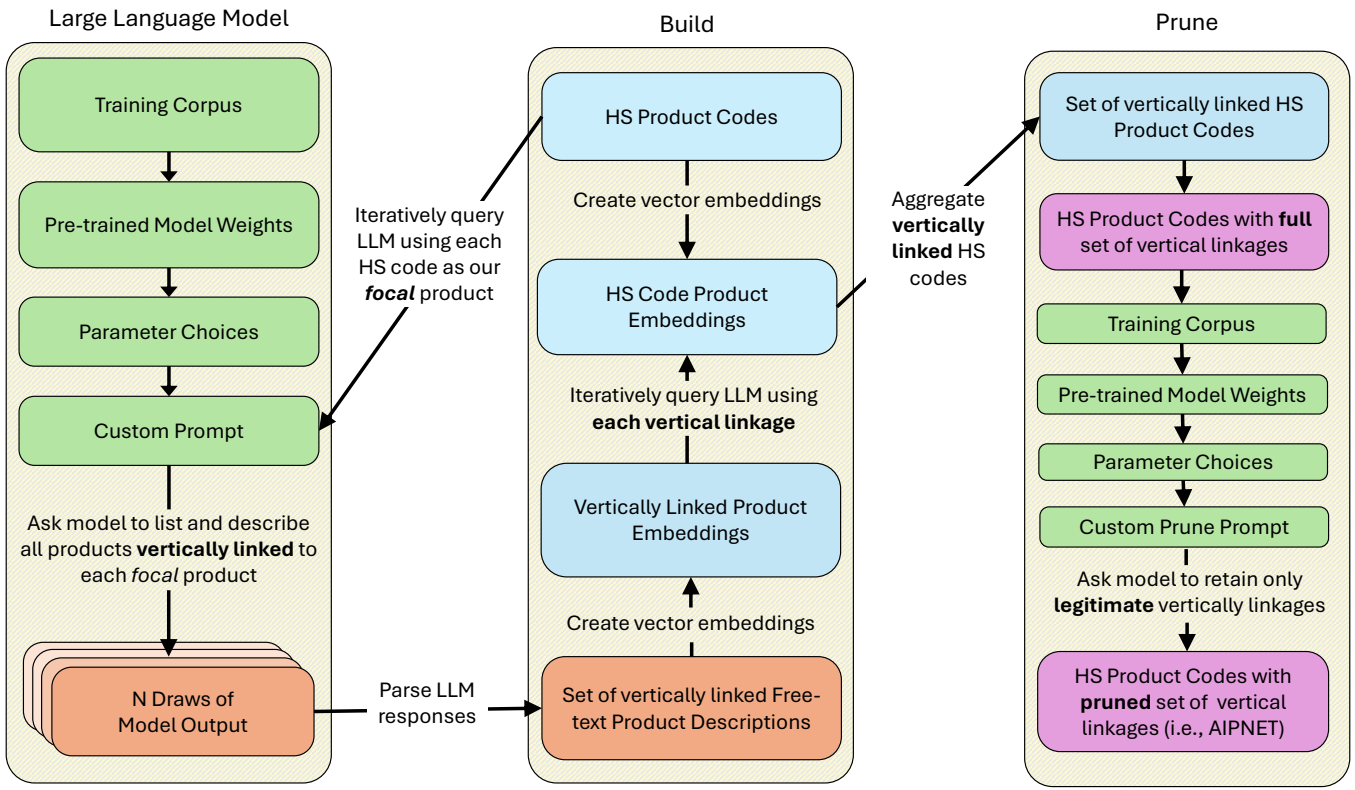
Note: This figure presents the **one-degree ego network** of Full-fat Milk and Cream (HS: 0401.50) in AIPNET, shortened to "High-fat Milk". Nodes represent six-digit HS product codes, and the focal node (Milk) has a larger font size. Nodes are color-coded by the BEC5 end-use classification: green for consumption goods, blue for capital goods, yellow for raw materials (intermediate-primary goods), and grey for intermediate-processed goods. Edges are colored by the source node. As a consumer-facing non-durable good, Full-fat Milk and Cream serves as an input to various dairy products, such as processed cheese (HS: 0406.30) and miscellaneous food preparations (HS: 2106.90). For a more extensive view of HS Code linkages, please refer to our dataset and https://aipnet.io/.

Figure 3: One-degree Network for Wind-powered Generators (HS: 8502.31)



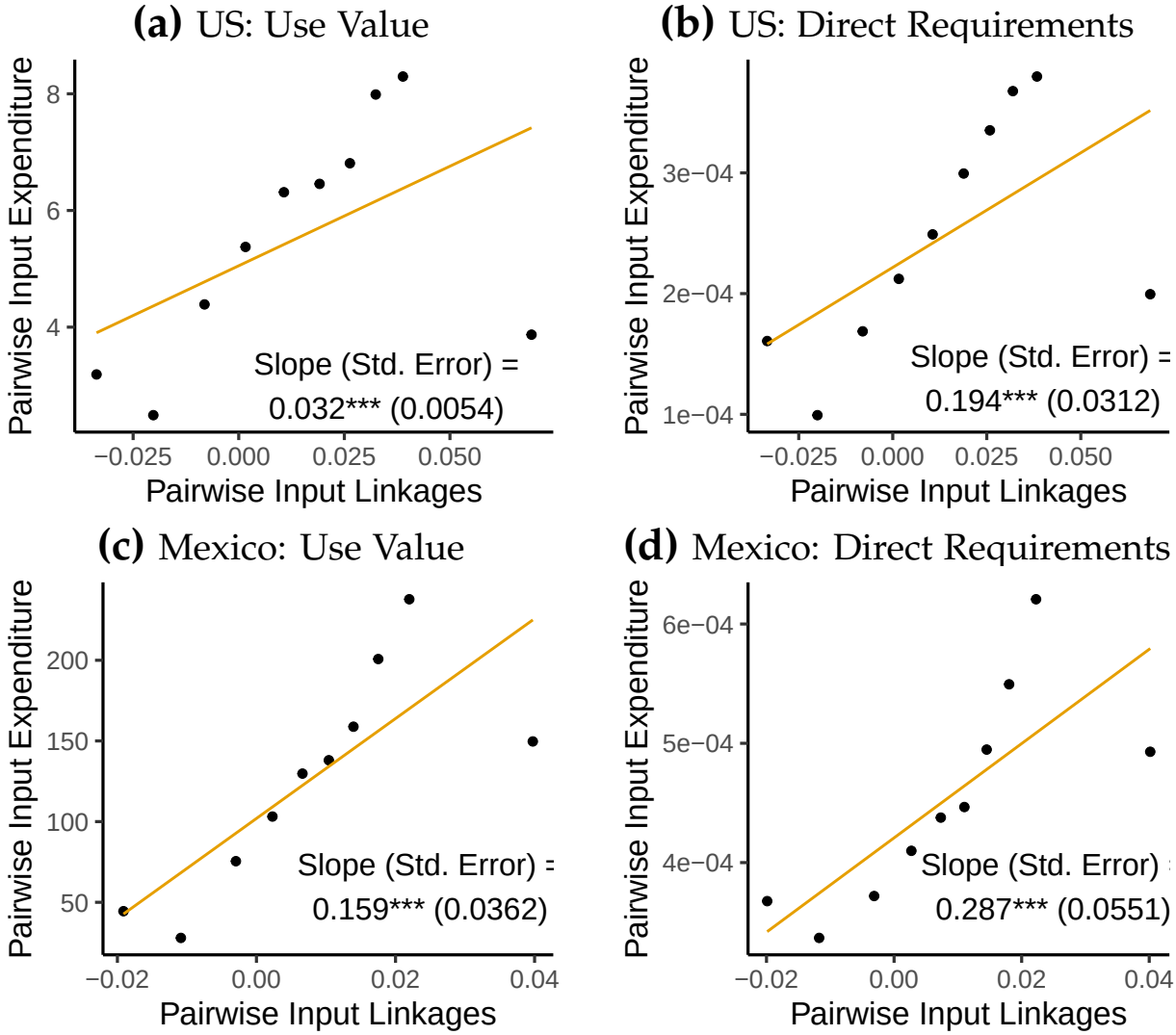
Note: This figure presents the **one-degree ego network** of Wind-powered Generators (HS: 8502.31) in AIPNET. Nodes represent six-digit HS product codes, with the focal node (Wind Generators) highlighted in larger font. Nodes are colored based on BEC5 end-use classification: green for consumption goods, blue for capital goods, yellow for raw materials (intermediate-primary goods), and grey for intermediate-processed goods. Edges are colored by the source node. As an essential capital input, Wind Generators require various intermediary goods, such as electrical control parts (HS: 8538.90) and hot-rolled stainless steel (HS: 7219.23). For more comprehensive HS Code connections, visit https://aipnet.io/.

Figure 4: Flowchart of steps to build AIPNET



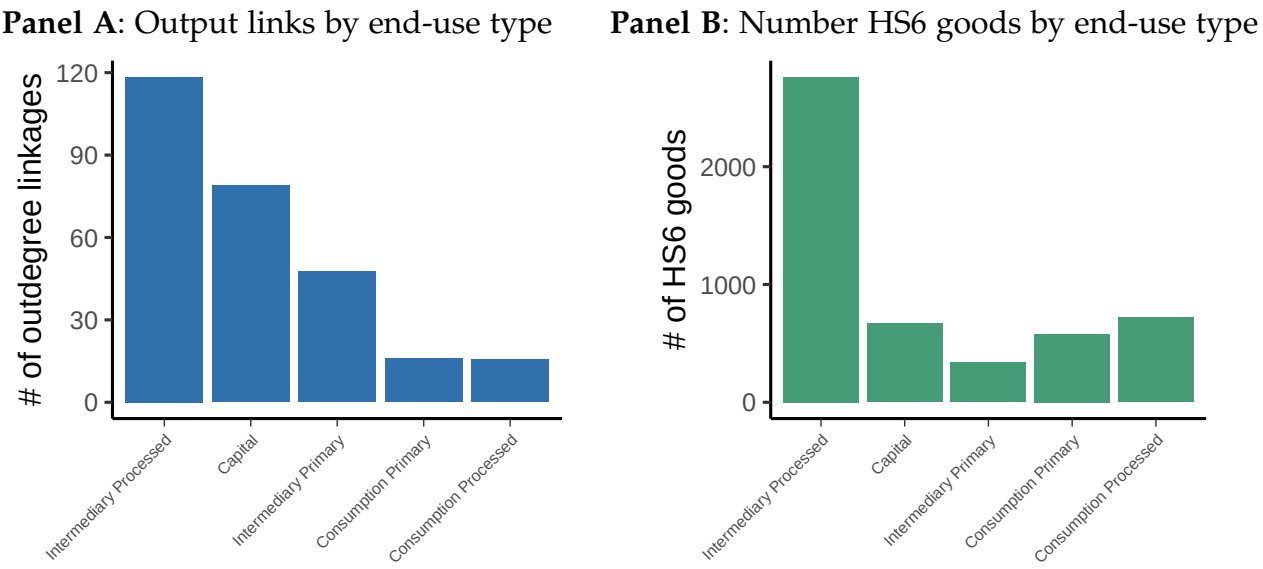
Note: This flowchart illustrates the structured pipeline used to construct AIPNET, integrating advanced AI techniques with rigorous data processing. The process is divided into three main components, each represented by color-coded sections. The leftmost section, shaded in green, represents the Large Language Model (LLM) setup and querying process. It begins with a pre-trained model, fine-tuned through parameter selection and custom prompting to identify input-output relationships for each HS code. The LLM iteratively generates descriptions of vertically linked products, which are aggregated across multiple draws to account for the model's probabilistic nature. The middle section, shaded in beige and blue, represents the build stage, where product embeddings for both HS product codes and their vertically linked free-text descriptions are created. These embeddings capture detailed representations of each product, which assist in matching relationships accurately. The rightmost section, shaded with final node shaded in pink, shows the aggregation of vertically linked HS codes, followed by a pruning process. In this stage, aggregated linkages are refined through additional AI querying to retain only legitimate input-output relationships. This step ensures that AIPNET's connections are robust, with edge overlap used as a hyperparameter to align network sparsity with U.S. Input-Output tables and maximize predictive accuracy (R^2). This approach culminates in a refined AIPNET network that accurately represents real-world production linkages.

Figure 5: Recasting AIPNET into coarse Input-Output Relationships across industry-classes is strongly correlated with Official Data



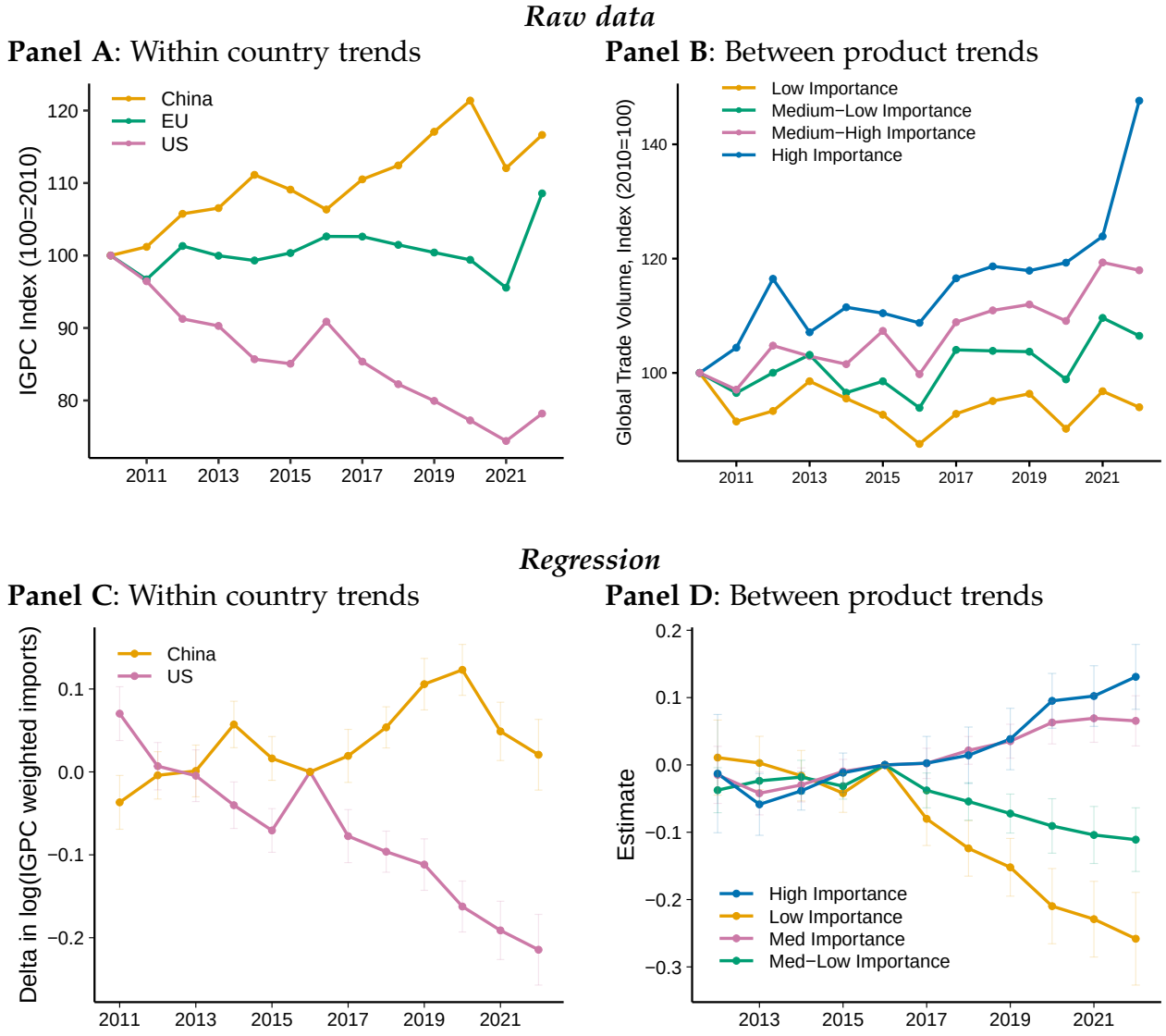
Note: The binscatter plots compare our core network to (a) the Use table, and (b) Direct Requirements. Each plot shows the relationship between residualized, standardized Input-Output table values and standardized network scores, both aggregated from the HS6 level to the BEA-industry level using 10 bins based on the network score. The points represent the mean of binned observations, and the red line shows the linear fit across these bins. We exclude observations with residuals in the top and bottom 5% to avoid outliers. The slope coefficient (standard error) from the linear regression for each plot is: (a) 0.035*** (0.005), (b) 0.219*** (0.033), and (c) 0.208*** (0.031), indicating that the network score explains a significant proportion of the variation in Input-Output relationships, especially in the Requirements tables. *** denotes significance at the 0.001 level.

Figure 6: Relationship Between BEC Classification and Number of Output Uses for HS6 Goods



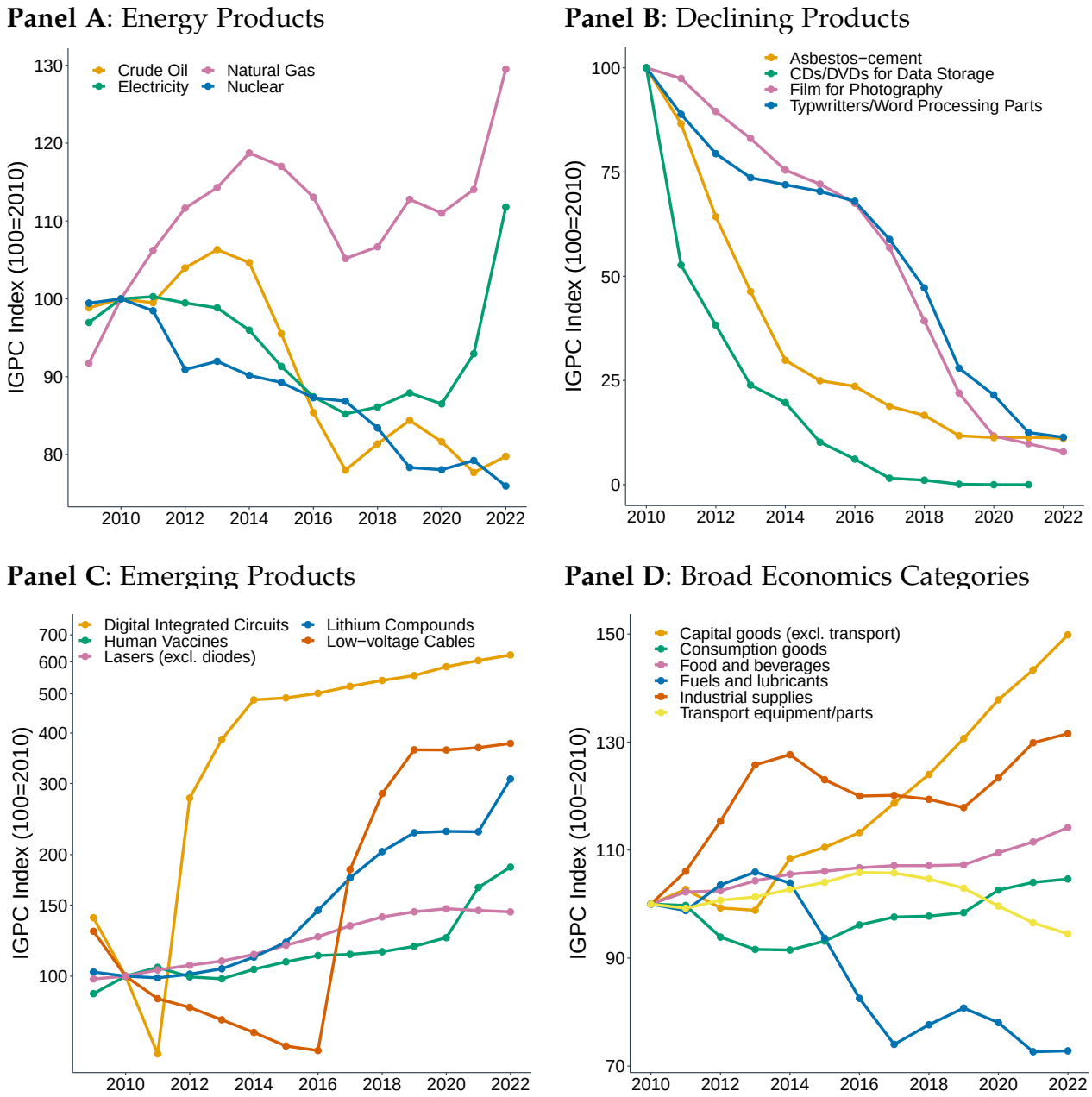
Note: This figure illustrates how HS6 goods classified under different Broad Economic Categories (BEC) relate to the number of output uses in the AIPNET network. **Panel A** shows the average number of output links (i.e., the number of goods each good is used to produce) for each end-use type. Intermediary processed goods, capital goods, and primary intermediary goods have, on average, many more output uses, indicating that they serve as inputs in numerous production processes before reaching final consumption. **Panel B** displays the count of HS6 goods by their end-use classification, highlighting the distribution of goods across different categories. Goods classified for final consumption have notably fewer output uses, reinforcing their role as end products in the supply chain. Understanding the positioning and connectivity of goods within the network provides insight into their roles in production processes and supply chains, which is critical for analyses related to industrial policy and economic shocks.

Figure 7: Diverging trends in IGPC weighted imports across countries and product centrality levels



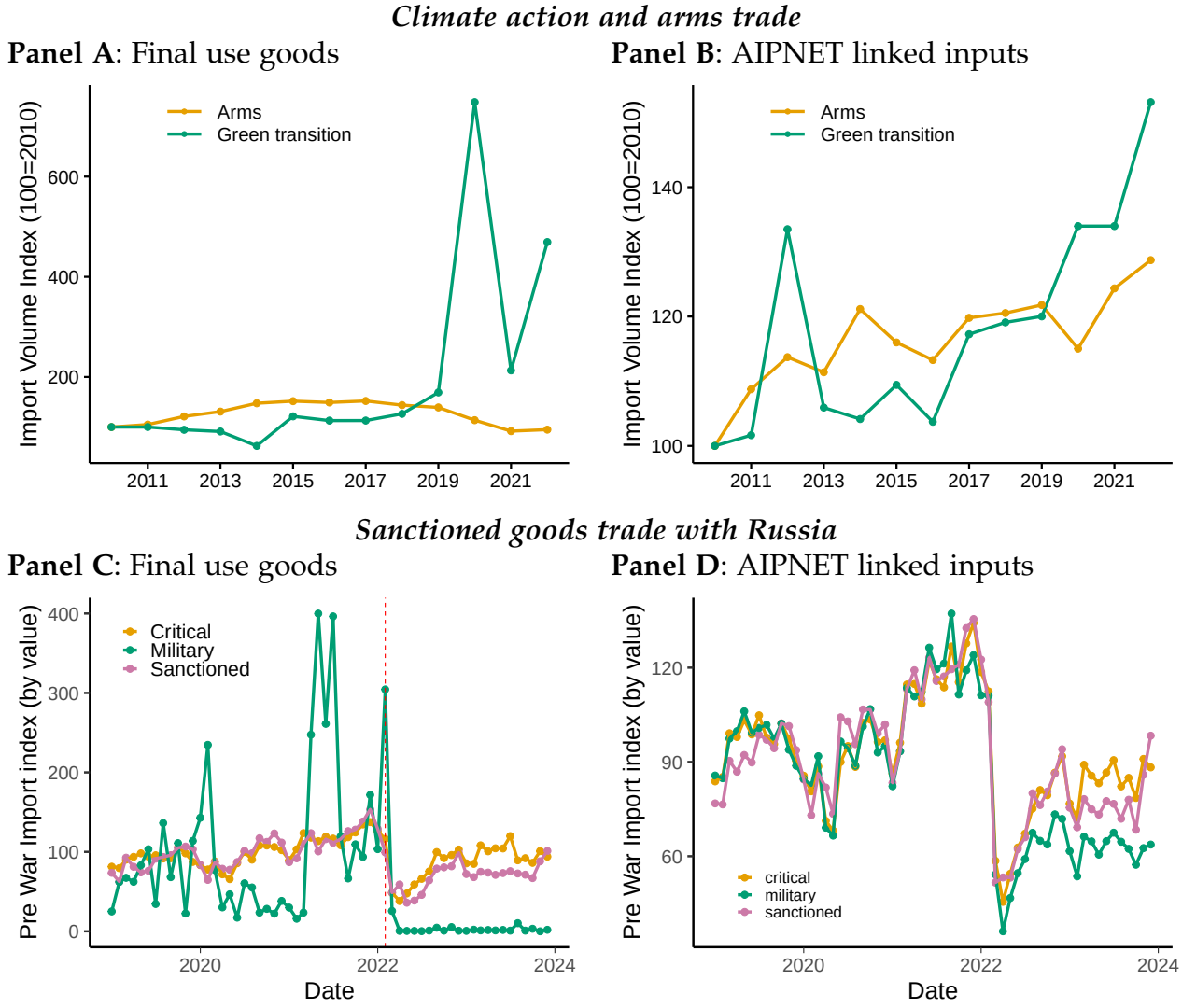
Note: Figure presents trends in the IGPC weighted import volumes between countries or products. The top two panels focus on raw country-level or product-level aggregated data. The bottom panel presents similar evidence within a regression framework working with a disaggregated country-by-product panel dataset. Panel A shows within-country IGPC index trends for China, the EU, and the US, reflecting the centrality of each country's imports in global production networks. Panel B presents trends across products grouped by IGPC importance quartiles, showing that products with higher baseline centrality have generally increased in importance over time.

Figure 8: Changes in Product Importance in Global Supply Chains (IGPC Measure)



Note: This figure shows changes in product "importance" in global supply chains using the IGPC (Integrated Global Product Centrality) measure (constructed by combining AIPNET and trade-data, see Section 3.5), indexed to 2010=100. Panel A shows trends for key energy products including electricity, natural gas, nuclear power, and crude oil. Panel B displays products that have experienced significant decline in "importance", such as CDs/DVDs, photographic film, asbestos-cement, and typewriter parts. Panel C presents emerging products with increasing importance, including lithium compounds, vaccines, digital integrated circuits, low-voltage cables, and lasers. Panel D aggregates products by BEC (Broad Economic Categories) classification, showing trends across major product categories. The IGPC measure captures a product's importance as an input in global production networks, weighted by trade volumes.

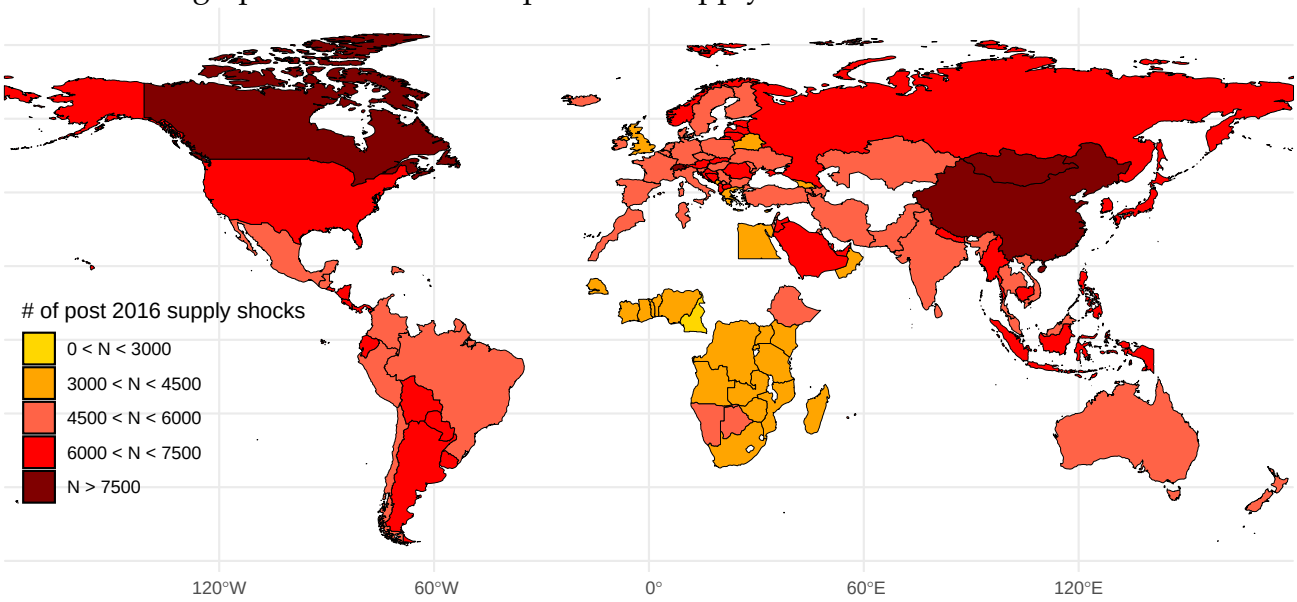
Figure 9: Import volume index of broad group of final use goods as seen through AIPNET linkages



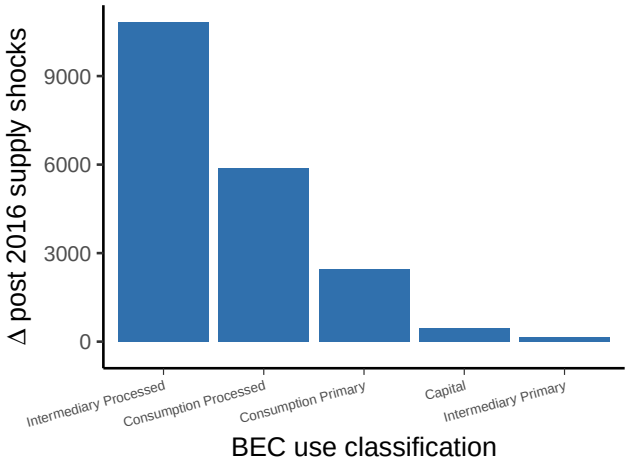
Note: Figure shows the trend in imports globally (top panel) or into Russia (bottom panel) of a set of goods and the AIPNET linked inputs. The top row focuses on import volume of goods related to climate action and contrasts this with arms. A real import volume index is constructed using an imputed constant average global unit price. We focus on Climate Action Goods, 841861/841581 - Heat Pumps, 850231 - Wind Turbines, 854140 - Solar Panels, 840110 - Nuclear Reactors, 841011 - Hydroelectric Turbines, 870380 - Electric Vehicles, 850760 - Electric Storage; for the Military goods we consider 930190 - Military weapons, 930200 - Revolvers and pistols, 1930320 - Shotguns, 930120 - Artillery weapons 930621/930630 - Cartridges, 930690 - Munitions, 871000 - Armored vehicles, 930700 - Swords and bayonets, 880220/880230 - Helicopters, 880240 - Unmanned aircraft, 880250 - Spacecraft 880220/880230/880212 - Military drones, 880240 - Military aircraft, 880330 - Aircraft parts. The bottom panel focuses on imports into Russia by third countries of goods in nominal terms relative to an import value index prior to the war in Ukraine starting. We note that critical or sanctioned goods direct import value decrease only modestly. Yet, imports of inputs for military goods contrasts much more pronouncedly in Panel D.

Figure 10: Characterisation of the Supply Shock

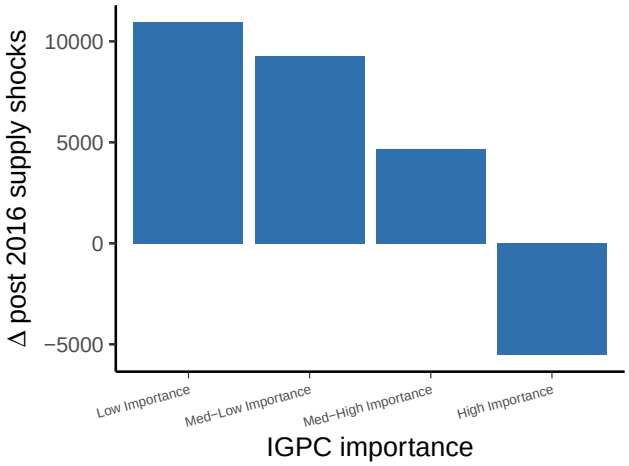
Panel A: Geographic distribution of post 2016 supply shocks



Panel B: Characterisaton by BEC

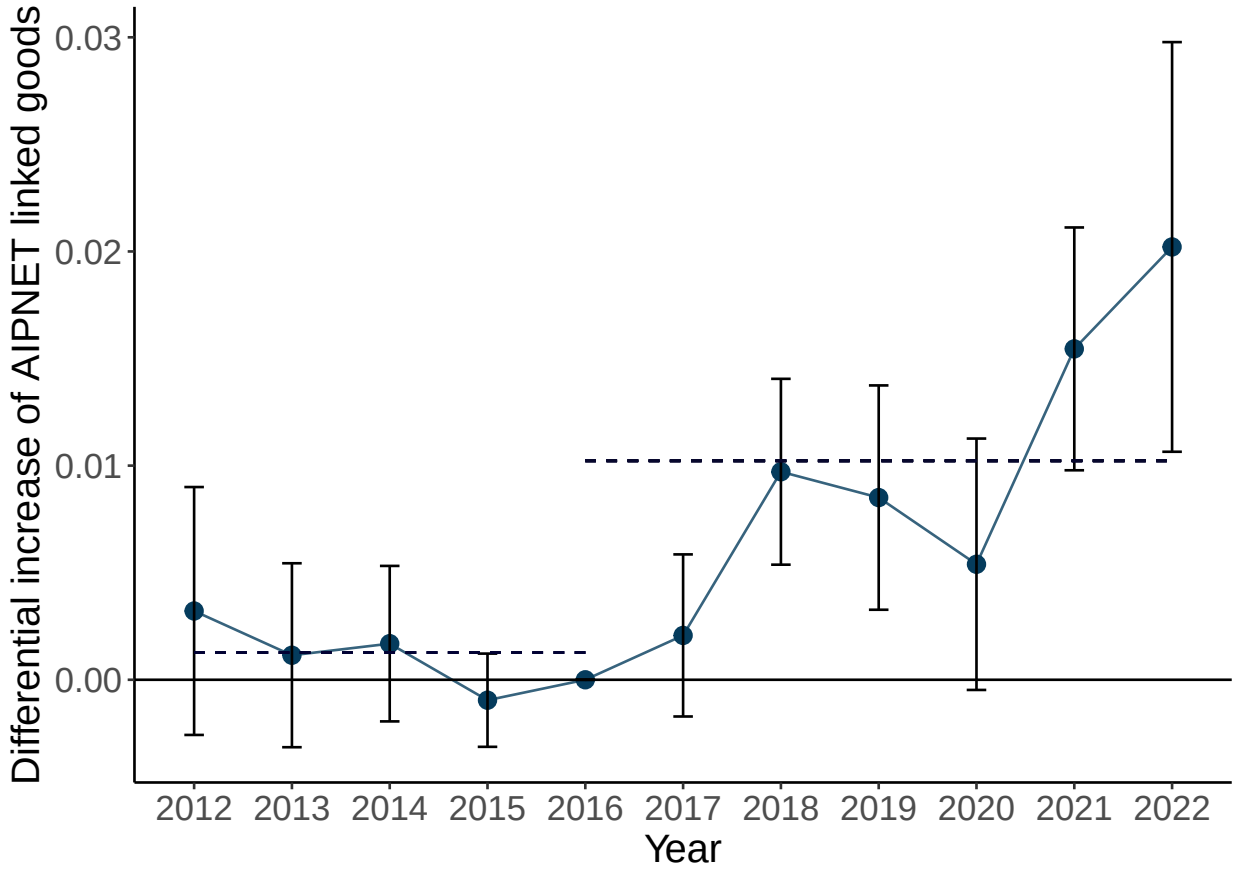


Panel C: Characterisation by IGPC



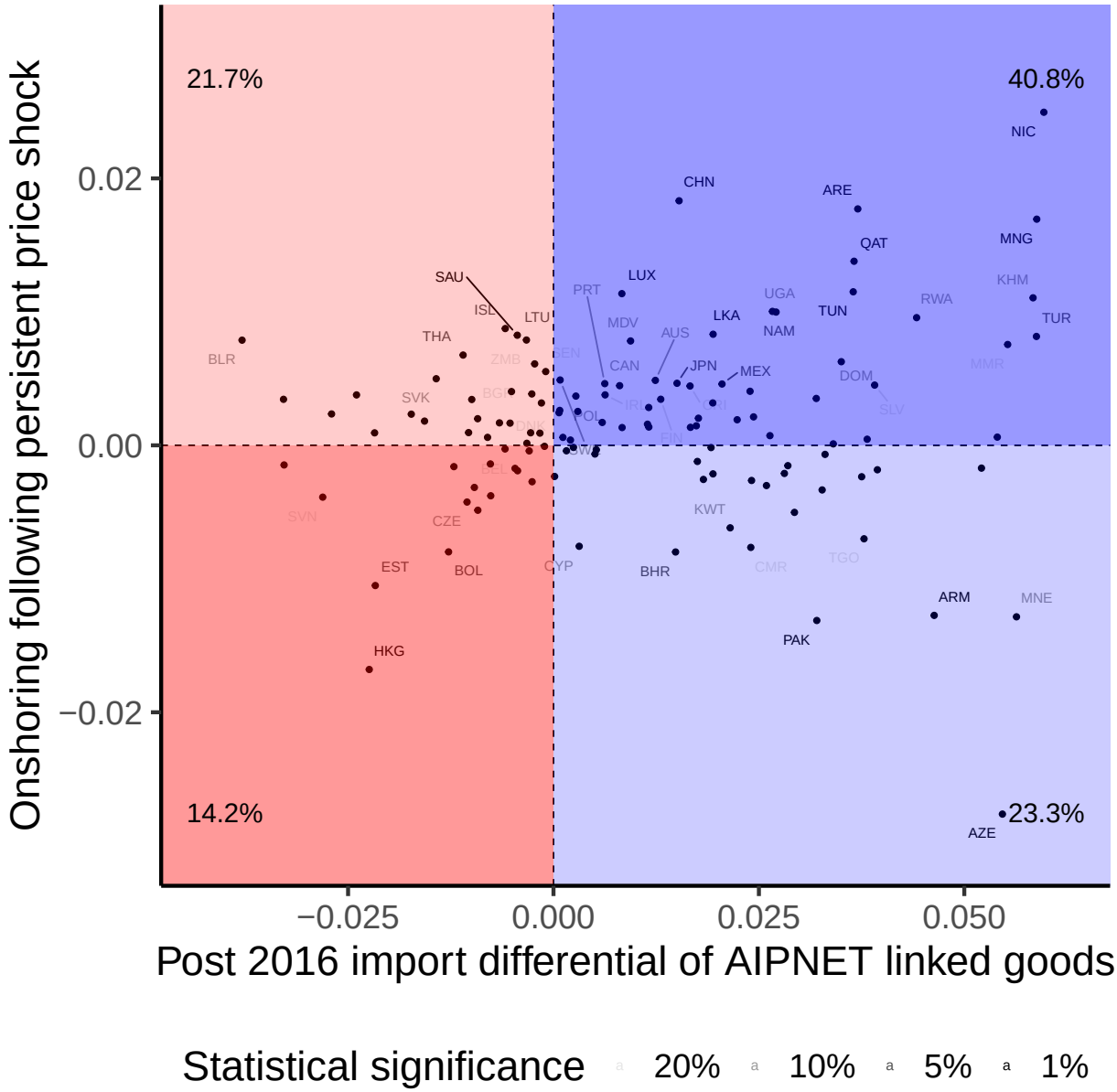
Note: Figure characterises the increase in the identified supply shocks that have been presented in Panel D of Figure 1 since 2016. We count the number of supply shocks at the commodity level by number of countries before and after 2016 and characterise the product level difference. This is then aggregated up and presented by the HS code's BEC classification in Panel A or by the HS codes IGCP quartile in Panel B.

Figure 11: Global trend towards increasing imports of goods with many AIPNET downstream linkages since 2016



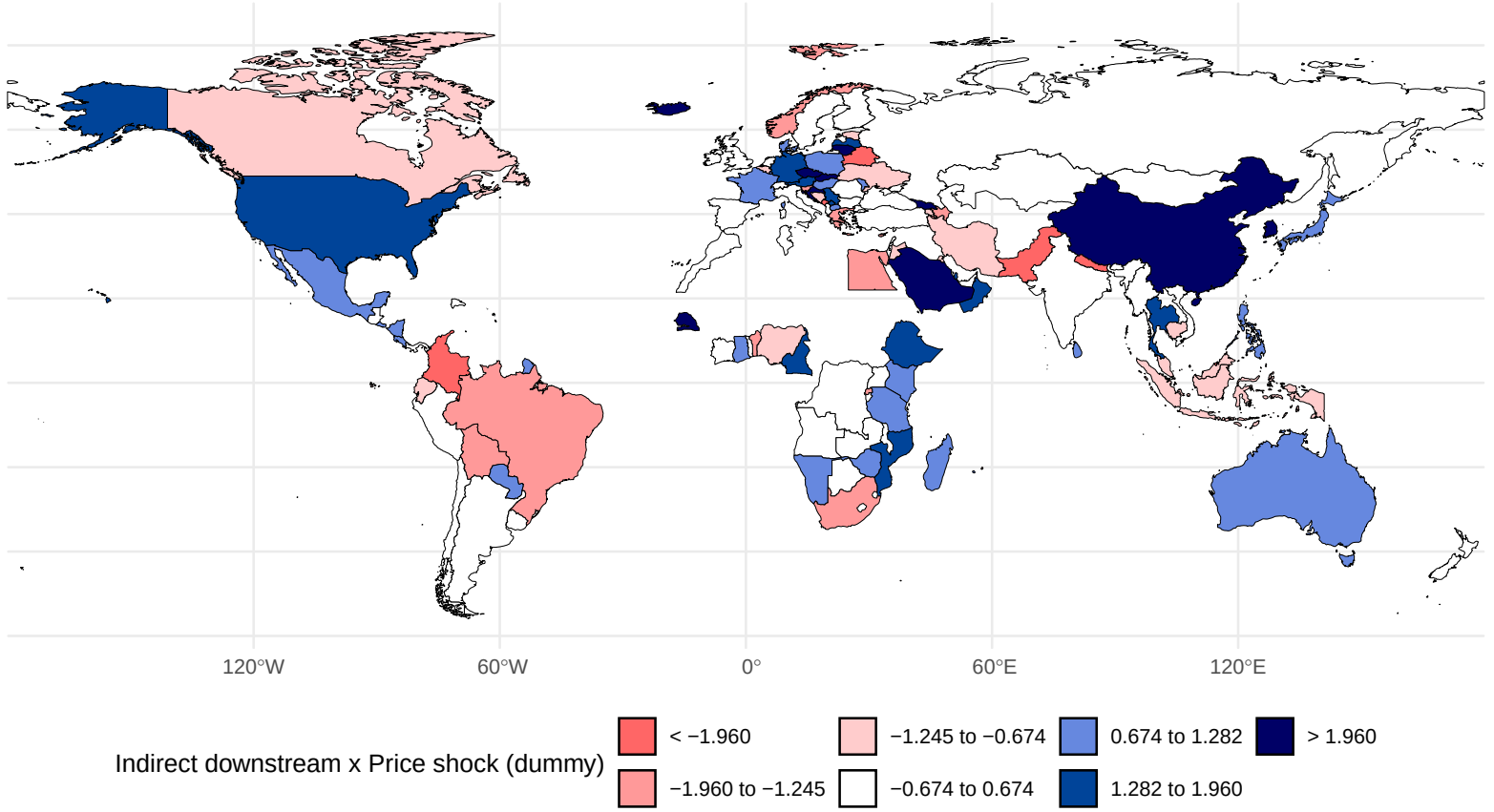
Notes: Figure presents results from a regression analysis documenting a trend across the globe suggesting that countries, on average, notably increased their imports of goods that have many AIPNET I/O linkages, $\omega_{u,d} = 1$, compared to goods with fewer or no such linkages. The dataset that is used for the estimation is a fully dyadic panel with 120 (countries) x 1183 (HS4 goods) x 1183 (HS4 goods) x 11 (years) observations. The regression includes country-by-dyad and country by HS2 by year fixed effects. We note that from 2016 onwards there has been a notable change in trend suggesting increased imports of goods that have more I/O linkages. Appendix Figure A.3 presents the same specification but for the extensive margin. Standard errors are obtained from clustering at the HS2 upstream x HS2 downstream pair across all countries with 95% confidence intervals being indicated.

Figure 12: Countries exhibiting post 2016 differential increase in goods with many AIPNET linkages also exhibit notable onshoring response following downstream supply shocks



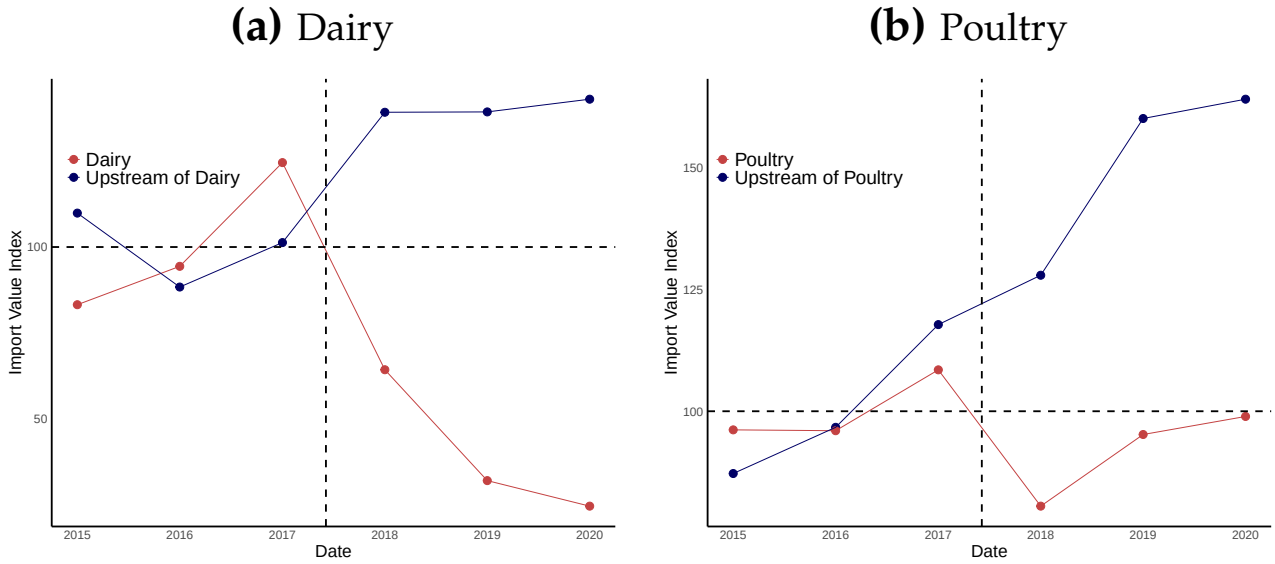
Note: Figure plots a distribution of 120 country-specific regression coefficients that attempt to proxy for onshoring. The horizontal axis captures the differential increase in imports of upstream goods that have many downstream uses or applications since 2016. The vertical axis displays the estimated increase in imports of upstream goods following a shock to import prices of downstream goods. We note that empirically, the distribution of point estimates has asymmetric support on the x-axis, this is picking up the global trend that was detected in Figure 11. We note excess mass in the top right quadrant and take particular note of Qatar. Standard errors are clustered at the upstream HS2 x downstream HS2 digit pair. Country points are labeled based on their statistical significance.

Figure 13: Estimated supply-shock induced onshoring response across countries



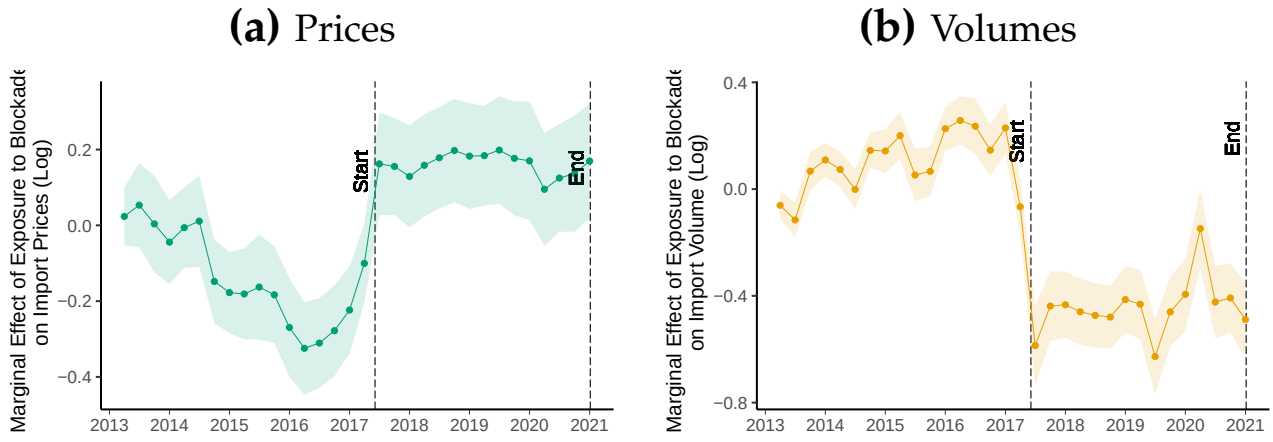
Note: Map displays the t-statistic that corresponds to the point estimates plotted on the vertical axis of Figure 13 capturing the extent to which a countries imports of upstream goods are increasing following a downstream supply shock, indicative of on onshoring response. Clusters of onshoring are found in Central Europe, China, the Middle East, East Africa and North America.

Figure 14: Examples of Onshoring in Qatar Measured through AIPNET Linkages



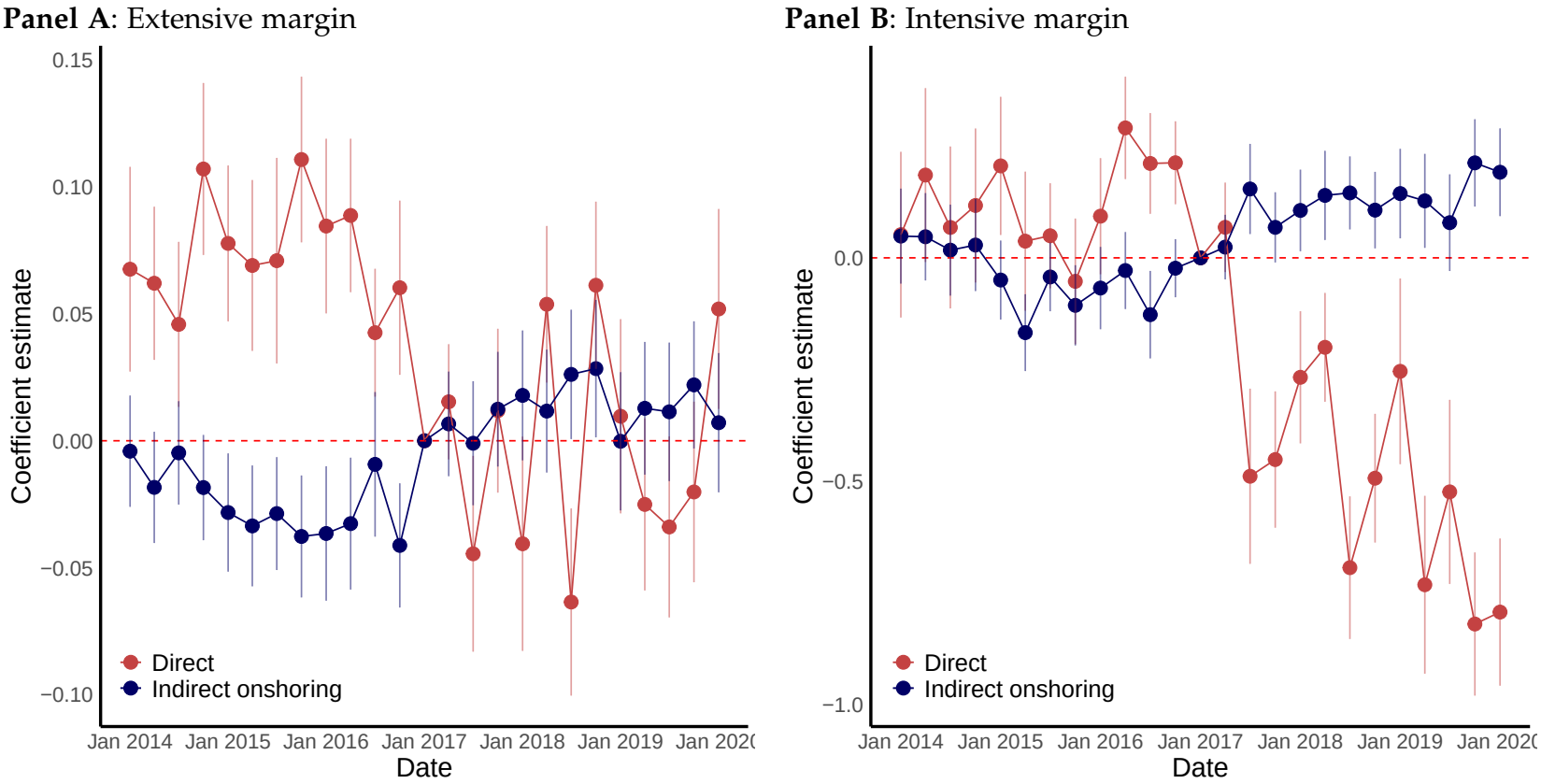
Note: Product-specific patterns of onshoring in response to the blockade for dairy (a) and poultry sector (a) in raw data studying import volumes of good as well as goods that are supply chain linked. We index 4-digit products' annual imports w.r.t. the 2016 import level and plot the weighted average for a given group of products using the import values in 2016 as weights. We define upstream goods based on AIPNET.

Figure 15: Qatar Blockade Caused Sudden Shift in Import Prices and Volumes



Note: Regression evidence of the Qatar blockade on import goods' prices (a) and import volumes (b). Using monthly 8-digit product-level imports to Qatar, we obtain the year-specific effect of a one-unit blockade exposure increase on a good's logarithmic import price and volume. We control for product- and year-month-fixed effects. The shaded areas represent the 95% confidence intervals.

Figure 16: Goods with higher supply shock saw a stronger direct import contraction and notably stronger onshoring response



Note: Figure presents dynamic version of the estimates of column (3) and (6) from Table 4. The underlying data is the dyadic dataset capturing imports of goods that have an AIPNET input/output relationship. The underlying data resolution is at the HS4 level and at a quarterly temporal resolution.

Table 1: Example of Two Prompts used in the 'Build' Stage

Section	Prompt A: Measuring Outputs	Prompt B: Measuring Inputs
System Guidance	You are an assistant with expertise in supply chains and production processes. When given a product description, identify all products (narrowly defined physical products) which this **[focal product is used to produce i.e. list output goods for which this focal product is the input]**. Prioritize outputs for which the focal input good is an essential input. Consider both **['capital', 'intermediary', and 'final' outputs]**. When responding, you will follow the JSON schema provided in the response format. For the subfield 'product_description', you should use similar language to the detailed Harmonized System (HS) of traded goods products. Some examples of these product descriptions include:	You are an assistant with expertise in supply chains and production processes. When given a product description, identify all products (narrowly defined physical products) which this **[focal product requires as an input i.e. list input goods which are used to produce this focal product]**. Prioritize inputs which are essential to producing the focal good. Consider both **['raw materials', 'intermediary', and 'capital' inputs]**. When responding, you will follow the JSON schema provided in the response format. For the subfield 'product_description', you should use similar language to the detailed Harmonized System (HS) of traded goods products. Some examples of these product descriptions include:
Examples Drawn from HS Product Codes	• Portland cement, aluminous cement (ciment fondu), slag cement, supersulphate cement and similar hydraulic cements, whether or not coloured or in the form of clinkers • Fruit, edible; apples, fresh • Engines; pneumatic power engines and motors, other than linear acting (cylinders) • Milking machines and dairy machinery ...	• Portland cement, aluminous cement (ciment fondu), slag cement, supersulphate cement and similar hydraulic cements, whether or not coloured or in the form of clinkers • Fruit, edible; apples, fresh • Engines; pneumatic power engines and motors, other than linear acting (cylinders) • Milking machines and dairy machinery ...
Structuring Output	Only include products that are physical in nature and traded internationally between countries. List up to 20 output products, focusing on the goods which **[require the direct contribution of the focal input product]**. You will list your 20 responses in a JSON format provided in the response format. This format includes a description of the **[output product]**, an importance score (between 1 and 10), and a classification of the **output** product as either 'intermediary' or 'capital'.	Only include products that are physical in nature and traded internationally between countries. List up to 20 input products, focusing on the goods which are **[directly required to produce the focal product]**. You will list your 20 responses in a JSON format provided in the response format. This format includes a description of the **[input product]**, an importance score (between 1 and 10), and a classification of the **input** product as either 'raw material', 'intermediary', or 'capital'.
Focal Product	**[PRODUCT DESCRIPTIONS]**	**[PRODUCT DESCRIPTIONS]**

Table 2: Network Properties for First Iteration at HS6 and HS4 Levels

	(1)	(2)
Property	**HS6**	**HS4**
Number of Nodes	5,633	1,190
Number of Edges	980,018	48,212
Average Degree	347.96	81.03
Network Density	0.03	0.03
Global Clustering Coefficient	0.37	0.32
Assortativity	0.14	0.06
Number of Connected Components	1	1
Size of Largest Component	5,633	1,190

Note: This table presents key structural characteristics of the AI-generated Production Network (AIPNET) at both the HS6 and HS4 levels of product aggregation. Column (1) displays metrics for the HS6 level (6-digit codes), and column (2) for the HS4 level (4-digit codes). The properties include the number of nodes (unique products), edges (potential input-output relationships), average degree (typical connectivity), network density (interconnectedness), global clustering coefficient (degree of clustering), assortativity (tendency for similar nodes to connect), and connectivity measures like the number and size of connected components. The higher number of nodes and edges at the HS6 level reflects a more detailed network structure. The moderate clustering coefficients and positive assortativity values indicate the presence of product clusters and a mild preference for nodes to connect with similar nodes. Both networks consist of a single connected component. These metrics offer insights into how AIPNET models production relationships at different levels of aggregation.

Table 3: Impact of persistent supply shocks affecting downstream goods prices on imports of upstream goods

	(1)	(2)	(3)	(4)
Dependent variable	log(import)		Import > 0	
Post 2016 $\times$ AIPNET linkage	0.0092***		0.0013***	
	(0.0027)		(0.0003)	
AIPNET linkage $\times$ Indirect upstream price shock (dummy)		0.0014**		0.0003***
		(0.0007)		(6.32×10^{-5})
AIPNET linkage $\times$ Direct own price shock (dummy)		-0.0348***		
		(0.0081)		
R^2	0.94329	0.94421	0.74324	0.74622
No. of dyad FE	171,179,618	167,866,385	174,364,440	171,163,960
No. of time FE	121,433	121,433	121,440	121,440
Observations	1,818,906,530	1,727,632,808	1,916,673,000	1,818,407,660
Regression specification:				
Country x Dyad	X	X	X	X
Country x HS2 x Time	X	X	X	X

Notes: Table presents regression results documenting the impact of persistent price shocks on downstream goods on upstream imports. Standard errors provided in parentheses are clustered at the HS2 by HS2 pair level with stars indicating *** p <0.01, ** p <0.05, and * p <0.1.

Table 4: Impact of product-level blockade exposure on imports along AIP-NET linked supply chains

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	upstream good import > 0			log(upstream good import)		
Panel A: HS4 x quarterly panel						
Blockade $\times$ Direct Exposure	-0.0685***		-0.0755***	-0.5595***		-0.5809***
	(0.0120)		(0.0142)	(0.0718)		(0.0718)
Blockade $\times$ Downstream Exposure		0.0267***	0.0286***		0.0881**	0.1616***
		(0.0084)	(0.0075)		(0.0409)	(0.0430)
R^2	0.66689	0.66701	0.66743	0.86734	0.86695	0.86815
Observations	991,150	976,650	976,650	787,950	776,058	776,058
Panel B: HS6 x quarterly panel						
Blockade $\times$ Direct Exposure	-0.0297**		-0.0379***	-0.4081***		-0.4154***
	(0.0120)		(0.0126)	(0.0521)		(0.0531)
Blockade $\times$ Downstream Exposure		0.0138*	0.0082		0.0204	0.0605
		(0.0073)	(0.0106)		(0.0370)	(0.0380)
R^2	0.56880	0.56778	0.55651	0.78021	0.77878	0.77958
Observations	9,640,300	9,405,825	9,405,825	5,833,222	5,699,916	5,699,916
Dyad FE	X	X	X	X	X	X
Time FE	X	X	X	X	X	X

Notes: **TO FIX** This table reports results from running the dyadic regression defined in Section C.4 Equation (8). We test for the effect of 'Blockade Exposure' on Import patterns of up-stream HS8 products. 'Direct Exposure' measures the focal upstream goods proportion of total value of trade that came from blockading countries in 2016. 'Down-stream Blockade Exposure' measures the exposure of the down-stream product. We interact exposure with an indicator variable denoting the start of the blockade. We run these specifications for up-stream goods which fall into Sections 1, 10, 15, 16, and 19 of HS product nomenclature, which represents the majority of raw material inputs and capital equipment. Panel (a) contains fixed effects for HS8 product codes and dummy variables for each month, to control for seasonality. Panel (b) adds an additional fixed effect for time (year-month dummies), which are interacted with HS6 Product Codes. We cluster standard errors to account for correlation on both sides of the dyadic pairs. Dyadic-clustered standard-errors in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1.

A Appendix Figures and Tables

Table A.1: Product Examples in the Harmonized System (HS)

Hierarchy	HS Code	Product Description
HS2	01	Live Animals
HS4	0101	Horses, asses, mules, and hinnies; live
HS6	0101.21	Horses; live, pure-bred breeding animals
HS2	02	Meat and edible meat offal
HS4	0203	Meat of swine; fresh, chilled, or frozen
HS6	0203.12	Meat; of swine, hams, shoulders, and cuts thereof, with bone in, fresh or chilled
HS2	03	Fish and crustaceans, molluscs, and other aquatic invertebrates
HS4	0302	Fish; fresh or chilled, excluding fillets
HS6	0302.13	Fish; fresh or chilled, Pacific salmon
HS2	84	Machinery and mechanical appliances; parts thereof
HS4	8402	Steam or other vapor generating boilers
HS6	8402.11	Boilers; steam or other vapor generating, water-tube boilers with a steam production exceeding 45 tons per hour

Note: This table presents examples of product nodes within the Harmonized System (HS) classification at HS2, HS4, and HS6 hierarchies. Each entry demonstrates the progression from broad categories (HS2) to more specific classifications (HS6), highlighting the diversity of goods from simple (e.g., live animals) to complex (e.g., machinery).

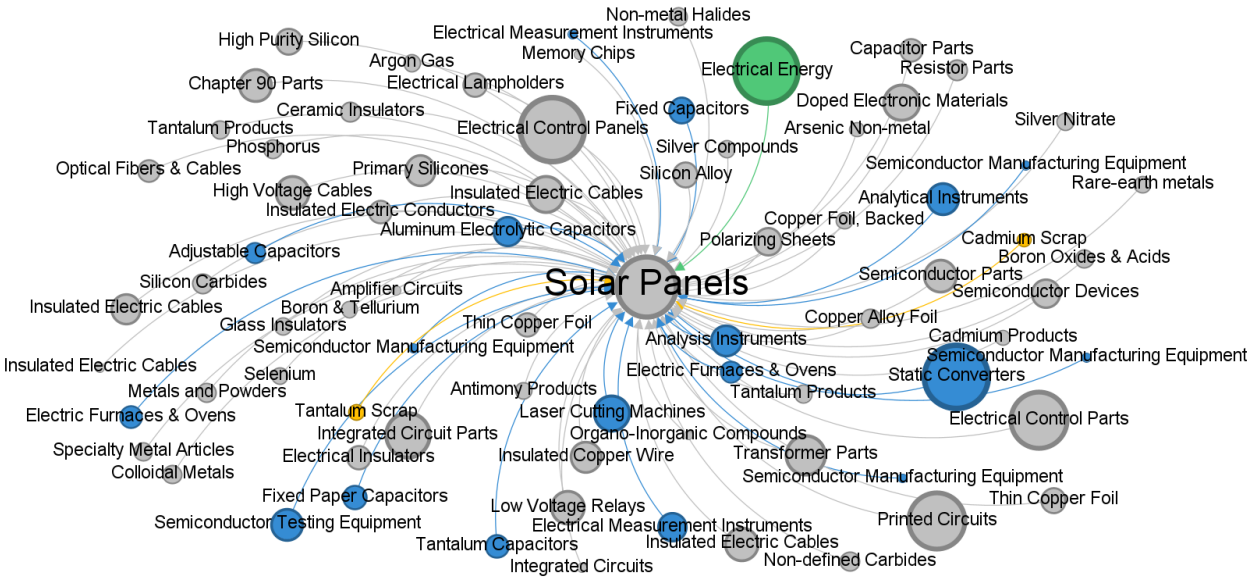
Table A.2: Top 30 Imports by volume across HS2 sections affected by the blockade

HS2	Description	Imports in 2016 from		E_i	Import share
		All countries	blockading countries		
73	Iron or steel articles	1,443.43	412.04	28.5%	4.9%
85	Electrical machinery and equipment and p	3,006.73	327.79	10.9%	10.3%
74	Copper and articles thereof	391.26	289.88	74.1%	1.3%
39	Plastics and articles thereof	718.35	261.21	36.4%	2.5%
72	Iron and steel	509.40	258.95	50.8%	1.7%
4	Dairy produce; birds' eggs; natural hone	393.48	245.25	62.3%	1.3%
84	Nuclear reactors, boilers, machinery and	4,984.55	141.15	2.8%	17.0%
27	Mineral fuels, mineral oils and products	181.39	137.96	76.1%	0.6%
68	Stone, plaster, cement, asbestos, mica o	362.90	124.15	34.2%	1.2%
76	Aluminium and articles thereof	225.93	117.44	52.0%	0.8%
32	Tanning or dyeing extracts; tannins and	172.86	104.11	60.2%	0.6%
7	Vegetables and certain roots and tubers;	268.68	97.61	36.3%	0.9%
71	Natural, cultured pearls; precious, semi	603.02	94.68	15.7%	2.1%
34	Soap, organic surface-active agents; was	165.11	94.06	57.0%	0.6%
94	Furniture; bedding, mattresses, mattress	835.00	93.74	11.2%	2.9%
87	Vehicles; other than railway or tramway	3,690.80	89.24	2.4%	12.6%
22	Beverages, spirits and vinegar	164.63	86.89	52.8%	0.6%
33	Essential oils and resinoids; perfumery,	389.23	85.57	22.0%	1.3%
15	Animal or vegetable fats and oils and th	104.87	76.48	72.9%	0.4%
38	Chemical products n.e.s.	378.74	74.59	19.7%	1.3%
19	Preparations of cereals, flour, starch o	205.14	72.86	35.5%	0.7%
48	Paper and paperboard; articles of paper	178.95	67.70	37.8%	0.6%
25	Salt; sulphur; earths, stone; plastering	107.24	64.42	60.1%	0.4%
2	Meat and edible meat offal	430.22	58.79	13.7%	1.5%
30	Pharmaceutical products	527.57	56.63	10.7%	1.8%
69	Ceramic products	213.33	49.67	23.3%	0.7%
21	Miscellaneous edible preparations	169.32	48.96	28.9%	0.6%
8	Fruit and nuts, edible; peel of citrus f	229.49	44.44	19.4%	0.8%
70	Glass and glassware	166.14	44.38	26.7%	0.6%
1	Animals; live	215.35	43.84	20.4%	0.7%

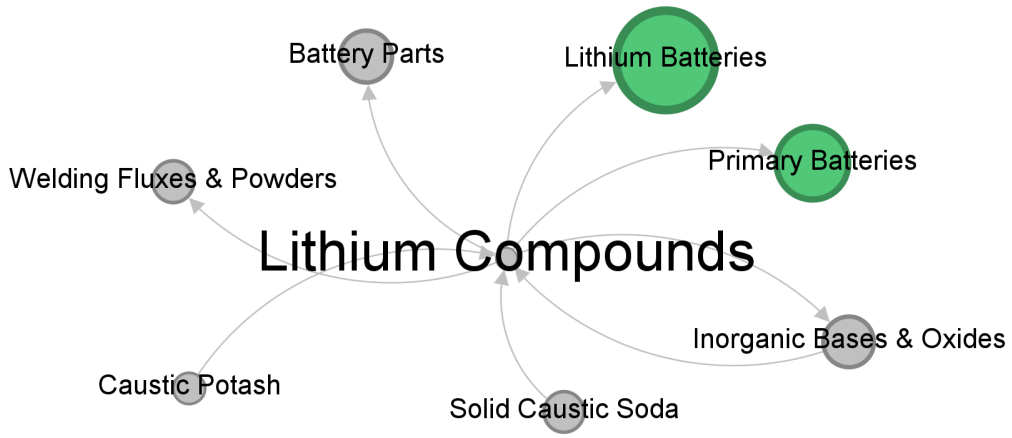
Notes: Table presents the top 30 imports by volume that have been affected by the blockade of Qatar. Trade volumes ($ million) are measured in 2016, the year before the blockade was announced. HS2 section descriptions are truncated to 40 characters in length.

Figure A.1: Illustrative examples: Solar Panels and Lithium Compounds

(a) Solar Panels (HS: 8541.40)

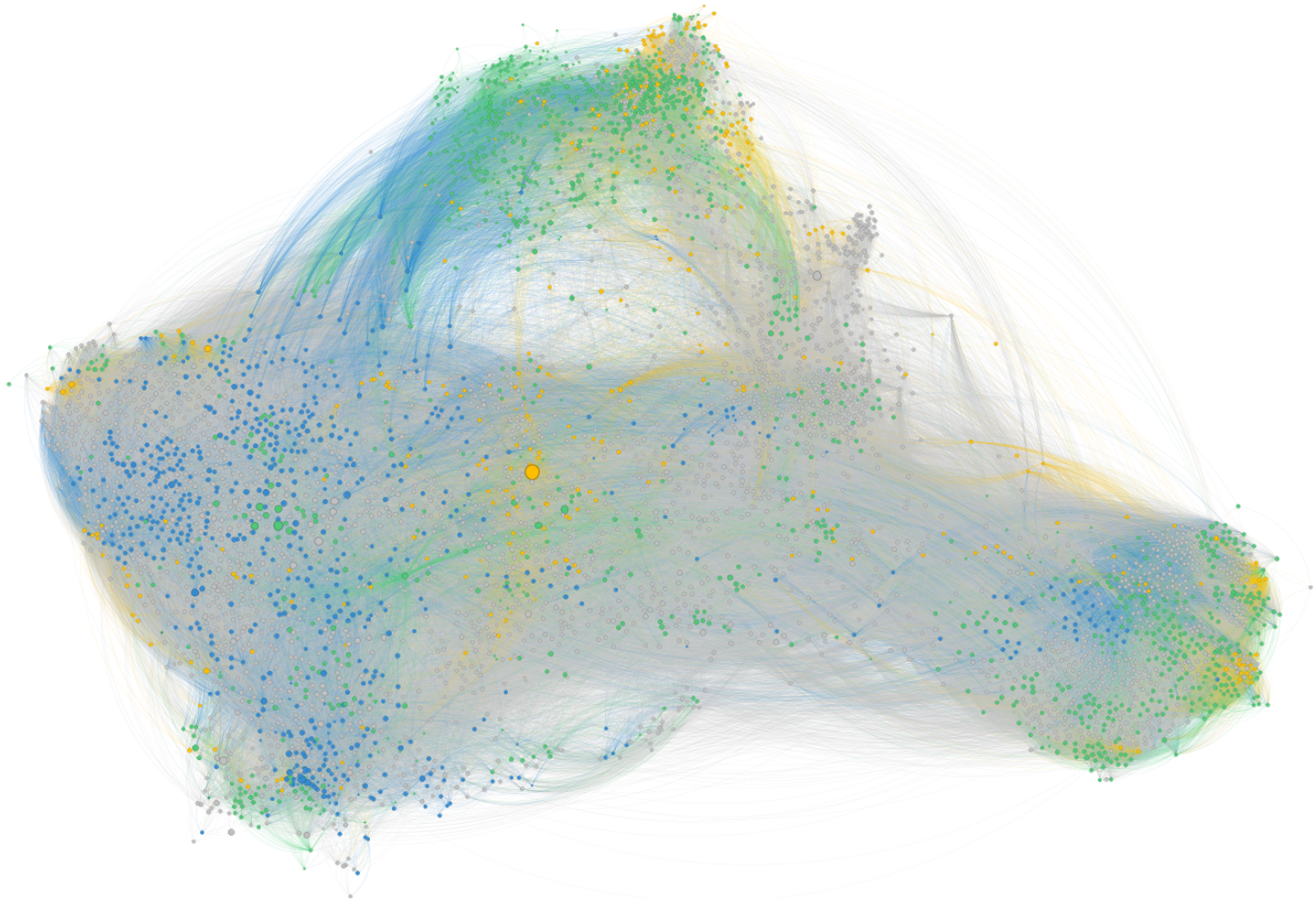


(b) Lithium Compounds (HS: 2825.20)



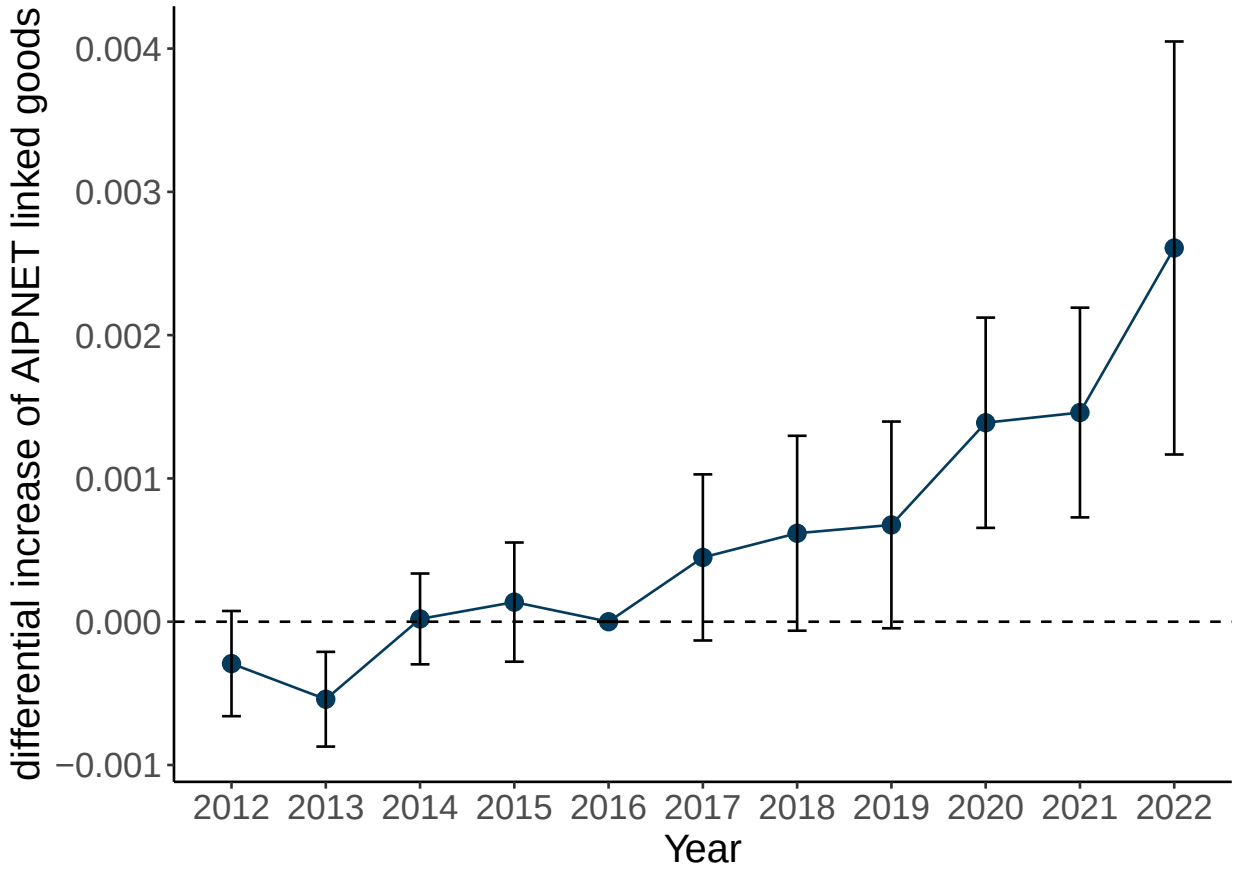
Note: Each panel presents a **one-degree ego network** of the focal product in AIPNET. Each node represents a single six-digit HS product code. Code descriptions have been abbreviated for brevity. The full set of HS Code linkages can be found in attached dataset and on our website https://aipnet.io/. Nodes are colored by their predominant linkage type according to the BEC5 end-use classification: green for consumption goods, blue for capital goods, yellow for raw materials (intermediate-primary goods), and grey for intermediate-processed goods. The focal node in each sub-figure has larger font size. Edges are colored by the source node.

Figure A.2: AI-generated Production Network (AIPNET) visualization



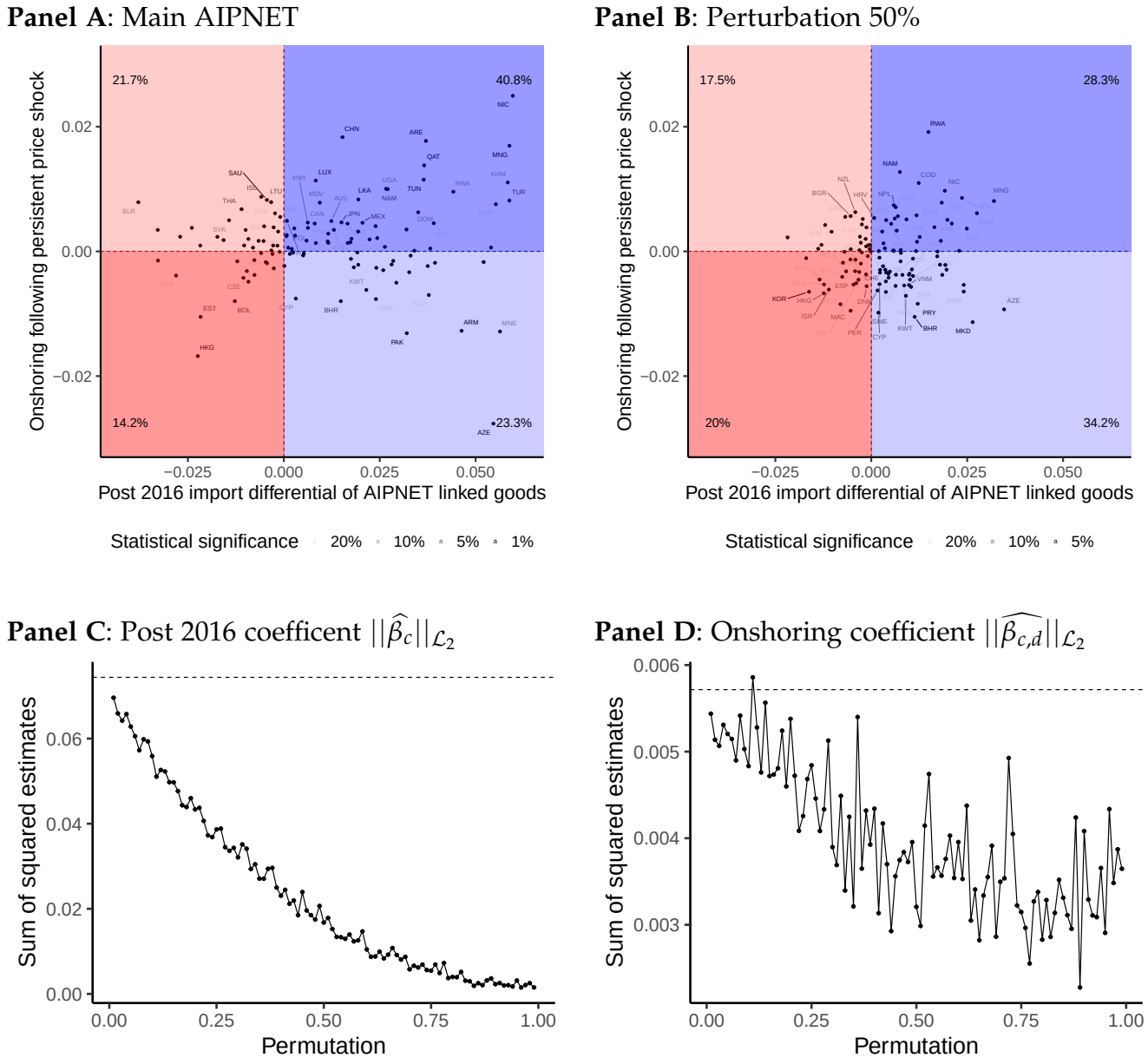
Note: This figure presents a visualization of the production network, where each node represents a Harmonized System (HS) 6-digit product code, and edges represent vertical production relationships. The layout, derived using the force-atlas 2 algorithm, organizes nodes into clusters based on the Broad Economic Categories (BEC) end-use classification. Nodes are color-coded: green represents consumption goods, blue represents capital goods, yellow represents raw materials (intermediate-primary goods), and grey represents intermediate-processed goods. The network reveals distinct clusters: raw materials are dispersed across the network, with a clear cluster of final goods (green) on the top and to the right. The central left cluster is composed of capital goods (blue) and intermediary goods (grey). This clustering highlights the central roles of different sectors within the global production network, emphasizing the interconnectedness of various product groups.

Figure A.3: Notable differential increase in imports of goods with many AIPNET I/O linkages since 2016



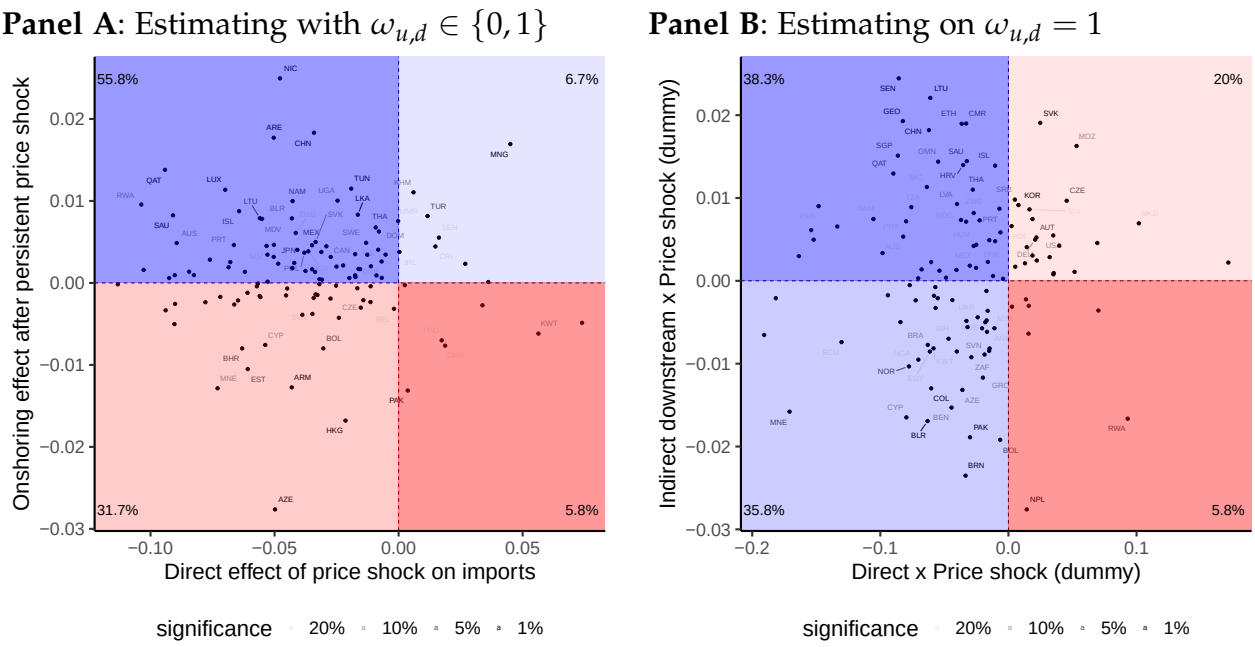
Notes: Figure presents results from a regression analysis documenting a trend across the globe suggesting that countries, on average, notably increased their imports of goods that have many AIPNET I/O linkages, $W_{ij} = 1$, compared to goods with fewer or no such linkages. The dataset that is used for the estimation is a fully dyadic panel with 120 (countries) x 1183 (HS4 goods) x 1183 (HS4 goods) x 12 (years) observations. The regression includes country-by-dyad and country by HS2 by year fixed effects. We note that from 2016 onwards there has been a notable change in trend suggesting increased imports of goods that have more I/O linkages. Figure 11 presents the same specification but for the intensive margin.

Figure A.4: Evolution of cross-country coefficients across perturbed production networks



Note: Figure showcases how the post 2016 and the onshoring coefficient distribution across countries is shrinking towards zero as we increase the amount of noise that is infused into AIPNET through varying the perturbation parameter. Panel A presents the main network results for reference, Panel B showcases what the distribution looks like with the perturbation parameter to be set to 50%. Panel C plot the evolution of the average squared length of the estimated coefficient $\widehat{\beta}_c$ as a function of the permutation paramter with permutations. The dashed line represents the coefficient for the main network. Panel D presents the evolution of the average squared length of onshoring coefficient as a function of the perturbation parameter. We note the convergence to the origin highlighting that as the production network becomes full noise we fail to detect an onshoring effect.

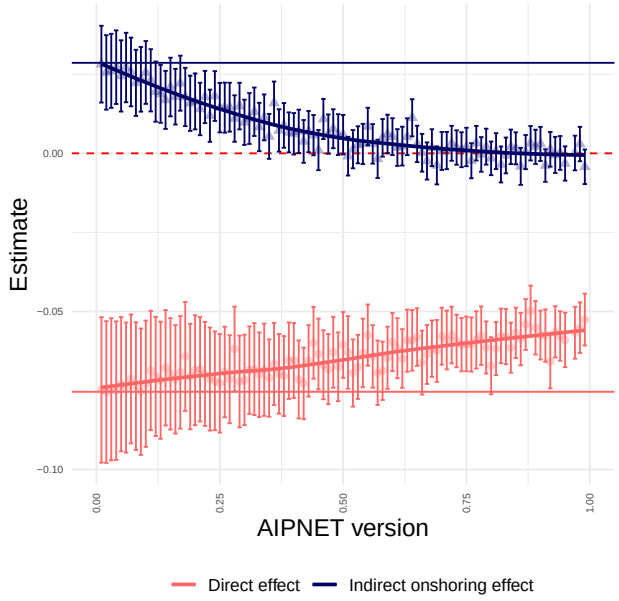
Figure A.5: Estimating of onshoring regression focusing on $\omega_{u,d} = 1$ network



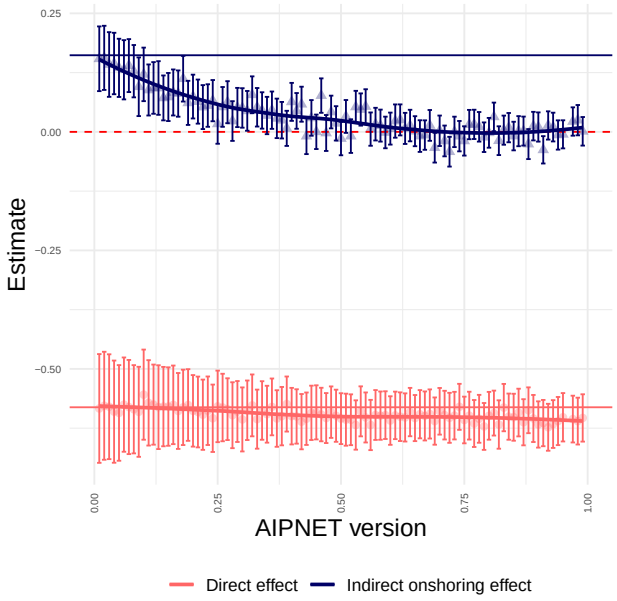
Note: Figure showcases how the post 2016 and the onshoring coefficient distribution across countries is shrinking towards zero as we increase the amount of noise that is infused into AIPNET through varying the perturbation parameter. Panel A presents the main network results for reference, Panel B showcases what the distribution looks like with the perturbation parameter to be set to 50%. Panel C plot the evolution of the average squared length of the estimated coefficient $\widehat{\beta}_c$ as a function of the permutation paramter with permutations. The dashed line represents the coefficient for the main network. Panel D presents the evolution of the average squared length of onshoring coefficient as a function of the perturbation parameter. We note the convergence to the origin highlighting that as the production network becomes full noise we fail to detect an onshoring effect.

Figure A.6: Evolution of Qatar natural experiment estimates across perturbed production networks

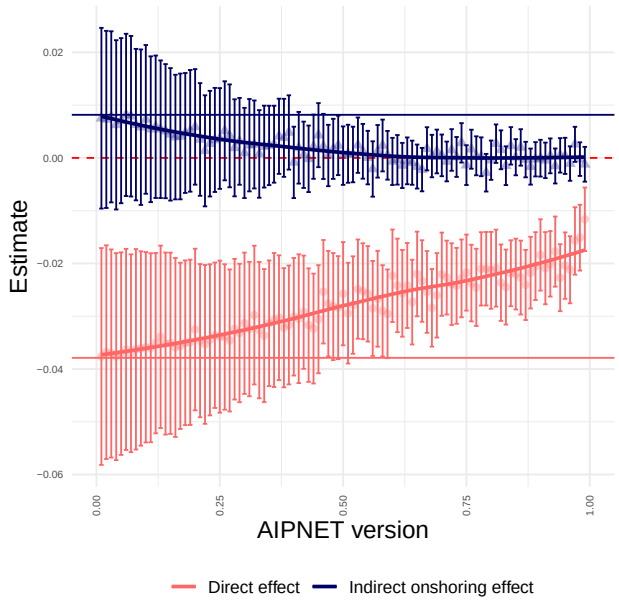
Panel A: HS4 Extensive margin



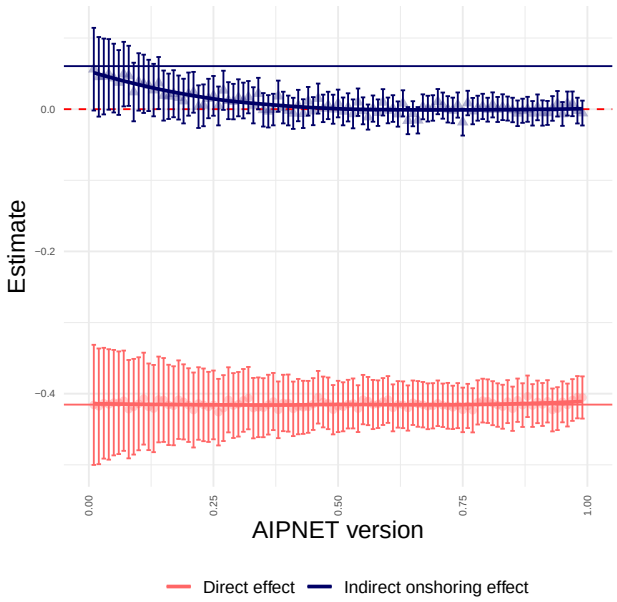
Panel B: HS4 intensive margin



Panel A: HS6 Extensive margin



Panel B: HS6 intensive margin



Note: Figure visually displays estimated coefficients from estimating the Qatar case study with different AIPNET production networks that become increasingly more noisy. The onshoring coefficient capturing imports of upstream goods following downstream blockade-induced supply disruptions is plotted in blue, while the direct effect of upstream price shocks on imports of upstream goods is illustrated in red. The solid horizontal lines provide the point estimates from the main AIPNET production network that can be found in Table 4.

B Detailed Description of AI-generated Production Network (AIPNET)

In this section, we provide full details on the construction of the AI-generated Production Network (AIPNET). These details offer a blueprint for utilizing AI/LLM tools in the construction of novel graphical datasets, which can be broadly applied in various fields. Potential applications include linking occupational classifications, structuring scientific and patent classes, and connecting financial reports or earnings calls from public firms, and so on.

Our approach to building graphical data using AI follows four key steps, which are each documented in a subsection of this appendix. Subsection B.2 provides some background on LLMs and discusses how we directly interact with our chosen model, including our prompting strategy. Subsection B.3 covers our approach to recasting the LLM's responses into a codified nomenclature to recover a graphical network structure. Subsection B.5 discusses our bootstrapping methodology to address potential sensitivity of results to unusual LLM behavior.

B.1 Background on LLMs

LLMs establish linkages beyond traditional Natural Language Processing (NLP) methods, which typically depend on keyword overlap. Pre-trained on extensive datasets, including detailed descriptions of production processes and product characteristics from sources like Wikipedia, LLMs bring a nuanced understanding of language that enables them to apply human-level judgment at scale. For instance, the Wikipedia entries for 'Electrical energy' and 'Dairy' are just two examples of the rich content informing LLM training, enhancing their ability to discern vertical relationships between products.[43]

[43]One important source for training LLMs is Wikipedia. This source alone provides highly informative content on the production process of many goods, for example, the entries for 'Electrical energy' and 'Dairy'. Countless sources such as these enter the LLMs training data. The net result is an ability to apply human-level judgment, at scale, when determining vertical relationships between products.

Deploying LLMs has become cost-effective and scalable. Their use results in high accuracy.

B.2 Interacting with the LLM

The process of constructing AIPNET began by interacting with the LLM using specific zero-shot prompts tailored to extract detailed input-output relationships for each product within the Harmonized System (HS) codes. Our approach directly queries the LLM to generate free-text responses, which are then systematically matched to HS codes based on the cosine similarity between text embeddings of the free-text descriptions and embeddings of HS code descriptions.

This sequence is visually summarized in Figure 4, which provides a flowchart of the entire process. The remainder of this section will describe this process, along with properties of the retrieved network and comparisons with existing network-inspired measures and correlation with input-output tables.

To facilitate the extraction of input-output linkages, we crafted two distinct types of prompts: one for identifying inputs required to produce a focal product and another for identifying outputs that the focal product is used to produce. The system instructions are designed to guide the LLM in generating structured, relevant responses.

The key steps in interacting with the LLM are as follows:

1. **Prompt Setup:** We prepared two types of prompts—Focal Finished and Focal Unfinished—to query the LLM. The prompts were designed to extract detailed information about inputs needed to produce a finished product and outputs that can be produced using an unfinished product as an input for each HS6 product. We provided the LLM with examples of similar product descriptions to guide the responses.

2. **Response Generation:** For each HS6 product code, the LLM generated a list of related products along with an importance score (1 to 10), a free-text description,

and a classification of the linkage type (intermediary, capital, or final). The LLM was prompted to list up to 20 goods, focusing on those with higher importance scores and ensuring that only internationally traded products were included.

3. **Iteration Process:** The entire process was repeated 10 times for each focal good to create 10 iterations each, capturing a range of possible relationships and ensuring robustness against potential LLM-induced variability.

B.2.1 Full System Instructions

In the design of the system instructions for the assistant, our primary goal was to ensure clarity, consistency, and depth in the responses pertaining to various products. Below are the full system instructions used to guide the LLM in generating the required input-output relationships:

For Focal Finished Products

> You are an assistant with expertise in supply chains and production processes.
>
> When given a product description, identify all products (narrowly defined physical products) which are used to produce this focal product i.e. list input goods for which this focal product is the output.
>
> Prioritize inputs that are well recognized as essential for the production of the focal output good.
>
> Consider both 'capital' and 'intermediary' inputs. When responding, you will follow the JSON schema provided in the response format.
>
> For the subfield 'product_description', you should use similar language to the detailed Harmonized System (HS) of traded goods products.
>
> Some examples of these product descriptions include: *insert example goods*

Only include products that are physical in nature and traded internationally between countries.

List up to 20 input products, focusing on the goods which most directly contribute to the production of the focal output product.

You will list your 20 responses in a JSON format provided in the response format, which includes a description of the input product, an importance score (between 1 and 10), and a classification of the input product as either 'intermediary' or 'capital'.

For Focal Unfinished Products

You are an assistant with expertise in supply chains and production processes.

When given a product description, identify all products (narrowly defined physical products) which this focal product is used to produce i.e. list output goods for which this focal product is the input.

Prioritize outputs for which the focal input good is an essential input.

Consider both 'capital', 'intermediary', and 'final' outputs. When responding, you will follow the JSON schema provided in the response format.

For the subfield 'product_description', you should use similar language to the detailed Harmonized System (HS) of traded goods products.

Some examples of these product descriptions include: *insert example goods*

Only include products that are physical in nature and traded internationally between countries.

List up to 20 output products, focusing on the goods which require the direct contribution of the focal input product.

You will list your 20 responses in a JSON format provided in the response format.

This format includes a description of the output product, an importance score (between 1 and 10), and a classification of the output product as either 'intermediary' or 'capital'.

We give the following examples, as a JSON list, in the area marked as "*insert example goods*":

- Pencils and crayons; with leads encased in rigid sheath
- Photographic plates and film; for colour photography (polychrome), in the flat, sensitised, unexposed, with no side exceeding 255mm, of any material other than paper, paperboard or textiles
- Oil-cake and other solid residues; whether or not ground or in the form of pellets, resulting from the extraction of linseed oils
- Portland cement, aluminous cement (ciment fondu), slag cement, supersulphate cement and similar hydraulic cements, whether or not coloured or in the form of clinkers
- Fruit, edible; apples, fresh
- Engines; pneumatic power engines and motors, other than linear acting (cylinders)
- Milking machines and dairy machinery
- Glass; cast glass and rolled glass, non-wired sheets, coloured through the mass (body tinted), opacified, flashed or having an absorbent, reflecting or non-reflecting layer
- Sanitary towels (pads) and tampons, napkins and napkin liners for babies and similar articles, of any material
- Mathematical equipment; micrometers, calipers and gauges

- Styrene polymers; waste, parings and scrap
- Machines; for additive manufacturing, with digital input
- Lathes; for removing metal, horizontal, numerically controlled
- Industrial robots; for multiple uses, capable of carrying a weight exceeding 10 kg

Here are some reasons why we included these specific directives in the system instructions:

- **"Identify all products (narrowly defined physical products)..."**: This clarifies the scope of the task, ensuring the LLM focuses on specific, identifiable products rather than broad categories.
- **"Prioritize inputs/outputs that are well recognized as essential..."**: This encourages the LLM to focus on the most crucial linkages, improving the relevance and accuracy of the response.
- **"Consider all types of linkages: capital goods, intermediary goods, and final goods..."**: This ensures that the LLM considers the full spectrum of potential connections, capturing the complexity of production processes.
- **"Your response should be in JSON format..."**: This requirement instructs model to comply with a structured response format, making it easier to parse and integrate into the subsequent steps of network construction.
- **"Assign an importance score from 1 to 10..."**: The importance score encourages the LLM to think critically about the significance of each product, which helps in prioritizing the most relevant linkages.
- **"List up to 20 goods..."**: This sets a clear limit to avoid overwhelming the process with too many products while focusing on the most relevant ones.

- **"Only include products that are traded internationally..."**: This ensures the relevance of the output to global trade, which aligns with the overall objective of constructing a network of internationally traded goods.

- ***"example goods"***: We included a diverse range of example product descriptions to guide the LLM in understanding different types of goods to ensure that the LLM can handle various product categories effectively, improving the robustness and flexibility of the output.

B.3 Matching LLM Output to HS Product Codes

After the LLM generated free-text descriptions of related products, the next critical step was to accurately map these descriptions to the corresponding HS codes. Given the potential for inconsistencies in nomenclature and the challenge of directly asking the LLM for HS codes, we employed a text embedding approach to ensure precise matching.

Allowing the LLM to generate free-text responses enables it to establish semantic connections between products within the predefined product space. This approach prevents the model from being constrained by predefined categories, which might miss out on less obvious but still relevant linkages. By focusing on free-text, we retain control over the final structure of the network through post-processing techniques like embedding similarity thresholds and edge overlap parameters, rather than relying solely on the model's output. This method ensures that we can fine-tune the network's precision and recall based on the specific needs of the analysis, as described in the OpenAI post on structured outputs.[44]

Embeddings Overview Embeddings are vector representations of text that capture the semantic meaning of the text in a numerical form. OpenAI's text embeddings, such as those generated by the "text-embedding-3-large" model, are commonly used to measure the relatedness of text strings. Each text string is transformed into a high-dimensional

[44] **https://openai.com/index/introducing-structured-outputs-in-the-api/**

vector (in our case dimensions are 3072), where each dimension captures a different aspect of the text's meaning. The similarity between two pieces of text can then be measured by the distance between their respective vectors, with smaller distances indicating higher similarity.

For example, the sentences 'Copper ore is smelted into copper' and 'Copper is extracted from copper ore' would have very similar embeddings because they describe the same production process. The cosine similarity metric is used to measure this relatedness, focusing on the direction in the vector space, which captures the essence of the process, rather than the exact wording. This is similar to how the abbreviation 'YMCA' would have a high similarity with 'Young Men's Christian Association' even though they share little in terms of characters, because embeddings capture the underlying semantic meaning

These embeddings are particularly useful for tasks such as identifying similarities between products in production networks, where similar processes or inputs can be grouped together; semantic search, where relevant processes or materials are ranked based on their similarity to a query; and classification, where products are categorized based on their most similar examples in the network.

Use Cases in Literature Embeddings have been widely used in various research and applied contexts. For instance, in the field of natural language processing, they have been employed for semantic search, where documents are ranked based on their relevance to a query (Mikolov, 2013; Rong, 2014; Goldberg & Levy, 2014). In addition, embeddings have been used in recommendation systems, where products or content are recommended based on the similarity of their embeddings to a user's preferences. This method has been employed by companies like Netflix and Amazon to personalize content delivery (Covington, Adams, & Sargin, 2016). Garg and Fetzer (2024) use them to match economics concepts to JEL codes for tractability.

Implementation Details We used the "text-embedding-3-large" model from OpenAI to create embeddings for both the LLM-generated descriptions and the official HS code descriptions. Each response line (i.e., vertical linkage free-text response) was converted into an embedding, and we did the same for each HS6 code description. The matching process involved calculating the cosine similarity between the embeddings of the LLM responses and the HS code descriptions. A cosine similarity threshold of 0.75 was used to determine a match. This threshold is a common standard in text similarity tasks, as it balances precision and recall effectively, ensuring that the matches are both relevant and accurate without being overly strict. This allows us to link the LLM output with the correct HS codes, maintaining a robust yet flexible network structure.

B.4 Pruning the Network

After constructing the initial network using the methods described in previous sections, we recognized a need to improve its precision by removing false positive connections. The Large Language Model (LLM) generated a range of potential input-output relationships, but some did not reflect true production linkages. These inaccuracies mainly arose from the inherent challenges of matching LLM output to HS product codes using vector embedding similarity. While this approach maximized recall, it came at the expense of precision. To address this, we introduced a pruning stage to filter out incorrect or unrealistic edges, thereby enhancing the accuracy of the network after optimizing for recall.

The pruning process involved a systematic assessment of each proposed input-output pair in the preliminary network. We employed an additional LLM-powered verification step, where the model evaluated whether a given upstream product could reasonably be used as an input in the production of a downstream product. This approach leveraged the LLM's understanding of supply chains and production processes to validate each linkage. To facilitate this verification, we designed a specific prompt to query the LLM for each potential edge. The prompt was crafted to elicit a definitive response from the

model, ensuring clarity and consistency. The prompt used was as follows:

> Potential input: '{HS_description_input}'. Potential output: '{HS_description_output}'?*You are an assistant with expertise in supply chains and production processes. You will be given pairs of products represented by their Harmonized System (HS) code descriptions. Each description refers to a product category describing types of traded goods. Your task is to assess whether a product within the first category (the potential input) is an important input into the production of a product within the second category (the potential output). You must return a response of 1 if the input product could reasonably be used in the production of the output product, and 0 if it could not. Return only the numerical response in a strict JSON format with no other output. Be careful to avoid cases where the potential input product is a substitute, but not an input, for the potential output product. If you are unsure, then you should return 0.*
>
> *Potential input: '{HS_description_input}'. Potential output: '{HS_description_output}'?*

In this prompt, {`HS_description_input`} and {`HS_description_output`} are placeholders for the HS code descriptions of the potential input and output products, respectively.

Implementation Details For the pruning stage, we used the GPT-4o-mini model, an efficient variant within the GPT-4o family designed to handle large-scale implementations. In this stage, the model processed each edge of the initial network, where each edge represents a potential upstream-downstream relationship. Unlike the build stage, where inputs were at the node level, inputs here were at the edge level. The model's response was restricted to a simple numerical value: 1 to indicate a valid input-output relationship or 0 to indicate that the input is not reasonably used in the production of the output. Responses were formatted in JSON to facilitate automated parsing.

By enforcing a strict response format and providing clear instructions, we minimized ambiguity and ensured that the model's outputs could be directly interpreted to accept or reject each edge in the network. Several considerations underpinned our approach:

- **Avoiding False Positives:** The initial network aimed for high recall, which risked including spurious linkages. The pruning stage focused on enhancing precision by filtering out unlikely or incorrect connections.
- **Handling Ambiguity:** We instructed the model to default to a response of 0 in cases of uncertainty. This conservative approach prioritized the reliability of accepted edges over completeness, acknowledging that some legitimate relationships might be excluded to avoid false positives.
- **Distinguishing Substitutes from Inputs:** The prompt explicitly cautioned the model to avoid confusing products that are substitutes for one another with those that have a direct input-output relationship, as substitutes are not part of the production process for the output product.
- **Consistency Across Evaluations:** Using the same model and prompt for all edge evaluations ensured consistency in the pruning process. The GPT-4o model's extensive training on supply chain information made it well-suited for this task.

Results of Pruning The pruning process resulted in a refined network with improved accuracy of the input-output relationships. Edges that did not represent valid production linkages were removed, enhancing the network's utility for subsequent analyses. The final AIPNET retained its cohesive structure, with each connection more reliably reflecting actual production dependencies among internationally traded goods. This pruning stage was essential to ensure that the AIPNET accurately models the true production relationships, thereby increasing the validity of any downstream applications or analyses utilizing this network.

B.5 Bootstrapping Approach to Building Graphical Data using LLMs

To enhance the robustness of AIPNET and to account for variability in LLM outputs, we employed a bootstrapping approach that involved generating 10 iterations of the network. Each iteration represented a separate draw of the network based on the LLM's responses to our prompts. By using multiple iterations, we were able to capture a range of possible relationships and reduce the impact of any unusual or outlier responses from the LLM.

The key aspects of this bootstrapping approach are:

- **Multiple Iterations:** We repeated the LLM querying process 10 times for each focal product, creating 10 versions of the network. This allowed us to observe the stability and variability of the relationships identified by the LLM.
- **Edge Overlap as a Hyperparameter:** The edge overlap, defined as the number of iterations in which a given edge appears, was used as a hyperparameter to refine the network. By adjusting this parameter, we aimed to align the network's sparsity with that observed in the US Input-Output (IO) tables, ensuring that our network structure was both parsimonious and empirically grounded.
- **Optimal Network Structure:** We found that the optimal network structure was the core network where a given edge appeared in all 10 iterations. This consensus-based voting approach provided the most reliable representation of the production network, balancing predictive accuracy with an appropriate level of sparsity.

This bootstrapping approach, akin to statistical resampling methods, provided a more robust and reliable network by ensuring that the final structure was not overly reliant on any single set of LLM outputs.

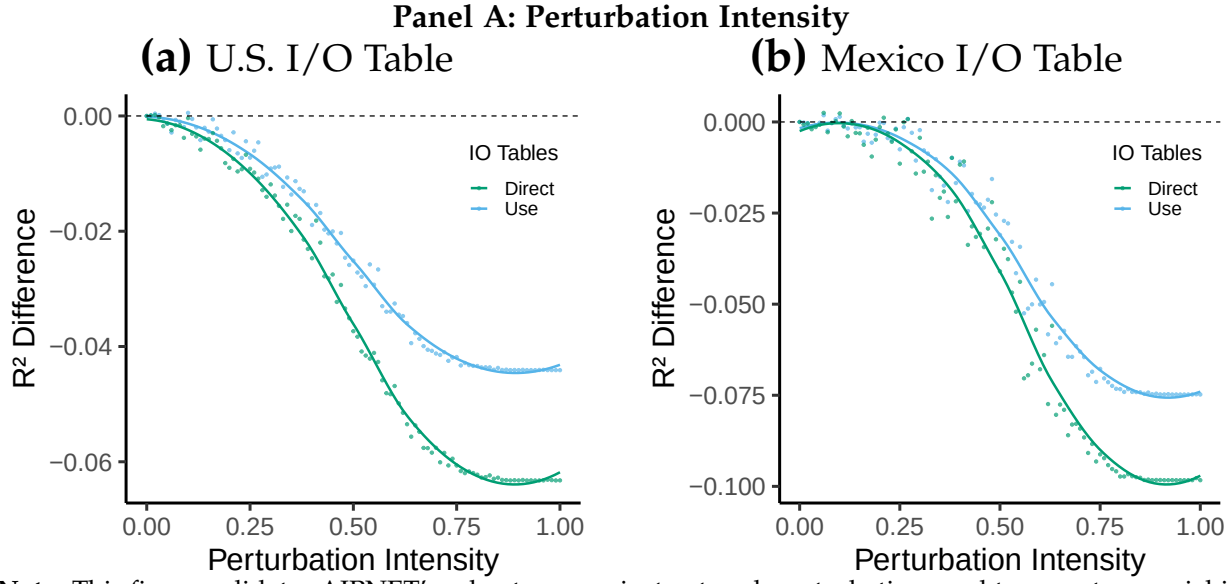
B.6 Validation of AIPNET with Perturbed Networks

Bartolucci, Caccioli, Caravelli, and Vivo (2024) find that the upstreamness and downstreamness measures as e.g. developed in Fally (2012) and Antràs et al. (2012), can exhibit a puzzling correlation that arises naturally due to structural constraints in input-output tables, even when the underlying production network is just constituted of noise. This raises concerns about whether our validation exercise leveraging input/output tables may suffer from a similar fate given that the IOT/SUT are much coarser compared to the granularity of AIPNET and as our validation implicitly involves aggregation of AIPNET to match to the coarser resolution of IOT/SUTs.

To allay that the correlation is spurious we further validated the robustness of AIPNET through a network perturbation approach that we also adopt as a robustness check to the substantive analysis. To do so, we systematically introduced varying degrees of edge rewiring of our AIPNET graph $\mathbf{G} = (\mathbf{W}, N, L)$. That is, we make $\mathbf{G}(p)$ with p capturing the fraction of edges that are being rewired from 1% to 100%, incremented by 1% at each step. The rewiring was conducted using the Erdos-Renyi random graph model, a method that introduces controlled randomness while preserving the network's structural properties (Erdos et al., 1960). Specifically, edges between nodes were rewired with a probability proportional to the perturbation intensity, resulting in a randomized network that maintains the original number of edges but with varying levels of structure degradation.

We then compared these perturbed networks against the Input-Output (I/O) tables by calculating the difference in R^2 between the original, unperturbed network (edge overlap = 10) and each perturbed version. The findings, presented in Figure B.7, demonstrate a clear trend: as the intensity of perturbation increases, the R^2 difference also increases monotonically, signifying a decrease in the network's predictive accuracy as it deviates from its original structure. This pattern holds consistently across both U.S. and Mexico I/O Tables—Direct and Use—highlighting that AIPNET's structure encapsulates meaningful economic relationships, which are progressively lost as noise is introduced.

Figure B.7: Validation of AIPNET: Impact of Network Perturbations



Note: This figure validates AIPNET's robustness against network perturbations and temperature variability, using U.S. and Mexico I/O Tables. Panel A examines perturbation intensity, with subfigures (a) and (b) showing the R^2 difference relative to the original, unperturbed network (edge overlap = 10) for the U.S. and Mexico I/O Tables. The R^2 difference increases monotonically with perturbation intensity, confirming AIPNET's sensitivity to meaningful economic relationships. Panel B explores temperature variability, with subfigures (c) and (d) illustrating the R^2 difference across a low (0.2) and high (0.1) temperature settings relative to a medium (0.6) temperature. The results show minimal impact to R^2 from temperature changes, indicating AIPNET's robustness to temperature-induced variability. The analysis uses the union of edges across iterations to fully capture the effect of temperature. Consistent results across I/O Tables from both countries highlight AIPNET's robustness in different economic contexts.

C Additional Results and Exercises for the Qatar exercise

C.1 Qatar Trade Data

Our study uses foreign trade data published by the Planning and Statistics Authority of Qatar.[45] For each month between January 2012 through June 2023 the data report value and volume (weight) of imports aggregated at the product category and country of origin level. The product categories follow the 8-digit Harmonised System (HS), embedded in the international 6-digit HS published in 2012 by the World Customs Organisation (WCO). We construct a balanced panel covering 102,715 combinations of product cat-

[45]The data is available at https://www.psa.gov.qa/en/statistics1/ft/pages/default.aspx.

egory and country of origin, giving 12.9 million observations across 10 years and 6 months. For exports, we similarly create a balanced panel covering 9,258 combinations of product category and target country, giving 1.3 million observations across 12 years and 6 months.[46]

C.2 Direct Effect of Blockade Exposure on Trade Patterns

We first show the results from a product-level difference-in-difference estimation approach outlined in Equation 9. Here, we use data for each 8-digit product in each month, collapsed across all countries. Our treatment is the blockade exposure of each good, defined as the share of import value sourced from blockading countries in 2016. Table C.4 reports results across two different levels of fixed-effect saturation. Panel C.4a includes 8-digit HS product level fixed effects, along with month dummies to account for seasonality, whereas C.4b additionally uses year-month fixed effects, interacted with 6-digit product codes. We find that the blockade is consistent with a large supply shock. Column (1) shows that for each additional percentage of exposure, post-blockade prices increased by around 0.35%. Columns (2) and (3) show that a one percent increase in exposure drove down import values and volumes by between 1.5-2.7%. Column (4) shows the probability of a non-zero trading month also fell by between 0.11 and 0.19 percent for each additional unit of exposure. Finally, above pre-blockade levels of monthly trade were also reduced by blockade exposure, as shown in Column (5).

(Table C.4)

We visualize the difference-in-difference setting for these direct effects of blockade exposure in Figure 15. We show that there is an abrupt and persistent increase in prices, and decrease in trade expenditures for goods which are more exposure to the blockade.

(Figure 15)

[46] As our treatment variable is defined at the product category level of imported goods, we restricted the export data to product categories that saw imports in 2016.

C.3 Descriptive Evidence

We document the blockade's impact on Qatar, highlighting two key facts: the blockade drove a large shift in trade patterns from blockading countries to the rest of the world and acted as a supply shock, leading to a 25 basis-point price increase and a 30 basis-point fall in trade volumes.

Appendix figure C.9 shows the rapid shift in import trade patterns that occurred after the blockade was unexpectedly introduced. Subfigure C.9a shows the total import trade value for all goods which are 'blockade-exposed', which we define as any HS 8-digit product code where the majority of import value in 2016 came from blockading countries. We see that the flow of trade abruptly falls after the introduction of the blockade, and a similar sized increase in trade value sourced from non-blocking countries. Subfigure C.9b shows the same analysis for all products, revealing a similar shift and highlighting how total trade volumes remained fairly stable.

Figure C.8 shows how prices and trade volumes (i.e. weight-volume) were affected after the blockade. Both figures depict an index constructed using 2016 as the base year. In both subfigures, we consider only those goods which are blockade-exposed. Each index is calculated by first calculating the HS 8-digit product level index. In each half-yearly period, we then aggregate across products using a weighted mean, weighted using the total value of trade for each product in 2016. Panel A depicts the abrupt 25 basis point increase in blockade-exposed products and that this price increase was also accompanied by a 30 basis-point fall in trade volumes. While these patterns are striking in and of themselves, we next outline an econometric design to test for the effects of blockade exposure–the share of 2016 imports sourced from blockading countries—on trade prices, expenditures, and volumes. The heterogeneous effects of the blockade are espoused in Panel B, which break both price and volume changes across broad HS section categories.

(Figure C.8)

C.4 Estimating direct effect of the blockade

In this section, we describe how we estimate the effect of the blockade on own-product import prices, trade values, and trade volumes. We further consider the impact on the importation of vertically related products, which are necessary in the domestic production of each good. Across all these outcomes, our treatment measure is the product-level $BlockadeExposure$.[47]

Direct Effect of Blockade Exposure Consider our baseline specification:

$$log(Y_{j,t}) = \beta \ PostJune2017_t \times BlockadeExposure_j + \alpha_j + \theta_t + \epsilon_{j,t} \tag{9}$$

where j is a product, defined at the HS 8-digit level, t represents each time period, PostJune2017$_t$ is a dummy variable which equals 1 for all year-months after June of 2017, BlockadeExposure$_j$ is a product's fraction of 2016 trade value that arrived from blockading countries, and α_j and θ_t are product and time fixed effects.[48]

The coefficient of interest is β, which represents the elasticity of our outcome measure with respect to the product-specific exposure to the blockade. A 1% increase in exposure leads to a β% increase in $Y_{j,t}$. The above specification can be further decomposed in order to measure the coefficient of interest, β, over time:

$$\log(Y_{j,t}) = \sum_{t}^{T} (\beta_t \times \theta_t \times \text{BlockadeExposure}j) + \alpha j + \theta_t + \epsilon_{i,t} \tag{10}$$

where again j is a product, defined at the HS 8-digit level, t represents each year-month, the sequences of $\{\beta_t\}_{t=1}^{T}$ contain the elasticity of product-specific exposure to the blockade on Y_{it} for each time period, including prior to and after the end of the blockade.

[47] Recall our $BlockadeExposure$ measure is the fraction of each focal product's total trade value which was sourced from blockading countries in 2016

[48] We also introduce alternative controls and fixed effects outlined in section **??**, but we omit them from the formula for brevity.

Lastly, we can decompose the specification across countries in order to recover heterogeneous changes in trade patterns across trading partners:

$$\begin{aligned}\log(Y_{j,c,t}) = \beta \times \text{PostJune2017}_t \times \text{BlockadeExposure}_j \times \text{Blockading}_c \\ + \gamma \times \text{PostJune2017}_t \times \text{BlockadeExposure}_j \times \text{NonBlockading}_c \\ + \alpha_{j,c} + \theta_{t,c} + \epsilon_{j,c,t}\end{aligned} \tag{11}$$

where again j is a product, defined at the HS 8-digit level, c represents the trading partner (country), t indexes month and year, β contains the elasticity of product-specific exposure to the blockade on Y_{jct} for trade with countries that participated in the blockade and γ contains the same elasticity for trade with other countries.

Table A.3: Our Network Predicts High Values in I/O Tables for the U.S. and Mexico

Panel A: United States						
Dependent Variables:	Use Value			Direct Requirements		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Network Score	0.0300***	0.0315***	0.0324***	0.1911***	0.1933***	0.1937***
	(0.0053)	(0.0055)	(0.0054)	(0.0328)	(0.0338)	(0.0312)
Fixed-effects						
Upstream Sector		Yes			Yes	
Downstream Sector		Yes			Yes	
Upstream Industry			Yes			Yes
Downstream Industry			Yes			Yes
Fit statistics						
Observations	41,209	41,209	41,209	41,209	41,209	41,209
R^2	0.04735	0.06054	0.10522	0.07678	0.08997	0.14393
Within R^2		0.04980	0.04697		0.07553	0.06879
Panel B: Mexico						
Dependent Variables:	Use Value			Direct Requirements		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Network Score	0.1560***	0.1551***	0.1590***	0.2658***	0.2734***	0.2867***
	(0.0376)	(0.0367)	(0.0362)	(0.0543)	(0.0538)	(0.0551)
Fixed-effects						
Upstream Sector		Yes			Yes	
Downstream Sector		Yes			Yes	
Upstream Industry			Yes			Yes
Downstream Industry			Yes			Yes
Fit statistics						
Observations	13,806	13,806	13,806	13,806	13,806	13,806
R^2	0.08778	0.09147	0.16077	0.10284	0.12014	0.16941
Within R^2		0.08464	0.08475		0.10702	0.10946

Note: This table reports the regression results according to equation 1. We regress the standardized Input-Output table values on our standardized network scores resulting from aggregating the HS 6-digit level network to the BEA-industry level for the U.S. and the corresponding industry level for Mexico. Columns (1) and (4) use no fixed effects, columns (2) and (5) employ upstream and downstream sector-fixed effects, and columns (3) and (6) employ industry-fixed effects. Both dependent and independent variables are winsorized at the bottom 0.5 percentile and top 99.5 percentile. The table shows that the Network Score is highly significant across all models, with coefficients ranging from 0.0334 to 0.2081 for the U.S. and from 0.1183 to 0.2735 for Mexico, depending on the specification. The R-squared values indicate the proportion of variance explained by the models, with the highest values observed when industry-fixed effects are included. Standard errors are clustered by both upstream and downstream industries. Significance Codes: ***: 0.001, **: 0.01, *: 0.05.

Table C.4: Effects of Product-level Blockade Exposure on Imports

(a) Controls for HS8 Product and Seasonality

Dependent Variables:	(1) Import Price (Log)	(2) Import Value (Log)	(3) Import Volume (Log)	(4) Nonzero Trade (Binary)	(5) High Trade (Binary)
Variables					
Blockade Exposure	0.35***	-1.6***	-1.5***	-0.11***	-0.07***
	(0.04)	(0.22)	(0.20)	(0.01)	(0.01)
Fit statistics					
Observations	661,664	941,472	941,472	941,472	941,472
R^2	0.759	0.586	0.601	0.532	0.589

(b) Controls for HS8 Product and HS6 Product×Year-Month

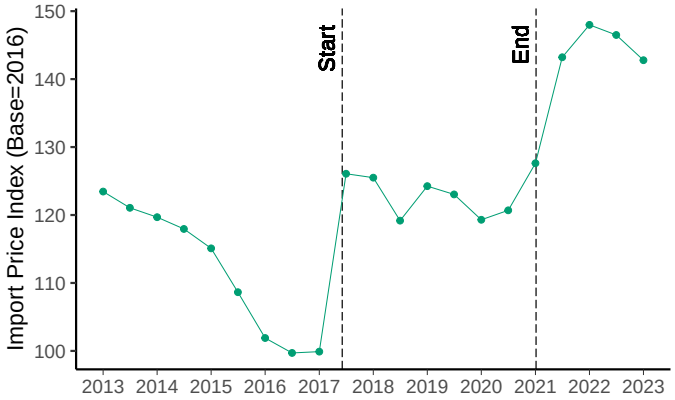
Dependent Variables:	(1) Import Price (Log)	(2) Import Value (Log)	(3) Import Volume (Log)	(4) Nonzero Trade (Binary)	(5) High Trade (Binary)
Variables					
Blockade Exposure	0.35***	-2.7***	-2.4***	-0.19***	-0.08***
	(0.05)	(0.26)	(0.23)	(0.02)	(0.01)
Fit statistics					
Observations	661,664	941,472	941,472	941,472	941,472
R^2	0.818	0.659	0.672	0.614	0.665

Note: This table reports results from running the regression defined in Section C.4 Equation (9). We test for the effect of 'Blockade Exposure', measured for each HS8 product as the proportion of the total value of trade that came from blockading countries in 2016. This is interacted with an indicator variable denoting the start of the blockade. The data used is at the monthly frequency of imports for each HS8 product. Columns (2) and (3) use the inverse-hyperbolic-sine transform to approximate log. Column (4) and (5) use as outcomes binary indicators for non-zero trade and high-trade respectively, where the latter equals 1 if trade volumes for that product in that year-month exceeded the 2016 average. Panel (a) contains fixed effects for HS8 product codes and dummy variables for each month to control for seasonality. Panel (a) adds an additional fixed effect for time (year-month dummies), which are interacted with HS6 Product Codes. Standard errors (clustered by HS4 product code) in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1.

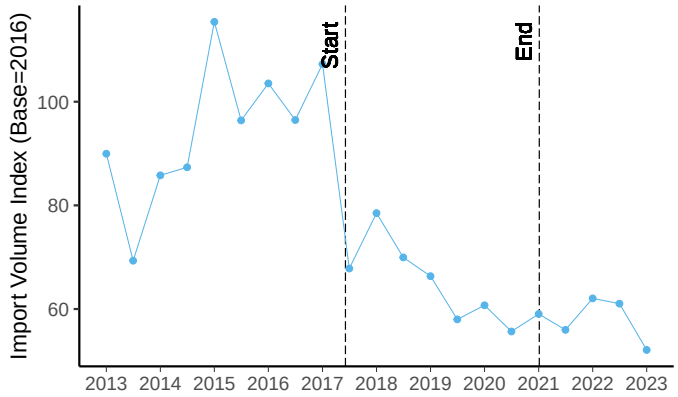
Figure C.8: Qatar Blockade Caused a Supply Shock for Trade-Exposed Products

Panel A: Aggregate patterns

(i) Prices Increased sharply after the Blockade



(ii) Volumes Contracted for All Product Sections



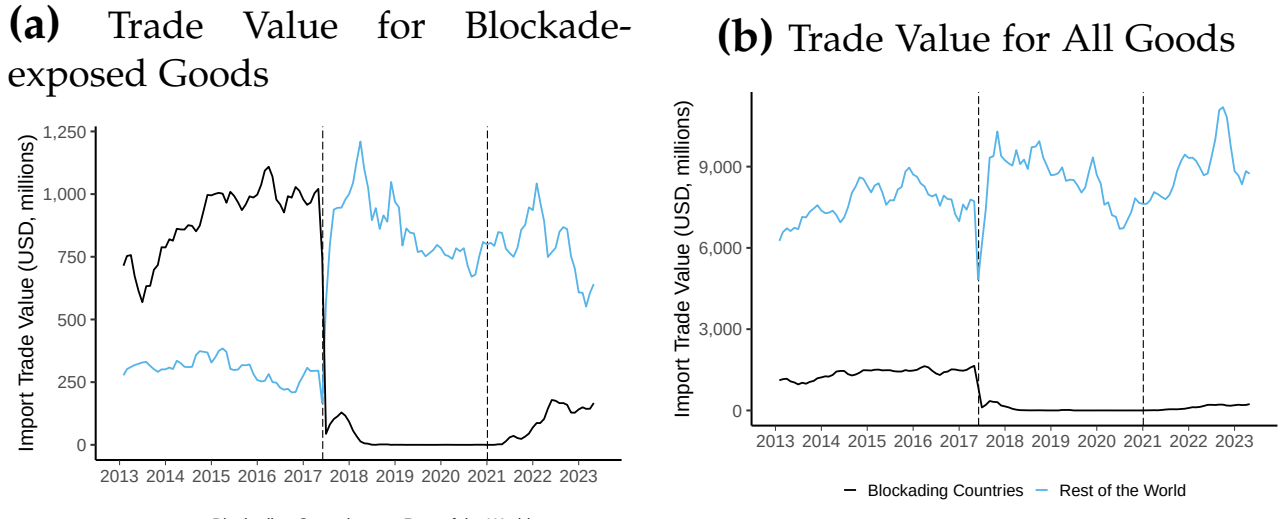
Panel B: Heterogeneity across goods

(i) Most Goods with sharp Increase in Prices after the Blockade

(ii) Volumes Contracted after the Blockade

Note: This figure shows the impact of the blockade on 'blockade exposed' goods and the heterogeneity across product sections. Panel A (a) displays a composite price index and (b) a composite volume index, both weighted by 2016 trade values. Vertical black dashed lines mark the start and end of the blockade. Panel B (c) shows percentage changes in average prices and (d) in total trade volumes between 2015-2016 and 2018-2019, with gradients indicating import value shares from blockading countries. "Works of art, collectors' pieces and antiques" is excluded for clarity.

Figure C.9: The Blockade substantially altered Trading Patterns



Note: This figure shows the monthly value of all goods imports, broken down by country-of-origin into two groups: Blockading Countries' includes Saudi Arabia, UAE, Bahrain, Egypt, Mauritania, Djibouti, and Maldives, and Rest of World', all remaining countries. Series are smoothed using a 3-month moving average, except during May-July 2017. Vertical black dashed lines mark the start and end of the blockade. Panel (a) includes only 'Blockade Exposed' 8-digit HS products, defined as goods where the majority of total import value was sourced from blockading countries in 2016. Panel (b) covers all goods.

D Relevant Network Statistics

This section describes a variety of relevant network statistics which are common in the literature, which we utilise to document shifts in the composition of trade flows. We leave our prefered and novel network measure, the Integrated Global Product Centrality (IGPC) to the next section proceeding this.

The most important of these is a measure of 'upstreamness', which was first introduced by Fally (2012) and Antràs et al. (2012). We extend this definition for the case of our highly granular product-level production network with discrete edges.

We also calculate degree-centrality and a page-rank centrality measure. These measures are implemented in R. For iterative measures like upstreamness, the computation stops when the change in scores falls below a threshold of 10^{-8}. Our implementation of these measures is publicly available.

Notation Let $G = (V, E)$ be a directed graph where V is the set of all HS6 products and E represents the binary input-output relationships between products. For each edge $(i, j) \in E$, product i is an input for product j. We define the adjacency matrix W as:

$$W_{ij} = \begin{cases} 1 & \text{if product } i \text{ is an input for product } j \ 0 \\ \text{otherwise} \end{cases} \tag{12}$$

Upstreamness Higher scores indicate products that are, on average, used as inputs in several stages of production before reaching final consumption. We define our unweighted upstreamness measure as follows:

$$\mathbf{u} = \mathbf{1} + \sum_{k=1}^{\infty} k \cdot \alpha^k (P^k \cdot \mathbf{1}) \tag{13}$$

where P is the row-normalized version of W:

$$P_{ij} = \frac{W_{ij}}{\sum_k W_{ik}} \tag{14}$$

and α is a decay factor (set to 0.5 in our implementation) to ensure convergence of the power series and to modulate the influence of distant connections in the production network.

Upstreamness (Weighted) We extend the upstreamness measure by incorporating global trade volumes. Let $\mathbf{v}$ be a vector of global trade values for each product. We define the weighted upstreamness as:

$$\mathbf{u_w} = \mathbf{1} + \sum_{k=1}^{\infty} k \cdot \alpha^k (P_w^k \cdot \mathbf{1}) \tag{15}$$

where P_w is a weighted transition matrix:

$$P_{w,ij} = \frac{W_{ij} \cdot v_j}{\sum_k W_{ik} \cdot v_k} \tag{16}$$

Degree Centrality Out-degree represents the number of products that use the given product as an input, while in-degree represents the number of products used as inputs for the given product. For a node i, we define:

$$\text{Out-degree centrality: } d_{out}(i) = \sum_j W_{ij} \quad \text{In-degree centrality: } d_{in}(i) = \sum_j W_{ji} \tag{17}$$

PageRank PageRank is a variant of Eigenvector Centrality that measures importance based on the quantity and quality of links to a node. Higher scores suggest products that are important inputs to other important products, considering both the quantity

and quality of connections in the network. It is defined recursively as:

$$PR(A) = (1-d) + d\sum_{i=1}^{n}\frac{PR(T_i)}{C(T_i)} \tag{18}$$

where d is a damping factor (typically set to 0.85), T_i are the nodes linking to A, and $C(T_i)$ is the number of outbound links from T_i.

These measures provide complementary views of a product's position and importance in the global production network, allowing for a comprehensive analysis of trade flow composition and structure.

E Integrated Global Product Centrality (IGPC) and AIPNET

The Integrated Global Product Centrality (IGPC) measure is fundamentally built upon the AI-generated Production Network (AIPNET). This section details how AIPNET is incorporated into the IGPC calculation and its significance in determining product centrality.

AIPNET is represented as a directed graph $G = (V, E)$, where V is the set of nodes, each representing a product in the Harmonized System (HS) classification, and E is the set of directed edges, representing input-output relationships between products. Each edge $e_{ij} \in E$ indicates that product i is an input in the production of product j. This network structure captures the complex interdependencies in global production processes.

The IGPC measure is defined by the equation

$$\mathbf{X} = (1-d)\mathbf{B} + d\mathbf{AWX} \tag{19}$$

where $\mathbf{X}$ represents the vector of IGPC scores, with X_i representing the IGPC score of product i. The parameter $d \in (0,1)$ is a damping factor (typically set to 0.85), $\mathbf{B}$ is the

base importance vector, $\mathbf{A}$ is the adjacency matrix derived from AIPNET, and $\mathbf{W}$ is a diagonal matrix of weight adjustments.

The AIPNET structure is integrated into the IGPC formulation through the adjacency matrix $\mathbf{A}$. The elements of $\mathbf{A}$ are defined as follows:

$$A_{ij} = \begin{cases} 1 & \text{if } e_{ji} \in E \text{ (i.e., product } j \text{ uses product } i \text{ as input)} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

This encoding in $\mathbf{A}$ ensures that the IGPC measure reflects the actual production relationships between products.

The integration of AIPNET into IGPC has several implications. The term $\mathbf{AWX}$ in the IGPC equation represents how importance flows from downstream products to their inputs, allowing product j to contribute to the importance of all its input products i where $A_{ij} = 1$. Products with high out-degree in AIPNET, serving as inputs to many other products, thus accumulate importance from a wider range of downstream products. The iterative nature of the IGPC calculation captures the significance of products that might not have high direct trade volumes but are critical in complex production chains. Additionally, the weight matrix $\mathbf{W}$ modifies the importance flow defined by $\mathbf{A}$, incorporating trade volumes (through Global Trade Share, GTS) so that high-volume trade relationships exert a stronger influence on importance propagation.

To further elucidate the role of AIPNET, the IGPC update equation for a single product i can be expressed as:

$$X_i^{(t+1)} = (1-d)B_i + d \sum_{j \in N_{out}(i)} \frac{W_{jj} X_j^{(t)}}{\sum_{k \in N_{in}(j)} W_{kk}} \tag{21}$$

where $N_{out}(i)$ is the set of products that use i as an input (outgoing edges in AIPNET), and $N_{in}(j)$ is the set of products used as inputs for j (incoming edges in AIPNET). Here,

B_i is the base importance of product i, defined as:

$$B_i = \left(\frac{TC_i - \min(TC)}{\max(TC) - \min(TC)} \right)^{\alpha} \times \left(\frac{GTS_i - \min(GTS)}{\max(GTS) - \min(GTS)} \right)^{\beta} \tag{22}$$

where TC_i is the Trade Concentration and GTS_i is the Global Trade Share of product i. The diagonal element W_{ii} in $\mathbf{W}$ corresponding to product i is given by

$$W_{ii} = \left(\frac{GTS_i - \min(GTS)}{\max(GTS) - \min(GTS)} \right)^{\gamma} \tag{23}$$

showing how the IGPC score of a product depends on its position in the AIPNET structure, specifically its relationship to downstream products.

The IGPC measure balances the structural information provided by AIPNET with empirical trade data. While AIPNET supplies the structure of product relationships and potential paths for importance propagation, GTS and TC provide product-specific weights that adjust this structure based on observed trade patterns. The damping factor d adjusts the relative importance of network structure versus intrinsic product importance (as measured by GTS and TC), and parameters α, β, and γ control the influence of TC, GTS, and their interaction in the network structure. Through this integration of AIPNET with trade volume and concentration data, IGPC provides a comprehensive measure of product centrality, reflecting both the complex structure of global production networks and the empirical realities of international trade.

F Estimating Structural Breaks in Unit Prices

Measuring persistent price shocks to finished goods To identify structural breaks in the unit price index for each country-product pair, we first residualise each each series absorbing out time invariant country and product-specific shifts. Specifically, we use the log of the unit prices as the outcome variable and run the following econometric

specification:

$$\log(p_{d,c,t}) = \nu_{d,c} + \eta_{d_{hs2},t} \tag{24}$$

The specification removes idiosyncratic good and country-specific level differences in prices, ($\nu_{d,c}$). Further, we remove HS2 good and country specific time fixed effects, $\eta_{d_{hs2},t}$, which may capture general supply- or demand-imbalances within similar class of goods.[49]

We then obtain the residuals from this regression, denoted as $\hat{\varepsilon}_{d_{hs4},c,t}$, and code a country-by-good-by-year specific observation as an outlier if a specific good d_{hs4} has both, an above-median residual that also is positive. That is, we create a binary indicator:

$$\mathbb{I}_{\hat{\varepsilon},t} = \mathbb{I}\left(\hat{\varepsilon}_{d,c,t} > \text{median}(\hat{\varepsilon}_{d,c,t}) \cap \hat{\varepsilon}_{d,c,t} > 0\right) \tag{25}$$

We focus on above median positive supply shocks that are *persistent*. That is, for a shock to classify as a shock, we require prices to be elevated for at least three consecutive periods:

$$\text{Shock}_{d,c,t} = \begin{cases} 1 & \text{if } \mathbb{I}_{\hat{\varepsilon},t} = 1, \mathbb{I}_{\hat{\varepsilon},t-1} = 1, \mathbb{I}_{\hat{\varepsilon},t-2} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

[49] This measure is rather ownerous, as it permits only shifts in idiosyncratic and granular product unit prices. As such, it may under-report larger global shocks.