Scientific Reports

https://doi.org/10.1038/s41598-026-47883-y

Article in Press

Inferring firm-level supply chain networks with realistic systemic risk from industry sector-level data

Received: 16 December 2025

Accepted: 3 April 2026

Published online: 29 April 2026

Cite this article as: Fessina M., Cimini G., Squartini T. *et al.* Inferring firm-level supply chain networks with realistic systemic risk from industry sector-level data. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-47883-y

Massimiliano Fessina, Giulio Cimini, Tiziano Squartini, Pablo Astudillo-Estévez, Stefan Thurner & Diego Garlaschelli

Inferring firm-level supply chain networks with realistic systemic risk from industry sector-level data

Massimiliano Fessina[1], Giulio Cimini[*2,3], Tiziano Squartini[1,4], Pablo Astudillo-Estévez[5,6,7], Stefan Thurner[6,8,9], and Diego Garlaschelli[1,4,10]

[1]IMT School for Advanced Studies, 55100 Lucca (Italy)
[2]Physics Department and INFN, University of Rome 'Tor Vergata', 00133 Rome (Italy)
[3]Enrico Fermi Research Center (CREF), 00184 Rome (Italy)
[4]INdAM-GNAMPA Istituto Nazionale di Alta Matematica, 00185 Rome (Italy)
[5]Colegio de Economía, Universidad San Francisco de Quito, 170901 Quito (Ecuador)
[6]Complexity Science Hub, A-1080 Vienna (Austria)
[7]Institute for New Economic Thinking, University of Oxford, OX1 3UQ Oxford (UK)
[8]Section for Science of Complex Systems, Medical University of Vienna, A-1090 Vienna (Austria)
[9]Santa Fe Institute, NM 87501 Santa Fe (USA)
[10]Lorentz Institute for Theoretical Physics, University of Leiden, 2333 CA Leiden (Netherlands)

ABSTRACT

Production networks form the structural backbone of modern economies, yet recent crises have exposed their intrinsic vulnerability. Quantifying the system-wide impact of local disruptions—i.e., systemic risk—requires detailed knowledge of supply chain networks, since such risks depend sensitively on their underlying topological patterns. In practice, however, microscopic network data are rarely available, forcing systemic-risk estimates to rely on coarse sector-level information, such as input/output tables, and aggregate production data at the firm-level. Here we assess the ability of maximum-entropy approaches to reconstruct supply chain networks with realistic level of systemic risk when only sector- and firm-level information is available. Leveraging the unique availability of the complete production network of Ecuador, we benchmark four reconstruction models with different formulations and data requirements, by comparing the inferred structural and systemic-risk properties (particularly, the Economic Systemic Risk Index) with those of the real network. We find that the *stripe-corrected gravity model*, which incorporates firm-specific input requirements disaggregated by sector, most accurately reproduces the systemic-risk content of the empirical network, underscoring the importance of capturing heterogeneity in firms' input profiles. Inferring the latter using sectorial input/output tables—through the *input-output gravity model* we introduce here—still yields satisfactory estimates while requiring substantially less information. Our results identify the minimal sector-level information needed to statistically generate synthetic production networks that faithfully encode firm-level systemic risk.

Introduction

Supply chains emerge since firms purchase goods from other firms as inputs for their own production. While the resulting network of relationships constitutes the backbone of any modern economy, it also leads to inherent fragility: disruptions can propagate through inter-firm linkages and amplify into economy-wide shocks[1,2]. Prominent examples are provided by the 2011 Fukushima earthquake, whose economic repercussions spread far beyond the region directly affected by the disaster[3–8], and the Covid-19 pandemic, when global supply chains faltered causing shortages of basic goods in multiple countries[4,7–10]. Developing tools that help governments and policymakers anticipate and mitigate the economic impact of such disruptions has therefore become a scientific and institutional priority.

Traditionally, the propagation of economic shocks has been studied using industry-level input/output (I-O) tables[11,12]. Only recently, firm-level production networks have been used to assess how shocks spreading across individual firms can generate aggregate fluctuations[13–16]. Early work in the supply chain management literature quantified the extent of failures caused by a propagating shock—the so-called *ripple effect*[17–19]. More recent approaches incorporate the explicit structure of the production network[20]. These include the *Nexus Supplier Index* (NSI), which evaluates the exposure associated with hidden suppliers deep inside supply chains[21,22], and the *Economic Systemic Risk Index* (ESRI)[23], which measures aggregate output losses using non-linear shock propagation dynamics ruled by generalized Leontief production functions.

*Corresponding author: giulio.cimini@roma2.infn.it

However, firm-level data are typically proprietary and rarely accessible at scale[24]. Global datasets such as those by Capital IQ[25,26], Compustat[27–29] and FactSet[30] cover only the largest publicly listed firms (mainly in the U.S.) and report only major customer connections, without link weights. The Tokyo Shoko Research dataset[31–34] is survey-based and again region-specific. Less geographically-biased datasets exist but tend to be restricted to specific sectors—for example, MarkLines for automotive supply chains[35,36]. National-level administrative datasets, based on VAT records or bank-mediated payment data, offer much richer coverage but only a few countries collect them as of today (e.g., Belgium[37], Ecuador[38], Hungary[23], Spain[39], Brazil[40], Japan[41,42], The Netherlands[43]). Additionally, access to such datasets remains heavily restricted due to confidentiality, as they contain highly sensitive information.

These limitations have spurred the development of methods to infer the unknown structural details of production networks from partial information[44]. On one hand, link prediction approaches use machine learning to estimate whether a buyer–supplier tie exists based on partial topology, financial indicators, sectorial and geographic similarity[38,45,46], or even news signals[47]. On the other hand, network reconstruction approaches do not rely on network snapshots. Some methods infer trading relationships from external information, such as correlations in firms' financial time series[48] or mobile phone interactions[49]. Other approaches, inspired by statistical physics[50], use aggregate firm-level information to reconstruct trade volumes from topology[51] or to infer both topology and link weights using firm-level information disaggregated by sector[43].

Here we focus on the latter approach, which belongs to the broader family of fitness-induced Exponential Random Graph (ERG) models[52] and has been successfully used in past research to reconstruct the structural properties of several kinds of economic and financial networks[53]. Specifically we assess four different formulations tailored to production networks. The recently proposed *Stripe-Corrected Gravity Model* (SCGM)[43], which constraints how much each firm sells overall and buys from each sector, plus the network density. The *Input-Output Gravity Model* (IOGM), introduced here as a SCGM variant operating with markedly reduced data requirements—by inferring the input-by sector of each firm using sector-to-sector flows. The *Density-Corrected Gravity Model* (DCGM)[54], the state-of-the-art for reconstructing trade and financial networks, a simpler version of the SCGM constraining only the total input of each firm. The *Stripe-Corrected MaxEnt Model* (SCMM), a standard MaxEnt formulation constraining input-by sector and total output of each firm but no topological information.

We test these models in reconstructing the Ecuadorian production network, with the aim of assessing how accurately each model can recover both its topological structure and its firm-level systemic risk content. In a nut, we show that models preserving firms' sector-specific inputs and outputs—i.e. SCGM and, approximately, IOGM—most successfully reproduce the observed distribution of firm-level systemic risk values (as measured by ESRI), thereby generating synthetic production networks with realistic risk profiles.

Results

Data description and model inputs

We derive the Ecuadorian production network from VAT data collected by the Internal Revenues Service (IRS) of Ecuador[55], covering year 2008. Data are cleaned (see Supplementary Materials S1) to represent a weighted, directed network with 2.189.066 transactions (links) and 60.488 firms (nodes), classified into 387 industrial sectors at the ISIC 4 digits level. The network is represented by an adjacency matrix W, where $w_{i\to j}$ is the amount of goods (products or services) firm i sells to firm j, measured by the total tax that j pays to i for the purchase. We then filter transactions by keeping only the links whose weight exceeds a threshold of 22.300 \$, due to computational feasibility reasons (see Supplementary Materials S2 for a robustness analysis against this threshold). By doing so, the number of connections reduces to 130.044, firms to 29.089 and sectors to 371 (see Supplementary Materials S2 for details). Note that we only have information on the industrial sector of firms but not on the type of good associated with each transaction. Hence, as in related studies[23,43], we assume that firms sell only one type of good, represented by its industrial sector (ISIC code).

Using the adjacency matrix of the production network, we compute firm-level and sector-level quantities that will constitute the input of the reconstruction models. DCGM constrains global firm-level quantities: the in-strength $s_i^{\text{in}} = \sum_j w_{j\to i}$ and out-strength $s_i^{\text{out}} = \sum_j w_{i\to j}$ of each firm i, respectively representing the total value of goods bought from and sold to other firms, plus the total link density of the network. IOGM further constrains the total value of goods sold by all firms in sector g_k to all firms in sector g_l: $s_{g_k\to g_l} = \sum_{i\in g_k}\sum_{j\in g_l} w_{i\to j}$, plus the link density within each sector. SCGM constrains the total out-strength plus the in-strength by sector of each firm i, $s_{g_k\to i} = \sum_{j\in g_k} w_{j\to i}$—that is, how much i buys from suppliers belonging to sector g_k—plus the link density within each sector. Finally SCMM has the same input of SCGM without link density information, thus generating dense random topologies. We remand to the Methods section for detailed definitions of the four models.

Reconstruction of structural network features

Here we assess to what extent the purely structural properties of the Ecuadorian production network are reproduced by the four reconstruction models (see Supplementary Materials S3 for the definition of the topological quantities considered). SCMM is not considered in this comparison as it does not preserve any topological information of the empirical network.

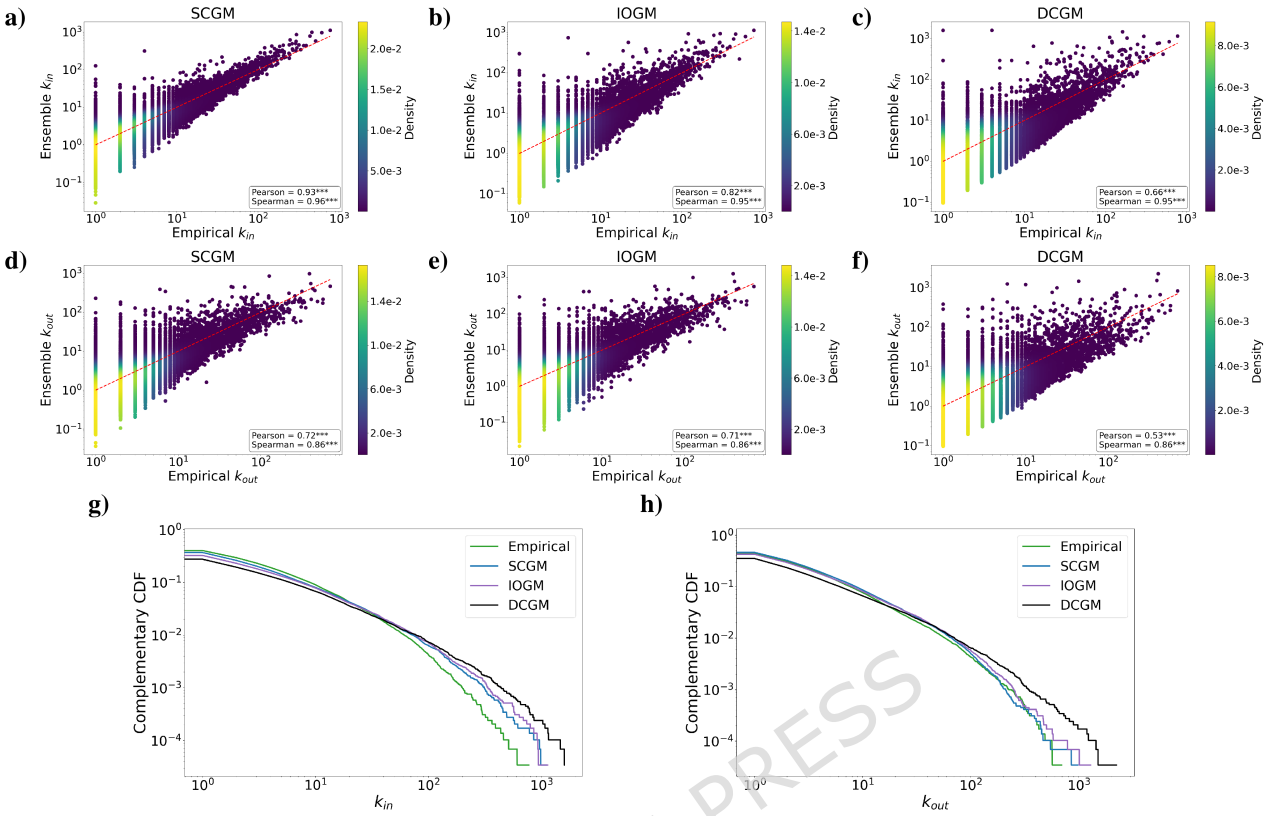

Figure 1. Reconstruction of node degrees. Scatter plot of empirical in-degrees (a-c) and out-degrees (d-f) versus reconstructed values, for SCGM (a,d), IOGM (b,e) and DCGM (c,f). The identity line is shown with red dasehs. Complementary Cumulative Distributions of in-degrees (g) and out-degrees (h) for the empirical production network and the three reconstruction models.

	$k^{in\text{-}in}$	$k^{in\text{-}out}$	$k^{out\text{-}in}$	$k^{out\text{-}out}$	$s^{in\text{-}in}$	$s^{in\text{-}out}$	$s^{out\text{-}in}$	$s^{out\text{-}out}$	C^3	C^4
SCGM	0.21	0.16	0.94	2.39	0.08	0.09	0.51	0.12	0.26	0.64
IOGM	1.12	1.83	2.12	0.79	0.10	0.16	0.56	0.11	0.86	2.83
DCGM	16.83	35.86	25.63	7.70	1.25	2.60	2.36	0.51	11.68	53.65

Table 1. Reconstruction of topological patterns. Agreement between empirical and reconstructed quantities, measured by the reduced chi-squared distance $\chi^2 = \frac{1}{n}\sum_i (y_i^{(emp)} - y_i^{(rec)})^2/(\mathrm{Var}_i^{(emp)} + \mathrm{Var}_i^{(rec)})$ among empirical ($y^{(emp)}$) and model ($y^{(mod)}$) curves, normalized by the respective dispersions, where i indexes the bins along the x-axis, and n is the number of bins (see plots in Supplementary Materials S3). The evaluated quantities are all the combinations of average nearest neighbours degrees and strengths, plus triangular and square clustering coefficient. $\chi^2 < 1$ signals a very good agreement between the curves, while values $\chi^2 \gg 1$ indicate large discrepancies. Note that χ^2 values are generally smaller for weighted quantities, since all models preserve node strengths on average.

We start from firm-level connectivity, measured by in-degrees k_{in} (number of suppliers) and out-degrees k_{out} (number of customers), which are not constrained by any model. Nevertheless, as Figure 1 shows, degrees are generally well reproduced thanks to the *fitness ansatz* (see Methods). As quantified by the values of the correlation coefficients between empirical and reconstructed degree values (computed from the scatter plots of Figure 1(a-f)), SCGM yields the best agreement, closely followed by IOGM, while DCGM displays the largest dispersion around the identity line, causing a drop in performance. Interestingly, all models perform better in the reconstruction of in-degrees rather than out-degrees. These results are confirmed by looking at the cumulative distributions of reconstructed in- and out-degrees (shown in Figure 1(g-h)), which closely resemble the empirical one. The fat tails of the distributions are in line with empirical evidence[24,34] and consistent with the *granular hypothesis* put forward in[15]. Still, all model distributions display a fatter tail than the empirical case, meaning that large

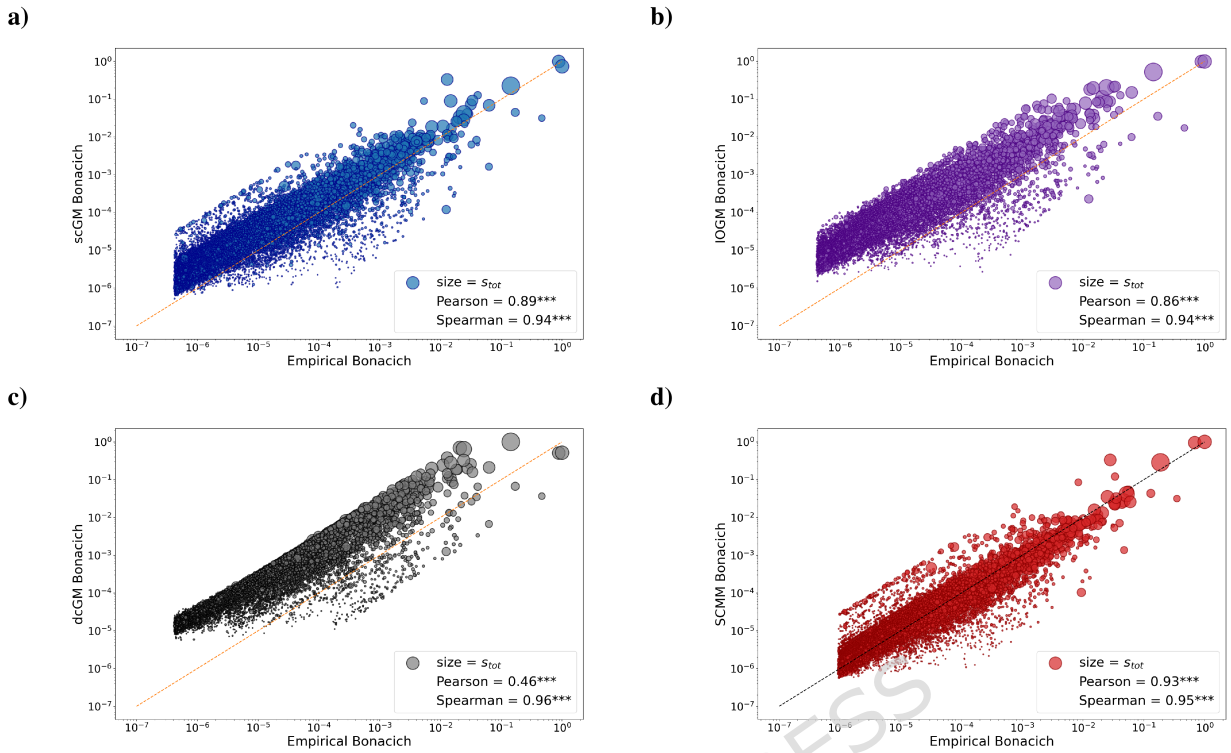

Figure 2. Reconstruction of firm-level systemic risk (Bonacich). Scatter plots of empirical versus reconstructed values, for SCGM (a), IOGM (b), DCGM (c) and SCMM (d), with values of Pearson and Spearman correlation coefficients. Dots' size is proportional to firms' out-strength (total sales).

(small) degrees are overestimated (underestimated). Moreover, the skewness of the distribution increases for models with less constraints. Indeed, models admit the presence of the link $i \to j$ only if $s_{g_i \to j} \neq 0$ (SCGM), $s_{g_i \to g_j} \neq 0$ (IOGM), $s_i^{\text{out}} \neq 0$ and $s_j^{\text{in}} \neq 0$ (DCGM). Therefore, firms tend to have more links under DCGM than SCGM.

Moving to higher-order topological properties (results reported in Table 1 and Supplementary Materials S3), analysis of average nearest neighbours degree and of average nearest neighbours strength reveals that the Ecuadorian production network is disassortative: firms having more customers (suppliers) tend to be connected with firms having fewer suppliers (customers), and vice-versa. This is a well documented property of national production networks[33,34]. SCGM is capable of capturing these trends, including the fluctuations of highly central firms, closely followed by IOGM. The analysis of the (binary) triangular and square clustering coefficients reveals a hierarchical organisation of the Ecuadorian production network: less connected firms participate in a higher fraction of closed triangular and square motifs with respect to highly connected ones. This trend is much less pronounced for the square clustering, showing that also firms with many connections tend to close a high number of square motifs—in line with the documented *complementarity-driven* self organization of production networks[33,36,56]. Despite their high-order nature, these patterns are reproduced by both SCGM and IOGM, showing that constraining sector-level information can boost model performance. Overall, as Table 1 shows, SCGM consistently provides the best agreement of reconstructed vs empirical quantities, sometimes at par with IOGM, while DCGM leads to largely deteriorated results.

As documented in the Figures of Supplementary Materials S2 and S3, all these topological quantities are well reconstructed by SCGM and IOGM also on the full, unfiltered network. The reconstruction accuracy, however, is enhanced by imposing the threshold, suggesting that removing small noisy transactions may improve the data quality.

Reconstruction of systemic risk

Bonacich centrality

Before considering ESRI, as a first quantification of systemic risk we use the Bonacich centrality[57,58], a mathematical equivalent of the influence vector[13] (see Methods). The Bonacich centrality of a firm quantifies the percentage change in total output of the economy, resulting from a 1% productivity shock to the firm, using linear propagation of losses.

As Figure 2 shows, all reconstruction models reproduce Bonacich centrality with high accuracy: SCGM and IOGM achieve

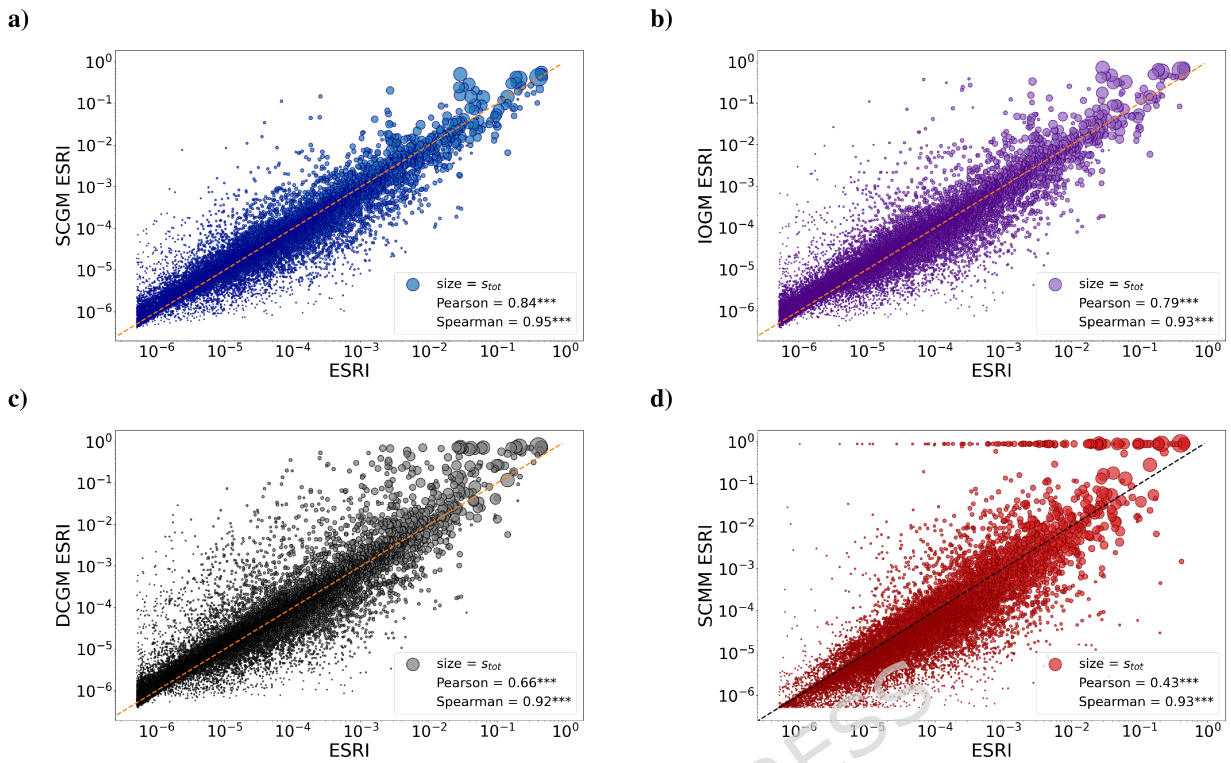

Figure 3. Reconstruction of firm-level systemic risk (ESRI). Scatter plots of empirical versus reconstructed values, for SCGM (a), IOGM (b), DCGM (c) and SCMM (d), with values of Pearson and Spearman correlation coefficients. Dots' size is proportional to firms' out-strength (total sales).

high correlations in terms of both values and rankings (Pearson = 0.89 and 0.86, Spearman = 0.94). DCGM maintains a high rank correlation (Spearman = 0.96) but performs less accurately on the absolute values (Pearson = 0.46). Surprisingly, the best performing model is the SCMM (Pearson = 0.93 and Spearman = 0.95), which disregards network topology entirely while exactly preserving the in-strengths by sector and total out-strength of all firms. The explanation can be traced back to the linear formulation of the Bonacich centrality: the first term in the series expansion of eq. (10) is directly proportional to firms' out-strength. This in turn implies that the influence vector disregards possible non-linearities stemming from the heterogeneous productive structure of firms, which are instead at the core of ESRI (see below). Overall, the constraint on the in-strengths by sector preserves the so-called technical coefficients of each firm; this, together with the constraint on firms' total revenue, is sufficient to capture propagations under linear dynamics, even without a network topology.

Economic Systemic Risk Index

The Economic Systemic Risk Index (ESRI) is a recently-introduced firm-level risk indicator[23], based on a shock propagation dynamics defined using a generalised Leontief production function: inputs from essential sectors set a hard constraint on the output of firms, while inputs from non-essential sectors are treated in a linear way. Briefly speaking, the systemic risk of firm i is evaluated by *a)* removing i from the network; *b)* running two, iterative processes accounting for both the upstream (demand) and downstream (supply) components of the propagating shock; *c)* computing the output reduction experienced by the whole production network as the sum of the output reductions for individual firms, after the convergence of the two shockwaves (see Methods for the precise mathematical formulation).

Figure 3 shows how ESRI values of firms computed on the empirical network compare with those computed on the reconstructed networks (numerically evaluated by averaging results of ESRI dynamics over 10^3 sampled configurations). Both SCGM and IOGM lead to a very good agreement, with Pearson correlation coefficient $\simeq 0.8$ for both models. DCGM and SCMM perform much worse due to a larger dispersion of values: Pearson correlation coefficient is $\simeq 0.7$ for DCGM $\simeq 0.4$ for SCMM. SCMM in particular wrongly assigns the maximum ESRI value to a large set of firms, whose empirical ESRI can be rather small. These firms correspond to suppliers of essential goods, whose failure halts production for their customers (due to the non-linear downstream propagation mechanism of ESRI). SCMM assigns these firm the maximum possible degree, so their failure causes the collapse of the entire economy, while empirical ESRI is sensitive to the specific topology of the network.

Results for the individual upstream and downstream components of ESRI (reported in Supplementary Materials S4) reveal that SCMM works well for the upstream component, which is computed via a linear propagation equation, while it fails for the downstream one, ruled by non-linear propagation. Together with the Bonacich centrality analysis, this confirms that realistic risk metrics involving nonlinear propagation of shocks do require realistic reconstruction of topology, while less-realistic metrics based on linear propagation do not. Another observation is that, overall, models tend to underestimate the upstream and overestimate the downstream components. This can be imputed to the density of the sampled configurations on which ESRI is computed. While models are designed to preserve the empirical link density *on average*, specific samples can contain disconnected nodes, leading to a larger density of the connected component with respect to the empirical one[59]. Then, while distributing s_i^{in} onto more incoming links (more suppliers) decreases the upstream impact of firm i, distributing s_i^{out} onto more outgoing links (customers) might increase its downstream impact, when the production of such customers is hard-constrained by the essential input provided by i. In order to test this intuition we inspected, for the empirical network and SCGM samples, the upstream and downstream *criticality* of firms. These are computed, for each firm, as the sum of the criticality of its upstream and downstream links, representing the impact of the firm's failure on its first neighbours (see Supplementary Materials S5 for the explicit derivation of these quantities). On one hand, concerning the upstream propagation, 60% percent of firms exhibit a higher criticality in the empirical network than in the model; on the other hand, 53% of firms are assigned by SCGM a higher downstream criticality than in the empirical network.

Analysis of the plateau. We further test the models in reproducing the ESRI *plateau*, namely the peculiar shape of ESRI values when plotted against the corresponding rankings[23]. Figure 4(a-d) reveals that all models do recover a plateau induced by the riskiest firms and the subsequent steep decrease of the trend. Still, the height and size of the plateau increase for less constrained models, which overestimate both the ESRI value and the number of highly risky firms in the network. This problem can be due (at least partially) to models' mathematical formulation. Since link probabilities and weights are proportional to the strengths of the involved firms, models tend to attribute more links and with larger weights to bigger firms. This effect increases for models with less constraints, as SCGM and IOGM connect firms only if their sectors are connected in the empirical network, while DCGM can connect any two firms. The number of links assigned to large firms reaches its maximum with SCMM which, despite following sectorial constraints, assigns a non-zero weight to each link compatible with these constraints. Consistently with this intuition, an inspection of the subgraphs of the most 15 risky firms in the empirical network and in SCGM ensemble reveals that the model fails to retrieve the high risk associated with 6 of the smallest firms in the subgraph, and replaces them with 6 much bigger firms (see Supplementary Materials S6).

The plateaux shown in the main panels of Figure 4(a-d) are obtained by plotting empirical and model ESRI values according to their own ranking. Empirical and model rankings are highly correlated (Spearman correlation coefficient larger than 0.9 for all models), but do not coincide perfectly. Thus we report in the insets the model plateaux when firms are ordered according to their empirical ranking (so that vertically aligned points represent the same firm). The superior performance of SCGM is even more evident now, as the other models substantially overestimate ESRI values of many non-risky firms, with SCMM showing the worst performance.

As the failure of the top ~ 10 firms belonging to the plateau would induce an output reduction of 40%-50% of the total value of the Ecuadorian economy, their identification represents an important test for the performance of our reconstruction models. Overall, as Figure 4(e) shows, the number of highly risky firms (in the top N positions of the ESRI ranking) that are correctly identified by the models decreases as we browse the ranking by increasing N. However, both SCGM and IOGM correctly identify 7 out of the top 10 (9 out of the top 15) firms, with SCGM providing the largest overlap for bigger values of N. More quantitatively, we can introduce the total relative error of reconstructed against empirical ESRI values, reading $\text{TRE}_m^N = \sum_{r=1}^N |1 - \langle \text{ESRI}_r \rangle_m / \text{ESRI}_r|$, with r indicating the r-th ranked firm according to the empirical ESRI values, N the number of firms considered and m the respective model values. As Figure 4(f) shows, the superior performance of SCGM becomes even more evident when scattering TRE_m^N versus N, for each model in our basket. Additionally, the plot confirms that IOGM works better in recovering the ranking of the firms than in reproducing their ESRI values.

Analysis of sector-level ESRI values. We assign an ESRI value to each of the 371 industrial sectors in the data (ISIC 4 digits classification) as the sum of the ESRI values of the firms belonging to the sector (see Supplementary Materials S1). This measure is just a proxy for the actual ESRI of a sector, which should be computed by simultaneously shocking all of its firms, but is useful to understand how the reconstruction models perform at a coarser level. Indeed, as Figure 5 shows, scattering the empirical versus model sector-level ESRI values returns a quite different picture from the scatter plots at the firm-level (Figure 3). Now SCGM and IOGM display an even better performance, while the performance of the other models drops significantly. Beside confirming the superior performance of SCGM, this result highlights the importance of achieving an accurate reconstruction at the micro-level, in order to obtain a proper reconstruction at the aggregate level as well.

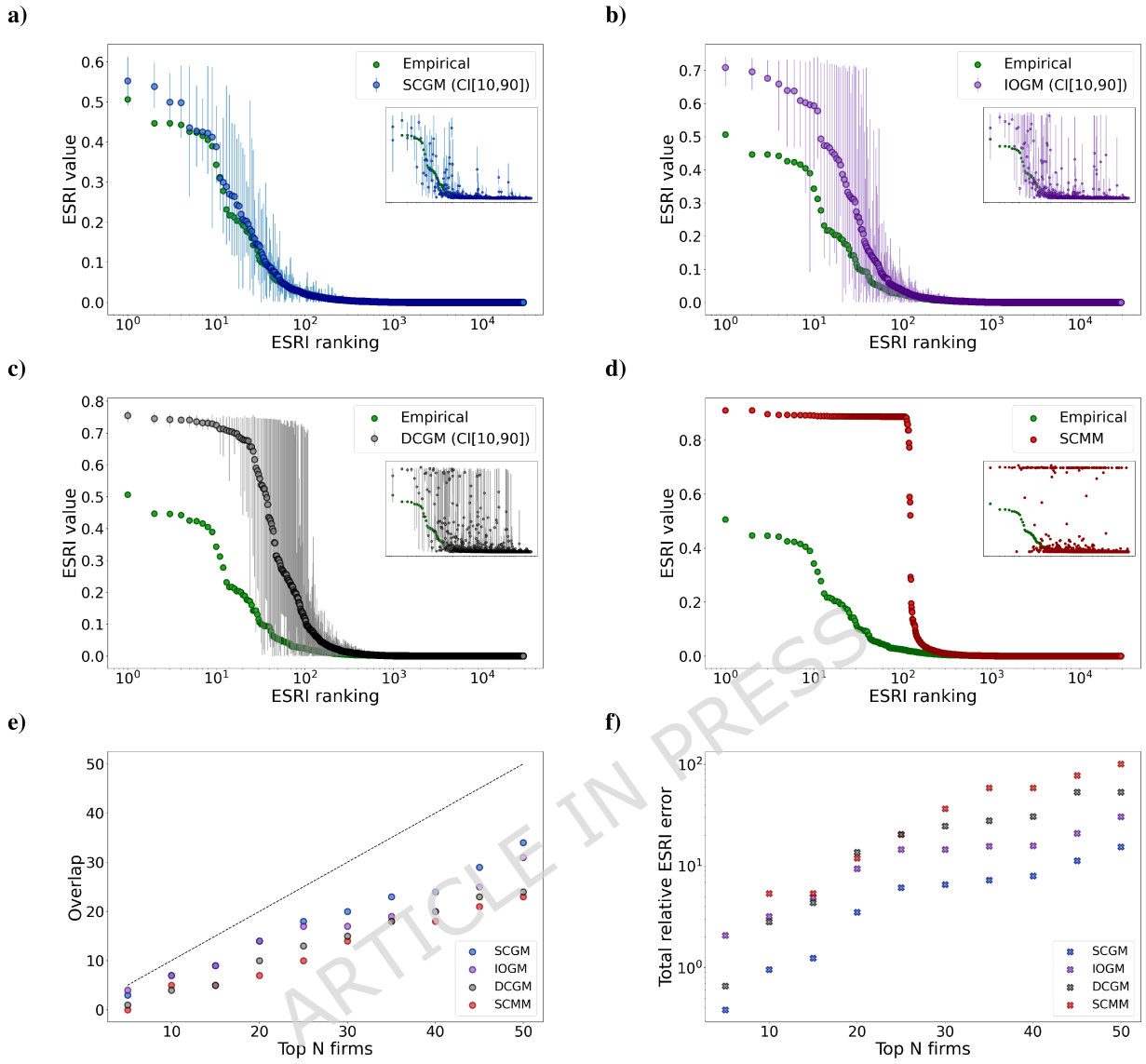

Figure 4. Reconstruction of ESRI ranking and plateau. (a-d) Scatter plots of empirical and reconstructed ESRI values versus their corresponding rankings, for SCGM (a), IOGM (b), DCGM (c) and SCMM (d). Insets: same plot of the main panel, with reconstructed ESRI values reordered according to the empirical ranking. (e) Overlap of the top N riskiest firms according to empirical and model ESRI rankings. (f). Total relative error of ESRI values for the top N riskiest firms in the empirical ranking.

Discussion

The absence of large-scale, publicly accessible firm-level datasets continues to limit systematic empirical analyses of production networks. As a result, methods for reconstructing such networks from partial information have become increasingly important. Although a variety of reconstruction techniques have been proposed, their validation has primarily focused on predicting the presence of supply links[38,44–46,60], reproducing structural properties[43,49], or recovering aggregate economic quantities in static settings[51]. A systematic assessment of their ability to capture firm-level systemic risk—a central question for understanding shock propagation in real economies—has remained largely unexplored. Our work addresses this gap by assessing four reconstruction models—the *Stripe-Corrected Gravity Model*[43], the *Input-Output Gravity Model*, the *Density-Corrected Gravity Model*[54] and *Stripe-Corrected MaxEnt Model*—on their ability to reproduce the Economic Systemic Risk index (ESRI)[23] values of firms in the Ecuadorian production network.

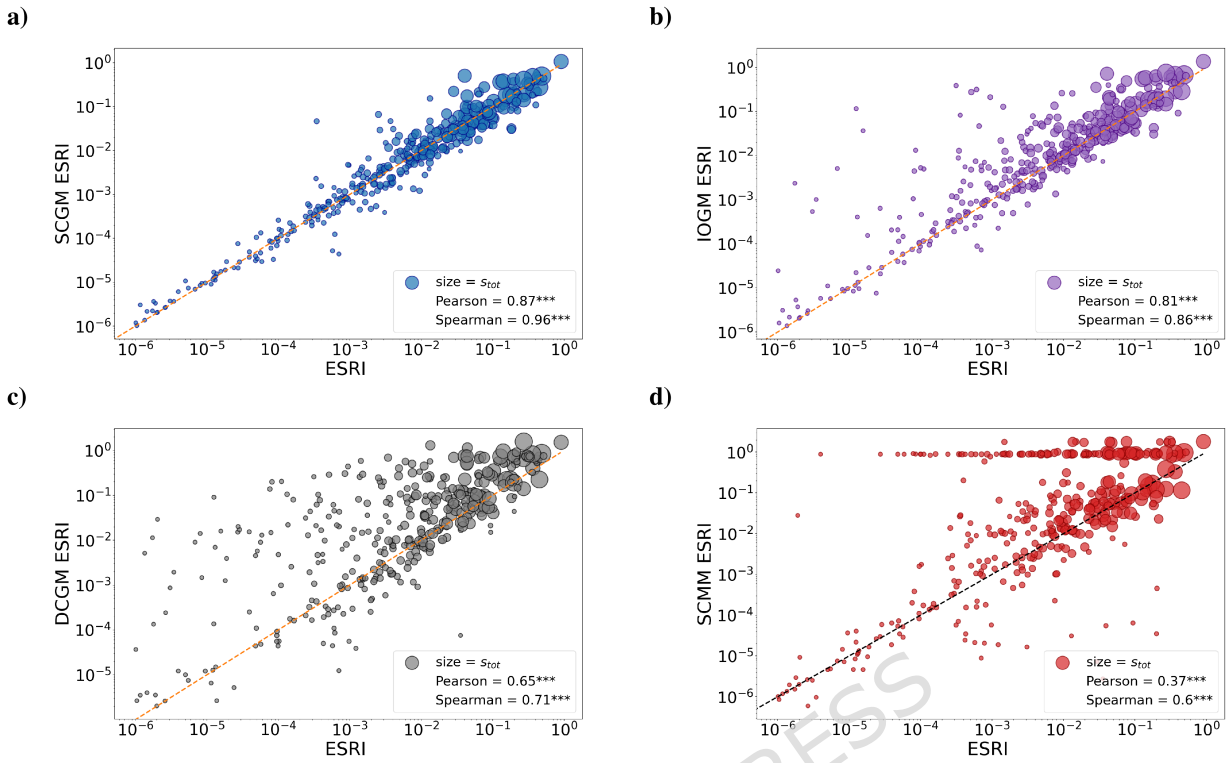

Figure 5. Reconstruction of sector-level systemic risk. Scatter plots of empirical versus model sector-level ESRI values, for SCGM (a), IOGM (b), DCGM (c) and SCMM (d).

We find that only the models that accurately reconstruct the empirical in-strength distribution across sectors (specifically SCGM and its approximated version IOGM) provide reliable estimates of firm-level ESRI. Notably, these models correctly identify most of the systemically important firms, demonstrating that they can serve as effective tools for systemic-risk assessment even when detailed topological information is unavailable. When ESRI values are aggregated at the sector level, SCGM distinctly outperforms the other models, closely followed by IOGM. This result underscores the importance of highly accurate micro-level reconstruction in order to preserve meaningful aggregate patterns. More broadly, it highlights the minimal amount of empirical information that must be disclosed to enable robust and policy-relevant reconstructions of national production networks.

We acknowledge that we validate our findings against the empirical production network of one specific country (Ecuador) on a single year (2008), which is also the year of the Global Financial Crisis. However, we remark that SCGM has been successfully used to reconstruct other production networks datasets, for the Netherlands in the original publication[43] and Hungary, after this work was finished[61]. Another potential issue is the presence of missing links in empirical data, which may arise from undeclared transactions, reporting failures, or sector-specific data limitations. While such omissions likely concern mainly small firms, they can bias the evaluation of reconstruction techniques, since models are calibrated and assessed on the empirical data and thus inherit its potential gaps: missing links may be mistaken for absent relationships, leading to more false positives. This underscores the crucial role of data quality.

Besides data limitations, it would also be useful to test reconstruction methods against alternative risk metrics, such as those based on link removal[62,63]. Nevertheless, comparing ESRI with the influence vector (proxied by Bonacich centrality) already provides useful insights. The influence vector is accurately recovered by all models, including the topology-agnostic SCMM. This stems from its linear formulation, which makes it well approximated by firms' out-strengths—quantities that are, by construction, reproduced by all models. By contrast, ESRI incorporates the nonlinearities arising from firms' production processes and can only be reproduced by network models that constrain the set of inputs required by firms. The key takeaway is that a realistic reconstruction of network topology becomes crucial for measuring nonlinear systemic risk, such as supply-chain contagion, whereas firm size alone may suffice to capture linear systemic effects.

Finally, we remark that SCGM achieves the highest reconstruction performance by exploiting more detailed information than its competitors, specifically the sector-level in-strengths (technical coefficients) of each firm. By contrast, IOGM approximates

these quantities using the technical coefficients of the corresponding sectors, while still achieving a satisfactory reconstruction. Thus, despite inheriting the well-known limitations of input–output tables[51,64,65], IOGM represents a viable and far more practical alternative to SCGM. Indeed, while firm-level technical coefficients are almost always undisclosed, input–output tables are routinely published in official statistics for most countries, making IOGM readily applicable in empirical settings.

Methods

Reconstruction models

The Stripe-Corrected Gravity Model (SCGM)

SCGM[43] works by extending DCGM recipe (see below) to reproduce the specific set of inputs each firm needs to produce its output. The model takes as input the total out-strength of each firm i, $s_i^{\text{out}} = \sum_j w_{i\to j}$, and the in-strength by sector (*stripe*) of each firm j, $s_{g_i\to j} = \sum_{k\in g_i} w_{k\to j}$. The latter quantity is the contribution to the in-strength of firm j coming from the sector g_i to which firm i belongs (i.e., how much j buys from suppliers belonging to sector g_i). SCGM employs this information to reconstruct the network, by assigning to the link $i \to j$ the weight

$$w_{i\to j}^{\text{SCGM}} = \frac{s_i^{\text{out}} s_{g_i\to j}}{W_{g_i}^{\text{out}}\, p_{i\to j}^{\text{SCGM}}}, \text{ with probability } p_{i\to j}^{\text{SCGM}} = \frac{z_{g_i} s_i^{\text{out}} s_{g_i\to j}}{1 + z_{g_i} s_i^{\text{out}} s_{g_i\to j}}; \tag{1}$$

naturally, $w_{i\to j}^{\text{SCGM}} = 0$ with probability $1 - p_{i\to j}^{\text{SCGM}}$. $W_{g_i}^{\text{out}} = \sum_{k\in g_i} s_k^{\text{out}} = \sum_j s_{g_i\to j}$ is the outgoing flux of sector g_i. The parameter z_{g_i} is determined by imposing

$$\langle L_{g_i}\rangle = \sum_{i\in g}\sum_{j(\neq i)} p_{i\to j}^{\text{SCGM}} = L_{g_i}^*, \tag{2}$$

i.e. that the expected number of links within sector g_i matches the empirical one. This formulation ensures that the out-strength and the in-strength by sector of each firm are preserved on average: $\langle s_i^{\text{out}}\rangle = s_i^{\text{out}}$ and $\langle s_{g_i\to j}\rangle = s_{g_i\to j}$, thus reproducing the productive structure of each firm (see Supplementary Materials S7). SCGM can be easily generalised to accommodate firms having 'multiple sector' outputs (see Supplementary Materials of [43] for further details).

The Input-Output Gravity Model (IOGM)

SCGM requires as input the in-strenght by sector of each firm. When this information is not available, it can be reformulated by employing input/output flows between sectors: the in-strength by sector of firm j is replaced by the expression

$$s_{g_i\to j}^{\text{IOGM}} = s_j^{\text{in}} \frac{s_{g_i\to g_j}}{W_{g_j}^{\text{in}}} = s_j^{\text{in}} \frac{s_{g_i\to g_j}}{\sum_{g_i} s_{g_i\to g_j}} \tag{3}$$

where $s_{g_i\to g_j} = \sum_{k\in g_i}\sum_{l\in g_j} w_{k\to l}$ is the total flux from sector g_i to sector g_j (how much g_j requires from g_i) and $W_{g_j}^{\text{in}} = \sum_{g_i} s_{g_i\to g_j} = \sum_{g_i}\sum_{k\in g_i}\sum_{l\in g_j} w_{k\to l}$ is the incoming flux of sector g_j (how much g_j requires from all other sectors). The expression above splits the in-strength of node j into sector-specific contributions, of which the i-th one is precisely the fraction of input provided by sector g_i. Then IOGM has the same functional form as SCGM:

$$w_{i\to j}^{\text{IOGM}} = \frac{s_i^{\text{out}} s_{g_i\to j}^{\text{IOGM}}}{W_{g_i}^{\text{out}}\, p_{ij}^{\text{IOGM}}}, \text{ with probability } p_{i\to j}^{\text{IOGM}} = \frac{z_{g_i} s_i^{\text{out}} s_{g_i\to j}^{\text{IOGM}}}{1 + z_{g_i} s_i^{\text{out}} s_{g_i\to j}^{\text{IOGM}}}; \tag{4}$$

naturally, $w_{i\to j}^{\text{IOGM}} = 0$ with probability $1 - p_{i\to j}^{\text{IOGM}}$. The parameter z_{g_i} is determined by imposing

$$\langle L_{g_i}\rangle = \sum_{i\in g}\sum_{j(\neq i)} p_{i\to j}^{\text{IOGM}} = L_{g_i}^*, \tag{5}$$

i.e. that the expected number of links within sector g_i matches the empirical one. The model preserves (on average) the out-strength and the in-strength of each firm and the total flux from sector g_i to sector g_j, i.e. $\langle s_i^{\text{out}}\rangle = s_i^{\text{out}}$, $\langle s_i^{\text{in}}\rangle = s_i^{\text{in}}$ and $\langle s_{g_i\to g_j}\rangle = s_{g_i\to g_j}$, but approximates the in-strength by sector of each firm, $s_{g_i\to j}$, with $s_{g_i\to j}^{\text{IOGM}}$ (see Supplementary Materials S7).

The Density-Corrected Gravity Model (DCGM)

DCGM[54] derives from the Directed Binary Configuration Model (DBCM), constraining the in- and out-degree sequences of the network. As degree information is often unavailable, DCGM relies on a *fitness ansatz*[66] that employs the strengths as parameters to set the values of the degrees[54,67,68]. DCGM is thus defined by imposing

$$w_{i\to j}^{\text{DCGM}} = \frac{s_i^{\text{out}} s_j^{\text{in}}}{W\, p_{i\to j}^{\text{DCGM}}}, \text{ with probability } p_{i\to j}^{\text{DCGM}} = \frac{z s_i^{\text{out}} s_j^{\text{in}}}{1 + z s_i^{\text{out}} s_j^{\text{in}}}; \tag{6}$$

naturally, $w^{\mathrm{DCGM}}_{i\to j}=0$ with probability $1-p^{\mathrm{DCGM}}_{i\to j}$. Notice that DCGM does not employ sector-specific quantities, as $s^{\mathrm{out}}_i=\sum_{j(\neq i)} w_{i\to j}$ is the out-strength of node i, $s^{\mathrm{in}}_j=\sum_{i(\neq j)} w_{i\to j}$ is the in-strength of node j and $W=\sum_i s^{\mathrm{out}}_i=\sum_i s^{\mathrm{in}}_i$ is the total weight of the network. The parameter z is determined by imposing

$$\langle L\rangle=\sum_i\sum_{j(\neq i)} p^{\mathrm{DCGM}}_{i\to j}=L^*. \tag{7}$$

i.e. that the expected number of links of the network matches the empirical one. The model preserves (on average) the out-strength and the in-strength of each firm, i.e. $\langle s^{\mathrm{out}}_i\rangle^{\mathrm{DCGM}}=s^{\mathrm{out}}_i$ and $\langle s^{\mathrm{in}}_i\rangle^{\mathrm{DCGM}}=s^{\mathrm{in}}_i$ (see[54] for further details).

The Stripe-Corrected MaxEnt Model (SCMM)

As last benchmark, we introduce SCMM, a topology-free model prescribing to pose

$$w^{\mathrm{SCMM}}_{i\to j}=\frac{s^{\mathrm{out}}_i s_{g_i\to j}}{W^{\mathrm{out}}_{g_i}}. \tag{8}$$

SCMM represents the deterministic version of SCGM, as it generates a single reconstructed network rather than an entire ensemble of reconstructed configurations. The model preserves the same quantities of SCGM but induces a topology which is unrealistically dense, since $w^{\mathrm{SCMM}}_{i\to j}$ equals zero only if firm i has no output or firm j does not receive any input from sector g_i.

Bonacich centrality

The influence vector, introduced in[13], measures the aggregate output elasticity to sector-level productivity shocks—i.e., the percentage change in total output resulting from a 1% productivity shock to a sector—in an input-output network. It is defined as

$$\vec{v}=\frac{\alpha}{N}\left[I-(1-\alpha)W\right]^{-1}\vec{1} \tag{9}$$

where W is the weighted adjacency matrix of the network[1], α is the labour share and N is the number of sectors in the network. The i-th component of the vector, v_i, represents the change in the network's output following a 1% shock to the productivity of sector i. We adapt this definition to our firm-level network, interpreting each components as the risk associated to a 1% shock to the corresponding firm. Note that the influence vector is mathematically equivalent to the Bonacich centrality[57,58], defined as

$$\vec{c}=\sum_{k=0}^{\infty}\beta^k W^k\vec{1}=\left[I-\beta W\right]^{-1}\vec{1} \tag{10}$$

where β is an attenuation parameter that rescales the contribution of increasingly longer paths on the propagation of the shock to a node[2]. Since our dataset does not include information on labour, we compute Bonacich centrality as a practical substitute of the influence vector.

The Economic Systemic Risk Index

The *Economic Systemic Risk Index* (ESRI) quantifies the systemic risk of a firm by evaluating the output reduction experienced by the whole production network (i.e. the sum of the out-strengths of all nodes) in case of its failure. More in detail, after removing firm i from the network, an upstream and a downstream shocks are propagated to any other firm $j(\neq i)$ through the two iterative equations:

$$x^d_i(t+1)=\min\left\{\min_{k\in\mathscr{I}_i^{\mathrm{Ess}}}\left\{\frac{1}{\alpha_{ik}}\sum_{j=1}^{n}W_{ji}h^d_j(t)\delta_{p_j,k}\right\},\beta_i+\frac{1}{\alpha_i}\sum_{k\in\mathscr{I}_i^{\mathrm{Not\ Ess}}}\sum_{j=1}^{n}W_{ji}h^d_j(t)\delta_{p_j,k}\right\} \tag{11}$$

$$x^u_i(t+1)=\sum_{j=1}^{n}W_{ij}h^u_j(t) \tag{12}$$

[1] In the original formulation, the influence vector is computed on the column-normalized adjacency matrix of the network W^{norm}, such that $\sum_i w^{norm}_{ij}=1$. Since, in this study, the reconstruction models are solved to replicate the full weighted adjacency matrix W, we use W rather than W^{norm} to fairly evaluate their reconstruction performance. Note that the same reconstruction performance would hold if we solved our models to replicate W^{norm} and computed Bonacich centrality on this matrix.

[2] In order for the Bonacich centrality to be computed on W, β has to be smaller than the inverse of the (absolute value of the) maximum eigenvalue of W, i.e. $\beta<1/\rho(W)$.

where t indicates the time step of the propagation, $x_i^d(t)$ is the out-strength of firm i at time t, following the propagation of the downstream shock and $x_i^u(t)$ is the out-strength of firm i at time t, following the propagation of the upstream shock, with $x_i^d(0) = x_i^u(0) = s_i^{\text{out}}$, $h_i^d(t) = x_i^d(t)/s_i^{\text{out}}$, $h_i^u(t) = x_i^u(t)/s_i^{\text{out}}$. The shock propagation is defined using a generalised Leontief production function: inputs from *essential* sectors (i.e. the sectors $k \in \mathscr{I}_i^{\text{Ess}}$) set a hard constraint on the output of firm i (enforced through the presence of the minimum) while the inputs from *non-essential* sectors (i.e. the sectors $k \in \mathscr{I}_i^{\text{Not Ess}}$) are treated in a linear way. The technical coefficients of the production function are calibrated on the empirical network: specifically, $\alpha_{ik} = \left(\sum_{j=1}^n W_{ji}\delta_{p_j,k}\right) / \left(\sum_{l=1}^n W_{il}\right)$ and $\alpha_i = \left(\sum_{j=1}^n W_{ji}\right) / \left(\sum_{l=1}^n W_{il}\right)$ while $\beta_i = \left(\sum_{k\in\mathscr{I}_i^{\text{Ess}}} \sum_{j=1}^n W_{ji}\delta_{p_j,k}\right) \left(\sum_{l=1}^n W_{il}\right) / \left(\sum_{j=1}^n W_{ji}\right)$ represents the fraction of output of firm i that can be produced with essential inputs only. The distinction between *essential* and *non-essential* inputs is derived from[69]. After the two, independent shocks have converged at $t = t^*$, the residual fraction of output of firm i is computed as $h_i(t^*) = \min\{h_i^d(t^*), h_i^u(t^*)\}$ and the ESRI value of firm i is given by

$$\text{ESRI}_i = \sum_{j(\neq)i} \frac{s_j^{\text{out}}}{\sum_k s_k^{\text{out}}} (1 - h_j(t^*)). \tag{13}$$

We refer to the original publication[23] for full information on the method.

Data availability

The VAT data used to construct the Ecuadorian production network were collected by the Internal Revenue Service (IRS) of Ecuador and provided to one of the authors by the Ecuadorian government. The authors do not have permission to make these data publicly available due to confidentiality restrictions. Researchers may contact the corresponding author for further information about possible access options; however, access to the full underlying data is restricted and subject to approval by the data owner(s). The reconstruction algorithms are available at the github repository https://github.com/mfessina/supchain_reconstruction.

Acknowledgements

DG acknowledges support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) and the Netherlands Organization for Scientific Research (NWO/OCW). The authors thank the Centro de Estudios Fiscales of Ecuador's Servicio de Rentas Internas (SRI), which provided the data for research purposes. We thank an anonymous reviewer for very valuable comments.

Funding

This work is supported by: the European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR), project 'SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics' - Grant IR0000013 (n. 3264, 28/12/2021); 'NetRes - Network analysis of economic and financial resilience', Italian DM n. 289, 25-03-2021 (PRO3 Scuole), CUP D67G22000130001 (https://netres.imtlucca.it); 'RENet - Reconstructing economic networks: from physics to machine learning and back', MUR PRIN 2022MTBB22 funded by European Union – Next Generation EU; 'C2T - From Crises to Theory: towards a science of resilience and recovery for economic and financial systems', MUR PRIN PNRR P2022E93B8 funded by European Union – Next Generation EU.

Author contributions statement

Study conception and design: MF, GC, TS, ST, DG. Data collection: MF, PE, ST. Data analysis: MF, GC. Discussion and interpretation of results: MF, GC, TS, PE, ST, DG. Draft manuscript preparation and revision: MF, GC, TS, PE, ST, DG.

Competing interests

The authors declare no competing interests.

References

1. Choi, T. Y. & Krause, D. R. The supply base and its complexity: Implications for transaction costs, risks, responsiveness, and innovation. *J. operations management* **24**, 637–652, DOI: https://doi.org/10.1016/j.jom.2005.07.002 (2006).

2. Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J. & Handfield, R. B. The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decis. Sci.* **38**, 131–156, DOI: https://doi.org/10.1111/j.1540-5915.2007.00151.x (2007).

3. Inoue, H. & Todo, Y. Firm-level propagation of shocks through supply-chain networks. *Nat. Sustain.* **2**, 841–847, DOI: https://doi.org/10.1038/s41893-019-0351-x (2019).

4. Guan, D. *et al.* Global supply-chain effects of covid-19 control measures. *Nat. Hum. Behav.* **4**, 577–587, DOI: https://doi.org/10.1038/s41562-020-0896-8 (2020).

5. Aldrighetti, R., Battini, D., Ivanov, D. & Zennaro, I. Costs of resilience and disruptions in supply chain network design models: A review and future research directions. *Int. J. Prod. Econ.* **235**, 108103, DOI: https://doi.org/10.1016/j.ijpe.2021.108103 (2021).

6. Carvalho, V. M., Nirei, M., Saito, Y. U. & Tahbaz-Salehi, A. Supply chain disruptions: Evidence from the great east japan earthquake*. *The Q. J. Econ.* **136**, 1255–1321, DOI: https://doi.org/10.1093/qje/qjaa044 (2021).

7. Ivanov, D., Tsipoulanidis, A. & Schönberger, J. *Supply Chain Risk Management and Resilience*, 485–520 (Springer International Publishing, Cham, 2021).

8. Chowdhury, P., Paul, S. K., Kaisar, S. & Moktadir, M. A. Covid-19 pandemic related supply chain studies: A systematic review. *Transp. Res. Part E: Logist. Transp. Rev.* **148**, 102271, DOI: https://doi.org/10.1016/j.tre.2021.102271 (2021).

9. Pichler, A., Pangallo, M., del Rio-Chanona, R. M., Lafond, F. & Farmer, J. D. Production networks and epidemic spreading: How to restart the uk economy? *arXiv preprint arXiv:2005.10585* DOI: https://doi.org/10.48550/arXiv.2005.10585 (2020).

10. Pichler, A. & Farmer, J. D. Simultaneous supply and demand constraints in input–output networks: the case of covid-19 in germany, italy, and spain. *Econ. Syst. Res.* **34**, 273–293, DOI: https://doi.org/10.1080/09535314.2021.1926934 (2022).

11. Miller, R. E. & Blair, P. D. *Input-output analysis: foundations and extensions* (Cambridge university press, 2009).

12. Leontief, W. *Input-output economics* (Oxford University Press, 1986).

13. Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. The network origins of aggregate fluctuations. *Econometrica* **80**, 1977–2016, DOI: https://doi.org/10.3982/ECTA9623 (2012).

14. Carvalho, V. M. & Tahbaz-Salehi, A. Production networks: A primer. *Annu. Rev. Econ.* **11**, 635–663, DOI: https://doi.org/10.1146/annurev-economics-080218-030212 (2019).

15. Gabaix, X. The granular origins of aggregate fluctuations. *Econometrica* **79**, 733–772, DOI: https://doi.org/10.3982/ECTA8769 (2011).

16. Diem, C., Borsos, A., Reisch, T., Kertész, J. & Thurner, S. Estimating the loss of economic predictability from aggregating firm-level production networks. *PNAS nexus* **3**, pgae064, DOI: https://doi.org/10.1093/pnasnexus/pgae064 (2024).

17. Dolgui, A., Ivanov, D. & Sokolov, B. Ripple effect in the supply chain: an analysis and recent literature. *Int. journal production research* **56**, 414–430, DOI: https://doi.org/10.1080/00207543.2017.1387680 (2018).

18. Ivanov, D., Sokolov, B. & Dolgui, A. The ripple effect in supply chains: trade-off 'efficiency-flexibility-resilience'in disruption management. *Int. J. Prod. Res.* **52**, 2154–2172, DOI: https://doi.org/10.1080/00207543.2013.858836 (2014).

19. Ivanov, D. Simulation-based ripple effect modelling in the supply chain. *Int. J. Prod. Res.* **55**, 2083–2101, DOI: https://doi.org/10.1080/00207543.2016.1275873 (2017).

20. Choi, T. Y., Dooley, K. J. & Rungtusanatham, M. Supply networks and complex adaptive systems: control versus emergence. *J. operations management* **19**, 351–366, DOI: https://doi.org/10.1016/S0272-6963(00)00068-1 (2001).

21. Yan, T., Choi, T. Y., Kim, Y. & Yang, Y. A theory of the nexus supplier: A critical supplier from a network perspective. *J. Supply Chain Manag.* **51**, 52–66, DOI: https://doi.org/10.1111/jscm.12070 (2015).

22. Shao, B. B., Shi, Z. M., Choi, T. Y. & Chae, S. A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index. *Decis. Support. Syst.* **114**, 37–48, DOI: https://doi.org/10.1016/j.dss.2018.08.008 (2018).

23. Diem, C., Borsos, A., Reisch, T., Kertész, J. & Thurner, S. Quantifying firm-level economic systemic risk from nation-wide supply networks. *Sci. reports* **12**, 1–13, DOI: https://doi.org/10.1038/s41598-022-11522-z (2022).

24. Bacilieri, A., Borsos, A., Astudillo-Estevez, P. & Lafond, F. Firm-level production networks: what do we (really) know? Tech. Rep. 2023-08, INET Oxford Working Paper (2023).

25. Chakraborty, A. & Ikeda, Y. Testing "efficient supply chain propositions"using topological characterization of the global supply chain network. *PLOS ONE* **15**, e0239669, DOI: https://doi.org/10.1371/journal.pone.0239669 (2020).

26. Chakraborty, A., Reisch, T., Diem, C., Astudillo-Estévez, P. & Thurner, S. Inequality in economic shock exposures across the global firm-level supply network. *Nat. Commun.* **15**, 3348, DOI: https://doi.org/10.1038/s41467-024-46126-w (2024).

27. Atalay, E., Hortaçsu, A., Roberts, J. & Syverson, C. Network structure of production. *Proc. Natl. Acad. Sci.* **108**, 5199–5202, DOI: https://doi.org/10.1073/pnas.1015564108 (2011).

28. Cohen, L. & Frazzini, A. Economic links and predictable returns. *The J. Finance* **63**, 1977–2011, DOI: https://doi.org/10.1111/j.1540-6261.2008.01379.x (2008).

29. Barrot, J.-N. & Sauvagnat, J. Input specificity and the propagation of idiosyncratic shocks in production networks. *The Q. J. Econ.* **131**, 1543–1592, DOI: https://doi.org/10.1093/qje/qjw018 (2016).

30. König, M. D., Levchenko, A., Rogers, T. & Zilibotti, F. Aggregate fluctuations in adaptive production networks. *Proc. Natl. Acad. Sci.* **119**, e2203730119, DOI: https://doi.org/10.1073/pnas.2203730119 (2022).

31. Saito, Y. U., Watanabe, T. & Iwamura, M. Do larger firms have more interfirm relationships? *Phys. A: Stat. Mech. its Appl.* **383**, 158–163, DOI: https://doi.org/10.1016/j.physa.2007.04.097 (2007).

32. Ohnishi, T., Takayasu, H. & Takayasu, M. Hubs and authorities on japanese inter-firm network: Characterization of nodes in very large directed networks. *Prog. Theor. Phys. Suppl.* **179**, 157–166, DOI: https://doi.org/10.1143/PTPS.179.157 (2009).

33. Ohnishi, T., Takayasu, H. & Takayasu, M. Network motifs in an inter-firm network. *J. Econ. Interact. Coord.* **5**, 171–180, DOI: https://doi.org/10.1007/s11403-010-0066-6 (2010).

34. Fujiwara, Y. & Aoyama, H. Large-scale structure of a nation-wide production network. *The Eur. Phys. J. B* **77**, 565–580, DOI: https://doi.org/10.1140/epjb/e2010-00275-2 (2010).

35. Brintrup, A., Barros, J. & Tiwari, A. The nested structure of emergent supply networks. *IEEE Syst. J.* **12**, 1803–1812, DOI: https://doi.org/10.1109/JSYST.2015.2493345 (2018).

36. Fessina, M., Zaccaria, A., Cimini, G. & Squartini, T. Pattern-detection in the global automotive industry: a manufacturer-supplier-product network analysis. *Chaos, Solitons & Fractals* **181**, 114630, DOI: https://doi.org/10.1016/j.chaos.2024.114630 (2024).

37. Dhyne, E., Magerman, G. & Rubínová, S. The belgian production network 2002-2012. Tech. Rep. 288, National Bank of Belgium Working Paper (2015).

38. Mungo, L., Lafond, F., Astudillo-Estévez, P. & Farmer, J. D. Reconstructing production networks using machine learning. *J. Econ. Dyn. Control.* **148**, 104607, DOI: https://doi.org/10.1016/j.jedc.2023.104607 (2023).

39. Peydró, J., Jiménez, G., Huremovic, K., Moral-Benito, E. & Vega-Redondo, F. Production and financial networks in interplay: Crisis evidence from supplier-customer and credit registers. Tech. Rep. 15277, CEPR Press Discussion Paper (2020).

40. Silva, T. C., Amancio, D. R. & Tabak, B. M. Modeling supply-chain networks with firm-to-firm wire transfers. *Expert. Syst. with Appl.* **190**, 116162, DOI: https://doi.org/10.1016/j.eswa.2021.116162 (2022).

41. Tamura, K. *et al.* Estimation of flux between interacting nodes on huge inter-firm networks. In *International Journal of Modern Physics: Conference Series*, vol. 16, 93–104, DOI: https://doi.org/10.1142/S2010194512007805 (World Scientific, 2012).

42. Fujiwara, Y. *et al.* Money flow network among firms'accounts in a regional bank of japan. *EPJ Data Sci.* **10**, 19, DOI: 10.1140/epjds/s13688-021-00274-x (2021).

43. Ialongo, L. N. *et al.* Reconstructing firm-level interactions in the dutch input–output network from production constraints. *Sci. Reports* **12**, 1–12, DOI: https://doi.org/10.1038/s41598-022-13996-3 (2022).

44. Mungo, L., Brintrup, A., Garlaschelli, D. & Lafond, F. Reconstructing supply networks. *J. Physics: Complex.* **5**, 012001, DOI: https://doi.org/10.1088/2632-072X/ad30bf (2024).

45. Brintrup, A. *et al.* Predicting hidden links in supply networks. *Complexity* **2018**, DOI: https://doi.org/10.1155/2018/9104387 (2018).

46. Kosasih, E. E. & Brintrup, A. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *Int. J. Prod. Res.* 1–14, DOI: https://doi.org/10.1080/00207543.2021.1956697 (2021).

47. Brockmann, N., Elson Kosasih, E. & Brintrup, A. Supply chain link prediction on uncertain knowledge graph. *ACM SIGKDD Explor. Newsl.* **24**, 124–130, DOI: https://doi.org/10.1145/3575637.3575655 (2022).

48. Mungo, L. & Moran, J. Revealing production networks from firm growth dynamics. *arXiv preprint arXiv:2302.09906* DOI: https://doi.org/10.48550/arXiv.2302.09906 (2023).

49. Reisch, T., Heiler, G., Diem, C., Klimek, P. & Thurner, S. Monitoring supply networks from mobile phone data for estimating the systemic risk of an economy. *Sci. Reports* **12**, 1–10, DOI: https://doi.org/10.1038/s41598-022-13104-5 (2022).

50. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620, DOI: https://doi.org/10.1103/PhysRev.106.620 (1957).

51. Bacilieri, A. & Austudillo-Estevez, P. Reconstructing firm-level input-output networks from partial information. *arXiv preprint arXiv:2304.00081* DOI: https://doi.org/10.48550/arXiv.2304.00081 (2023).

52. Cimini, G. *et al.* The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71, DOI: https://doi.org/10.1038/s42254-018-0002-6 (2019).

53. Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A. & Garlaschelli, D. Reconstruction methods for networks: The case of economic and financial systems. *Phys. reports* **757**, 1–47, DOI: https://doi.org/10.1016/j.physrep.2018.06.008 (2018).

54. Cimini, G., Squartini, T., Garlaschelli, D. & Gabrielli, A. Systemic risk analysis on reconstructed economic and financial networks. *Sci. Reports* **5**, DOI: https://doi.org/10.1038/srep15758 (2015).

55. Astudillo Estevez, P. *Towards a post-oil economy: a complexity approach to understanding natural resource dependency and economic diversification in Ecuador*. Ph.D. thesis, University of Oxford (2021).

56. Mattsson, C. E. *et al.* Functional structure in production networks. *Front. Big Data* **4**, 666712, DOI: https://doi.org/10.3389/fdata.2021.666712 (2021).

57. Bonacich, P. Power and centrality: A family of measures. *Am. journal sociology* **92**, 1170–1182, DOI: 10.1086/228631 (1987).

58. Bonacich, P. Some unique properties of eigenvector centrality. *Soc. networks* **29**, 555–564, DOI: 10.1016/j.socnet.2007.04.002 (2007).

59. Gabrielli, A., Macchiati, V. & Garlaschelli, D. Critical density for network reconstruction. In *From Computational Logic to Computational Biology: Essays Dedicated to Alfredo Ferro to Celebrate His Scientific Career*, 223–249, DOI: https://doi.org/10.1007/978-3-031-55248-9_11 (Springer, 2024).

60. Wichmann, P., Brintrup, A., Baker, S., Woodall, P. & McFarlane, D. Extracting supply chain maps from news articles using deep neural networks. *Int. J. Prod. Res.* **58**, 5320–5336, DOI: https://doi.org/10.1080/00207543.2020.1720925 (2020).

61. Mancini, A., Lengyel, B., Di Clemente, R. & Cimini, G. Evolution and determinants of firm-level systemic risk in local production networks. *arXiv preprint arXiv:2506.21426* DOI: https://doi.org/10.48550/arXiv.2506.21426 (2025).

62. Amico, A., Vaccario, G. & Schweitzer, F. Efficiency and resilience: key drivers of distribution network growth. *EPJ Data Sci.* **13**, 52, DOI: 10.1140/epjds/s13688-024-00484-z (2024).

63. Ivanov, D., Sokolov, B. & Dolgui, A. The ripple effect in supply chains: trade-off 'efficiency-flexibility-resilience' in disruption management. *Int. J. Prod. Res.* **52**, 2154–2172, DOI: 10.1080/00207543.2013.858836 (2014).

64. Colon, C. & Poledna, S. Constructing supply-chain input–output tables: Mapping distribution networks from trade margins. *SSRN preprint 5796664* DOI: https://ssrn.com/abstract=5796664 (2025).

65. Katafuchi, Y. *et al.* Construction of an enterprise-level global supply chain database. *Nat. Commun.* **16**, 11158, DOI: 10.1038/s41467-025-66083-2 (2025).

66. Caldarelli, G., Capocci, A., De Los Rios, P. & Munoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702, DOI: https://doi.org/10.1103/PhysRevLett.89.258702 (2002).

67. Garlaschelli, D. & Loffredo, M. I. Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.* **93**, 188701, DOI: https://doi.org/10.1103/PhysRevLett.93.188701 (2004).

68. Cimini, G., Squartini, T., Gabrielli, A. & Garlaschelli, D. Estimating topological properties of weighted networks from limited information. *Phys. Rev. E* **92**, 040802, DOI: https://doi.org/10.1103/PhysRevE.92.040802 (2015).

69. Pichler, A., Pangallo, M., del Rio-Chanona, R. M., Lafond, F. & Farmer, J. D. In and out of lockdown: Propagation of supply and demand shocks in a dynamic input-output model. *arXiv preprint arXiv:2102.09608* DOI: https://doi.org/10.48550/arXiv.2102.09608 (2021).