PNAS Nexus, 2025, **4**, pgaf102
https://doi.org/10.1093/pnasnexus/pgaf102
Advance access publication 26 March 2025
Research Report

Synergistic small worlds that drive technological sophistication

Hardik Rajpal[a,b,c,*] and Omar Guerrero[c]

[a]Centre for Complexity Sciences, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom
[b]Department of Mathematics, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom
[c]Public Policy Programme, The Alan Turing Institute, Euston Road, London NW1 2DB, United Kingdom
*To whom correspondence should be addressed: Email: h.rajpal15@imperial.ac.uk
Edited By Esteban Moro

Abstract

It is a well-known fact that economic growth goes hand in hand with improvements in technological sophistication. While critical to such sophistication, the nature and underlying structure of the input interactions taking place inside production processes remain opaque, at least in the study of large systems such as industries and entire economies. We develop a method to quantify the degree of input complementarity in production processes form input–output data. We propose that the information-theoretic concept of synergistic information is analog to economic complementarity and exploit this link to create a data-driven approach that does not require the ex ante assumption of production functions. In contrast to alternative empirical approaches, our method is able identify input–input interactions and to quantify their contribution to output, revealing an input–input synergistic interaction network that characterizes an industry's productive technology. We find that more sophisticated industries tend to exhibit highly modular small-world topologies; with the tertiary sector as its central connective core. Overall, countries and industries that have a well-established connective core and specialized modules exhibit higher economic complexity, higher output, and lower emissions. The proposed method provides a framework to identify key relationships in the economy that can enhance economic performance.

Keywords: economic complexity, technology, synergy, information, small-world networks

Significance Statement

The measurement of technological sophistication is critical for understanding industrialization and economic development. We develop an information-theoretic method to estimate an industry's level of technological sophistication by calculating the degree of synergy or complementarity among its inputs. Out method does not rely on ex ante assumptions on how inputs interact (like production functions) and accounts for input–input and input–output (IO) interactions. Using IO datasets, we reveal the productive structure of industries as networks of synergistic interactions. Their topology highlights the role of modular small-world architecture—similar to the human brain—in supporting technological sophistication in advanced economies. These results suggest an underlying universal architecture that is beneficial for integrating information in systems with high level of differentiation or specialization.

Introduction

The nature of production technology, understood as the way in which inputs interact to generate output, is central to the inner works of any economy. The ability of a technology to combine inputs to generate new output characterizes its degree of sophistication. Moreover, it is well known that economic growth often goes hand-in-hand with improvements in the sophistication of production technologies.[a] Therefore, quantifying the nature of the interaction among the inputs of a production process is key to identifying the drivers of technological sophistication.

The concept of input complementarity is particularly salient, as it describes the ability to take advantage of exiting inputs to produce new output. For example, in regional-development studies, where geographical spillovers and industrial diversity are key for economic growth (1, 2), it is known that effective economic development hinges in exploiting such spillovers in ways that make them complementary in production processes (3).[b] More generally, inferring complementarities in production processes is an important endeavor in various disciplines. Nevertheless, standard practices often involve proxy measures that only take output data into account (missing their relationship to inputs), or assume relationships between inputs through

Competing Interest: The authors declare no competing interests.
Received: January 16, 2024. **Accepted:** January 10, 2025

production functions (ex ante assuming the nature of input interactions).

In information theory, there exists a measure of interdependency that maps into the economic notion of input complementarity: *synergy*. Hence, taking advantage of this measure, we develop an information-theory-based approach to quantify input complementarities and infer their networked structure. In contrast with existing approaches, our framework does not assume production functions ex ante (common practice in economic modeling), and it accounts for the information of both input and output data. Interestingly, a parallelism between the concepts of "synergy" and "complementarity" have been previously established in the study of complementarities within organizations: "Complementarity, as we use the term, is a near synonym for synergy" (4, p.1).[c]

In essence, the proposed method measures the contribution of input interactions to the output of an industry, firm, or region. To the best of our knowledge, this is among the first approaches to infer the nature of technology at such granular level, without assuming functional forms, and accounting for both input–input and input–output (IO) relationships. By enabling the inference of pairwise complementarities, our method allows for the construction of synergy networks underlying different industries and regions, and analyze the topological features that characterize the structural principles of technological sophistication.

Our results suggest that more sophisticated industries (e.g. those in more developed countries or those that exhibit lower emission levels) exhibit small-world topologies. However, they are a special class of small-world networks: modular graphs that use tertiary-sector inputs as their connective core. This outcome is consistent with the literature on the central role of the service sector in economic and technological development (5, 6). In these synergy graphs, modularity appears to be crucial for technological sophistication alongside classic properties of small-world networks that enable global integration as well as functional segregation. Interestingly, these properties seem to be ubiquitous as they can be found in a wide class of complex systems (7), including social (8), technological (9), political (10), managerial (11), and biological (12) systems. Our results suggest that sophisticated production processes leverage the ability of small world networks to address the challenge of balancing specialization in a diverse set of inputs vs. integrating them together to generate new output. Thus, the proposed framework is a step towards opening the black box of technological sophistication and sheds new light on the principles behind input complementarity. In the rest of the article, we discuss the existing knowledge gap, describe the methods and datasets used, and present our results.

Knowledge gap

Our work contributes to filling gaps in various strands of literature that concern the quantification of input complementarity. Here, we review some ideas prevalent in the literature and highlight the limitations that our work aims to overcome.

First, in the literature of relatedness, it is assumed that products, services, or skills that complement each other can foster learning, innovation, and spillovers, making more of the product space accessible (13). In certain applications, relatedness measures have been extended to products (14), skills (15), patents (16), and social capabilities (17). Such metrics are derived by identifying the co-occurrence of two inputs. In essence, if two inputs (products, skills, patents, or social capabilities) exist together in a region/industry/firm with a frequency higher than expected at random, they are deemed complementary. While these metrics "can indirectly measure the technological sophistication of [a country's] production fabric." (18), their interpretation about the interaction between inputs relies on the strong assumption that, if two inputs co-occur statistically significantly, then such co-occurrence necessarily happens because of their complementary nature, without actually differentiating complementarity from other types of interactions such as redundancy (substitutability).[d] Therefore, relatedness looks at complementarity from the point of view of occurrence rather than from a perspective of production; and production necessarily entails explicitly linking inputs to output, something missing in co-occurrence methods.

Second, in the IO literature, analysts model production networks that assume, ex ante, the nature of the input–input and IO interactions through production functions (19, 20). This approach has the benefit of shifting the problem of modeling technological sophistication from one of inferring structures to one of fitting parameters. This shift comes with problems such as misspecification (21), estimation biases (22), and aggregation artifacts (23, 24). Thus, assuming (parametric) production functions instead of inferring interaction structures limits our capacity to quantify technological sophistication.[e]

Third, in systems engineering the concept of design structure matrices (DSMs) is a popular approach (25, 26). DSMs describe networks of interactions between the different components of a production processes. They provide a comprehensive tool to represent technological sophistication, but they require substantial knowledge about the process itself. Thus, DSMs build on a top-down approach that demands substantial ex ante knowledge; making it difficult to scale up to more aggregate levels (e.g. industries and regions) that allow a more general understanding of the nature of production technologies.

Framework and data

Next, allow us to provide a succinct description of the methodology and the data employed in our analysis, while leaving more specific details for Methods section. Broadly speaking, our method seeks to estimate the amount of information pairs of inputs provide about the output of a given industry (from IO tables). We decompose this contribution to the output into different information-sharing modes. We focus on a particular mode known as *synergistic information*, i.e. the information that cannot be obtained from any of the inputs alone as it exists only due to their interaction. This is the core idea behind the economic notion of complementarity. Hence, using this type of information, we produce a *synergy score* capturing how complementary are certain inputs in a production process.

A higher synergy score means that their complementary interactions provide more information about the target than what is expected by a substitutable combination of inputs. The presence of complementary interactions among inputs, in turn, relates to the degree of sophistication of the production process. This interpretation aligns with the established notion of enhanced learning and innovation afforded by complementary spillovers amongst different inputs. Thus, our synergy score provides a measure of technological sophistication that explicitly quantifies the complementary nature of interactions among inputs and outputs.

The synergy score

Consider an industry and three associated datasets. Data Y contain the growth rates of output of the industry, while X_1 and X_2 capture the growth rates of inputs 1 and 2, respectively. We are interested in quantifying how much information X_1 and X_2 provide about Y. This information can be quantified using the total mutual information $I(X_1, X_2; Y)$ between the output and the input growth rates. In a nutshell, the total mutual information is a measure of the amount of uncertainty in Y, that is reduced by knowing X_1 and X_2. Furthermore, it is possible to estimate how much of this uncertainty reduction is a result of the interactions between the inputs. We can achieve this by following the partial information decomposition proposed by Ref. (27), where the total mutual information I provided by X_1 and X_2 about Y can be decomposed as

$$I(X_1, X_2; Y) = Syn(X_1, X_2; Y) + Red(X_1, X_2; Y) + Unq(X_1; Y) + Unq(X_2; Y). \quad (1)$$

In Eq. 1, the mutual information is decomposed into synergistic, redundant, and unique contributions (we explain the nonsynergistic parts in Methods section). Synergistic information can only be obtained from the interaction between X_1 and X_2. If either input is removed from the production process, all the synergistic information would be lost from the output signal.

Let us refine our explanation by providing an example with logic gates, which map well to the different technologies of a production process. Consider two independent binary inputs into a logic gate X_1 and X_2. If the logic gate is an *XOR* gate, then it gives the output of 1 if and only if exactly one of the inputs is 1. Here, both inputs are necessary to predict the output exactly, and each input cannot predict the output by itself. Therefore, this gate can be considered as a model of unsubstitutable inputs in a production process. Now, an OR gate provides an output of 1 even if any one of the inputs is 1. Here, the output is partially predictable if any one of the inputs is known. Suggesting the presence of redundancy in the system. However, to predict the output exactly, we still need both inputs; therefore, some synergy exists as well. Therefore, an OR gate maps well to a production process with partially substitutable inputs. Finally, if we consider a special case of the OR gate with identical inputs, we refer to this variant as the OR* gate. This system is completely redundant, and the output is exactly predictable even if only one input is known, thereby capturing a production process with completely substitutable inputs. While this example uses binary variables, these information-theoretic measures have been extensively used in neuroscience on continuous neural time-series signals.[f] We present the synergy, redundancy, and the total mutual information scores of this example, calculated using the PID framework, in Table 1.

More sophisticated production processes involve more synergistic information because they generate more output by recombining the same inputs in complementary ways. Thus, we use this type of information as a synergy score and compute it for all unique pairs of inputs in the data.[g] Further details on the estimation method can be found in Estimating the synergy score section.

Using the pairwise synergy score for the inputs, we infer a weighted undirected network of the synergistic input interactions. These synergy networks capture the structure of the technology underpinning a production process. In contrast to the popular approach of assuming production functions in IO models, our synergy networks are inferred from the data. Thus, they provide a "model-free" approach to quantify the degree of technological sophistication and structure of interactions in the production process. Analyzing the topological properties of a large cross-section of synergy networks reveals key features underpinning technological sophistication.

Table 1. Synergistic and redundant information about the output provided by binary input variables into three cases of logic gates.

Gate	$I(X_1, X_2; Y)$	$Syn(X_1, X_2; Y)$	$Red(X_1, X_2; Y)$
XOR	1 bit	1 bit	0 bit
OR	0.8 bits	0.5 bits	0.3 bits
OR*	1 bit	0 bit	1 bit

Data

Our empirical application focuses on industrial technology through IO data, and places especial emphasis on the analysis of technologies with high heterogeneity in terms of their sophistication. To empirically capture such diversity, it is necessary to assemble a dataset with substantial variation in terms of economic development. Thus, having a large number of countries is key, as most technological variation would be expected between nations with different levels of development. We find such coverage in the Eora26 dataset (28), a global collection of IO tables with harmonized industries across a large number of countries. The subset extracted from Eora26 contains annual IO tables for 26 industries across 148 countries during the period 1995–2020. To test the robustness of the analysis, we replicate our findings on the Inter-Country Input Output dataset provided by the Organisation for Economic Co-operation and Development (OECD) (29). Although the OECD dataset has lower variation in levels of development (76 countries), it provides a more granular description of inputs over 45 industries. We use the time series of each input growth rates and the corresponding total output growth rates to compute the synergy scores and networks associated with a particular industry in a given group of countries. Thus, we exploit the temporal variation in the data to estimate the joint probability distribution of growth rates. Then, we compare country-groups and industries.[h] Further information on the data and its processing can be found in Methods section.

Validation

To validate our synergy score, we verify its alignment with well-accepted proxies of economic sophistication (30–32). The economic complexity literature offers different popular indices in this regard: the Economic Fitness Index (EFI) (33), the Economic Complexity Index (ECI) (14), and a generalization of the previous two, the GENEPY index (34).[i]

Typically these indices are estimated on a national level using international trade datasets. However, recently, they have been calculated at the industry level (35, 36).[j] Note that, to compute these validation indices, we employ data that are independent from the aforementioned IO tables: the BACI international trade dataset (37).

Before presenting our validation, let us elaborate on why, on a theoretical level, the EFI and the ECI are relevant measures for this exercise. The economics literature has long argued that development takes place when an economy diversifies the nature of its activities while exploiting their interlinkages and capabilities (38–40). More recently, the concept of relatedness has become the theoretical backbone of investigations into the mechanisms behind economic diversification in regional studies. According to Ref. (41), relatedness is conceptually viewed through the lens of "similarity" and "complementarity." Furthermore, Ref. (41) relatedness argues that the similarity approach has become dominant, while complementarity has rarely been investigated.

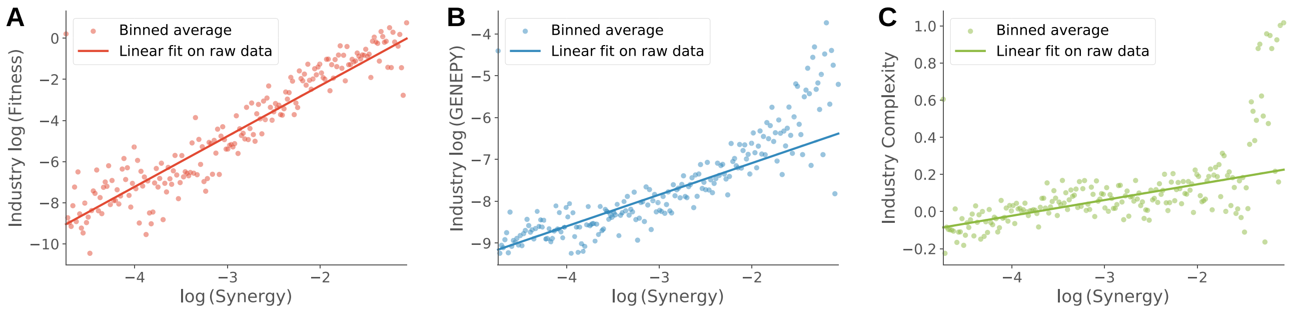

Fig. 1. Synergy score predicting export-based industry sophistication indices. With the purpose of a clear visualization, we display data binned according to the synergy scores. Each dot corresponds to the average value within the corresponding bin. Since the synergy distribution is right-skewed, we analyze the logarithm of the scores (on x-axis). The economic fitness index (A) and the GENEPY index (B) are also log transformed on y-axis. The economic complexity index (C) shows a bounded distribution and hence is not transformed. The linear fit is estimated using the entire data (not the binned one).

Table 2. Linear models predicting the economic fitness index of industries.

Predictor	EORA 1	EORA 2	EORA 3	EORA 4	OECD 1	OECD 2	OECD 3	OECD 4
Log synergy	2.4570**	2.3528**	2.6970***	2.1796***	4.3811***	4.2163***	4.4180***	4.5699***
	(1.0484)	(1.0053)	(1.0254)	(0.7582)	(1.2116)	(1.1707)	(1.1405)	(1.1588)
Log GDP per. cap.		0.6523	0.6403	−0.4684		0.6275	0.6192	0.3155
		(0.5745)	(0.5685)	(0.5618)		(0.5588)	(0.5560)	(0.5490)
Primary sector			−5.8356**	−7.9426**			1.4040	0.4751
			(2.7029)	(3.2428)			(1.8543)	(2.0307)
Secondary sector			−1.4582	−5.7442**			−2.5265	−3.4749
			(1.1045)	(2.3370)			(2.5887)	(2.5931)
Tertiary sector			−6.3603***	−8.9321***			1.0486	0.1175
			(2.0858)	(2.6226)			(1.8531)	(1.8746)
Log output				1.4227**				0.8703**
				(0.6162)				(0.3507)
Adjusted R^2	0.0237	0.0286	0.0454	0.0891	0.0297	0.0332	0.0406	0.0598
No. of observations	695,500	695,500	695,500	695,500	1,550,494	1,550,494	1,550,494	1,550,494

OLS regression coefficients. The model intercept is omitted. The dependent variable is the EFI, calculated for each industry in the dataset, and averaged across the years in the sample period. The stars denote the level of significance. The number in parenthesis is the clustered standard error. Confidence levels are indicated by * for 90%, ** for 95%, and *** for 99%.

As we have previously explained, a key issue with empirically measuring complementarity is that one needs to account for both inputs and outputs in the estimation. For instance, (42, 43) argue that complementarity is inherently a super-additive feature, which evidently involves being able to link output to the interaction of inputs. By construction, co-occurrence-based empirical approaches that are popular in similarity studies (such as the EFI and the ECI) do not account for input interactions, so complementarities cannot be properly estimated.[k] Our information-theory approach, in contrast, provides a formalism that explicitly links input interaction with output signals, enabling us to estimate complementarity through the lens of synergistic information. Hence, since both the similarity and complementarity routes represent alternative approaches to inferring the degree of economic sophistication, we expect that they correlate when inferred from data about the same system (economies).

Figure 1 validates our synergy score by showing a positive correlation with the three complexity-based indices. It can be seen that synergistic input interactions predict these proxies of technological sophistication. The measure is able to capture the cross-country and cross-industry variation produced by these indices, especially when these two distinct approaches use independent datasets. To enable this validation, we aggregated the synergy scores from pairwise measures to industry-level ones. In the Results section, we focus on the nuanced information provided by the previously unobserved network structure of synergistic interactions.

Table 2 provides statistical tests of this validation and shows that it is robust across the EORA and the OECD datasets. We perform these tests via regression analysis and controlling for the usual covariates. Our variable of interest is the logarithm of the synergy score. In the first model, we estimate the association between the synergy score and the EFI without any controls. Model 2 includes log GDP per capita, and aims at controlling for country-specific factors such as higher income, better public governance, and better infrastructure. Model 3 adds dummy variables indicating if the industry belongs to the primary, secondary, or tertiary sector (with the additional sector being "Other"). With this, we try to control for sector-specific factors like regulatory frameworks (e.g. tax prerogatives and unionization practices). In model 4, we include industry-specific total output.

The Supplementary information provides further validation results by using additional energy-related covariates that are available in the EORA dataset. We replicate this validation with the other complexity-based indices, different levels of aggregation, and alternative ways of computing the synergy score.

Results

One of the main advantages of our approach is the explicit calculation of synergy scores for all pairwise input interactions. This

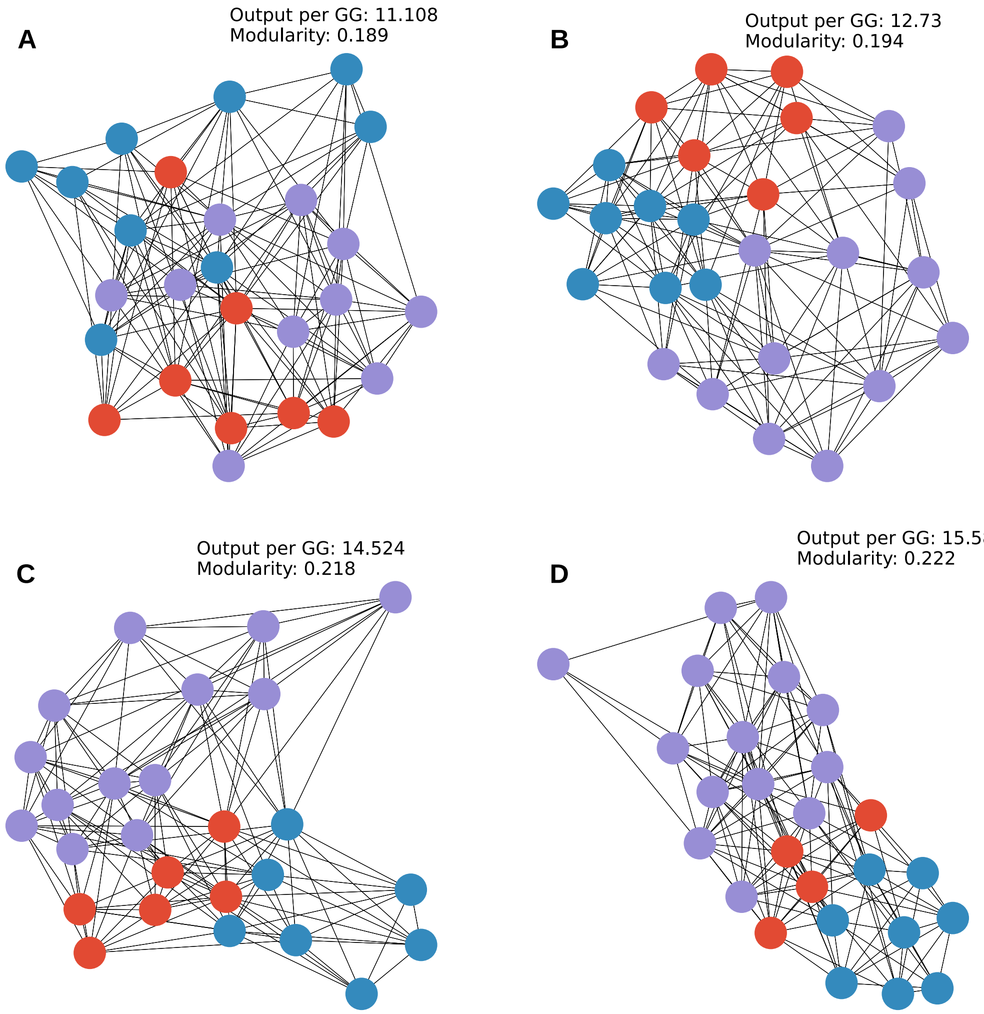

Fig. 2. Synergy networks of the transport industry. The node colors represent community labels. GG, greenhouse gas. The figures are aranged in the order of increasing output per GG, from A) to D). It can be seen that the networks become more modular and structured with increasing output per GG.

means that it is possible to construct an undirected network with weights determined by the synergy scores. These synergy networks capture the structure of the technology underpinning a production process. The details on how to construct synergy networks can be found in Estimating the synergy score section.

First, let us provide an example of one such network and how its structure differs across countries with different levels of technological sophistication. Figure 2 presents the case of the transport industry. Arguably, technologies with a lower rate of greenhouse gas (GG) emissions per unit output are more sophisticated. The plot displays four country clusters, ordered from the one with the least sophisticated technology (denoted by a larger output per GG emission) to the most advanced one. The nodes in each network correspond to industries (including transport itself) that sell inputs to the transport industry.

To better illustrate the structure of these networks, we apply the Louvain algorithm to detect three communities and display them using spring layouts. The nodes are colored according to each community. The first thing to notice is that, the higher the output per GG emission, the more distinguishable the communities become. This is confirmed by higher modularity scores. Second, the separation between communities is not random, as the red nodes seem to play a brokerage or intermediary role when technology is more sophisticated. This is fascinating as it points to a specialization-vs.-generalization trade-off in technological structure as an industry becomes more sophisticated. Furthermore, the potential existence of intermediary industries calls for their identification or discovery. We investigate these questions in a more systematic way and present our findings in the remainder of this section.

We analyze the topological properties of 100+ synergy networks inferred through our method. First, we investigate which are the industries that tend to be intermediaries. As the specific nodes with a brokerage role may vary between clusters and industries, we resort to a broader categorization based on the economic sector to which each input belongs: primary, secondary, or tertiary. Then, we estimate the mixing probabilities between sectors and perform a T tests for the mixing probabilities.[1]

Figure 3A presents the T values of every pair of sectors. As we can see, the tertiary sector interacts substantially more with the other two sectors. This suggests that industries in the tertiary sector (such as communication, hospitality, finance, education) tend to mediate synergies between industries from the other sectors of the economy.

For a robustness test, we estimate two "hub-ness" measures and corroborate if our findings regarding the tertiary sector are

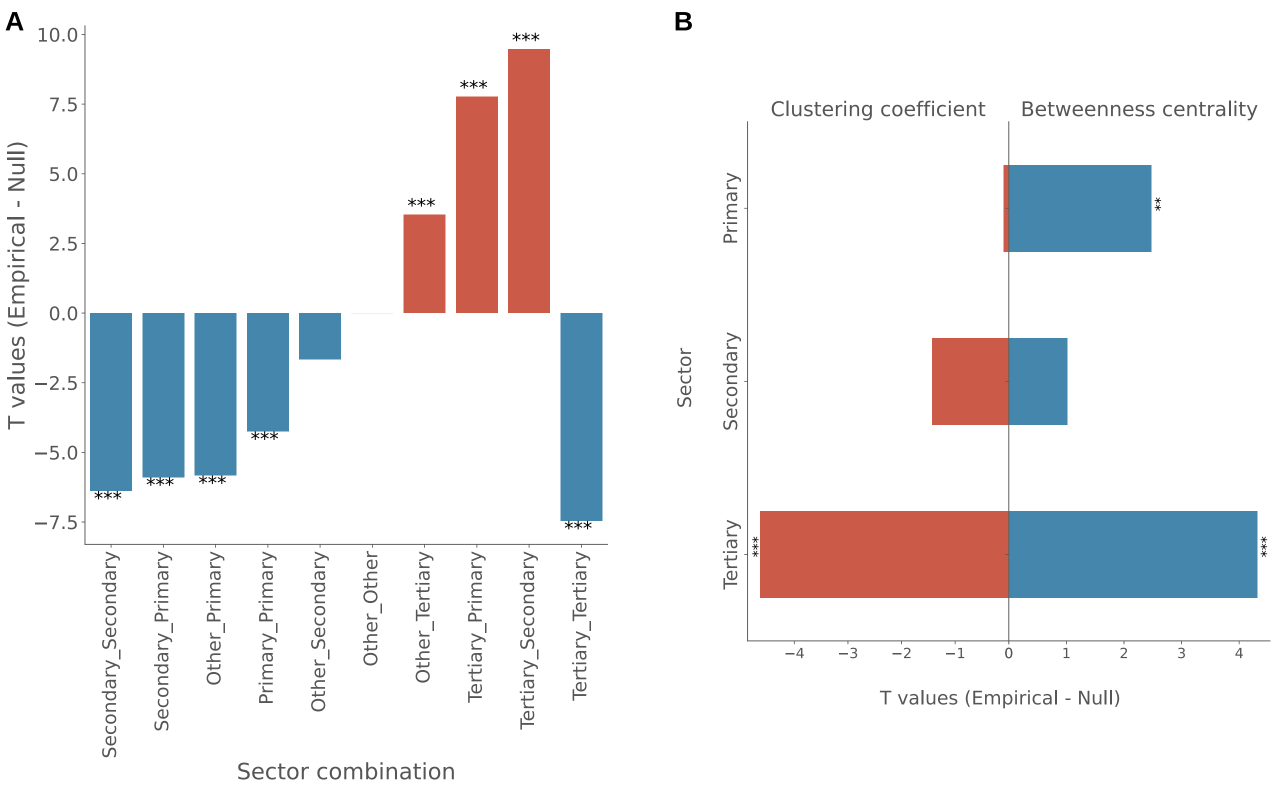

Fig. 3. Sector differences in network properties. Panel A) shows the difference in the pairwise assortativity coefficients between observed and the null synergy network, for all combinations of sectors. Panel B) shows the differences in the clustering coefficient and betweenness centrality at the sector level across empirical and null networks. These estimates are produced by computing the difference in the relevant metric between the empirical networks and the ones produced by the null model. Confidence levels are indicated by * for 90%, ** for 95%, and *** for 99%.

consistent. Hubs are usually characterized by high betweenness centrality and low clustering in networks. Figure 3B suggests that this is indeed the case for industries in the tertiary sector, distinct from those in the primary and secondary sectors (as compared to a null model).

The previous analyses confirm that a subset of inputs tend to mediate the production process of industries. Furthermore, such industries are classified into the tertiary sector. These findings are consistent with the accepted idea that membership to a particular sector conveys information about structural differences between industries and their underlying technologies, as they correspond to the main stages of production (44, 45). They are also validated by previous research showing the importance of having a more sophisticated tertiary sector to achieve higher economic complexity (18). Thus, our findings not only reveal structural features about technological sophistication but also confirm long-argued ideas about industrial development.

Now, let us analyze a broader set of (global) topological properties in synergy networks to uncover the features that underpin technological sophistication. To this end, we estimate 14 network measures that can be classified into three categories of network metrics: small-worldness, specialization, and global connectivity. Since several of these measures tend to correlate under certain topologies (46), we perform factor analysis to disentangle their contributions to each one of the three categories of network metrics.[m] We find that these factors significantly correlate with various output metrics that denote more sophisticated production processes, namely, total output, energy consumed, output per unit energy consumed, and output per unit GG emitted, as suggested by Figure 4B.

Using a robust linear model, we predict the four output measures with the leading factors. As observed in Figure 4, small-worldness and specialization are positively correlated with the output metrics. In contrast, global connectivity exhibits a negative correlation. Global connectivity, which relies on measures like algebraic connectivity, captures how resilient a network is to node removal. It can be inferred that the productive structures enabling more sophisticated processes are more fragile; as they rely on a few key nodes to keep the subnetworks connected.

Figure 4A suggests that a higher output per greenhouse gas emissions correlates with more modularity. In contrast, low output per emissions associates with a less modular structure overall. In the Supplementary information, we show that this characteristic structure is not directly observable in IO networks of normalized flows, and neither are the relationships between network structure and output efficiency documented here.

Discussion and conclusions

The quantification of technological sophistication in production processes is an elusive problem that is relevant to several disciplines. This article introduces the first data-driven method that explicitly addresses input–input and IO interactions, opening the black box of production processes. By estimating the amount of synergistic mutual information between the inputs of a production process, we quantify the degree of technological sophistication across various industries and countries. Moreover, we infer the structure of these technologies by constructing synergistic interaction networks and revealing features that characterize industrial sophistication. These networks provide empirical

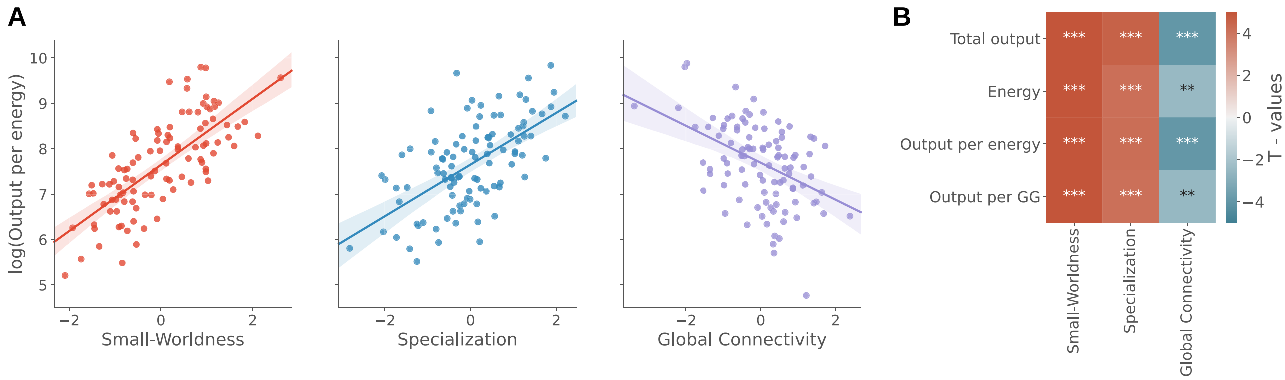

Fig. 4. Conditional correlations of salient factors with output efficiency. Output per GG stands for greenhouse gas. Panel A) shows the scatter plots of the identified network factors with output per energy and panel B) shows the T scores with all output measures considered for the analysis. It can be seen that the global connectivity is negatively correlated with all output measures, while small-worldness and specialization correlate positively with output.

grounds to select and justify production functions in IO models; something missing in this large body of literature and a major limitation in IO empirical studies. They also reveal the structural role of industries with a high degree of synergistic interactions. Finally, they suggest certain universality in the prevalence of modular small-world topologies across various classes of complex systems that perform sophisticated behaviors. For instance, in neuroscience, it has been shown that similar modular small-world structures support complex integration and consciousness in the brain, as described in the global neuronal workspace theory (47–50).

Cross-country comparisons of these network structures can help guide policymakers to prioritize targeted investments in key industrial relationships, facilitating enhanced spillover effects and spurring economic growth. For instance, policymakers often interpret predictors of economic growth such as complexity indices as signs of potential areas where a government should invest in the future (e.g. the location in the product space in which they should try to locate). Synergy networks, on the other hand can provide information about the interactions that are conducive to more productive activities. For example, by studying the synergy network of an exemplary industry (perhaps from an advanced economy), a middle-income country could gain a better understanding of which interactions in its own industry could become more synergistic, effectively guiding technology investments.

A limitation in this study is that the application of the method requires larger data than what is typically found in IO tables. This means that, in this article, we had to develop a clustering procedure (see the Methods section), and that our inferences are not for specific countries, but for groups. Another limitation is that we have to sacrifice the temporal dimension as we need to exploit time variation to estimate the underlying joint probability distributions. Ideally one would have high-frequency IO tables to compute the synergy scores by subperiods. This would allow us to study, for example, the evolution of technological sophistication and of its networked structure across countries and industries. Fortunately, new firm-transaction datasets are being generated as we write this manuscript, so the future for using the proposed approach looks very promising. Finally, the role of synergy in production process needs to be supplemented with an exploration of the role of redundancy in production process. This extension of the current study will provide a more holistic view of factors that promote resilience in addition to sophistication in the production process.

Methods

We adapt the framework of partial information decomposition (PID) to estimate synergy between inputs as a measure of complementarity. We use two inter-country IO datasets: The Eora global supply chain database (51) and the OECD inter-country IO tables (29). We discuss the specification of these datasets in IO datasets section.

PID requires estimating the joint probability distributions of inputs and outputs. However, IO datasets (prepared using national accounts data released by the governments) are only available yearly. To robustly estimate the joint probability distributions, we perform a data-clustering procedure based on the production similarity of countries for a given industry. Thus, we create a workflow that facilitates the preprocessing and inference tasks for multiple industries across both datasets. The workflow consists of two major steps, and an illustrative sketch is provided in Figure 5.

1. **Clustering**: Grouping countries that exhibit similar production patterns in a target industry, preferably in roughly equally sized groups. This means that two countries that are in the same cluster in a given industry A, may be in different clusters when analyzing another industry B.
2. **Estimation**: Using PID to estimate synergy scores for each pair of inputs in the target industry, and constructing its corresponding synergy interaction network.

Clustering countries with similar production technologies

We start by discussing the clustering procedure. To generate reliable estimates of the probability distributions, we pool together input and output growth rates across time (1995–2020) as well as from clusters of countries with similar metrics of production. These metrics include the marginal product of the inputs and factors that aid the production process such as labor market efficiency, quality of infrastructure, and the level of overall development of the country (indicated by Gross National Income per capita). These indicators help avoiding trivial clusters that could result from the coincidental similarity of marginal products due to nontechnological reasons such as exogenous shocks. For a target industry, we build a feature vector for all countries. This vector includes the median marginal product of all input industries to the target industry. It also includes auxiliary country level metrics relating to labor, infrastructure, and income

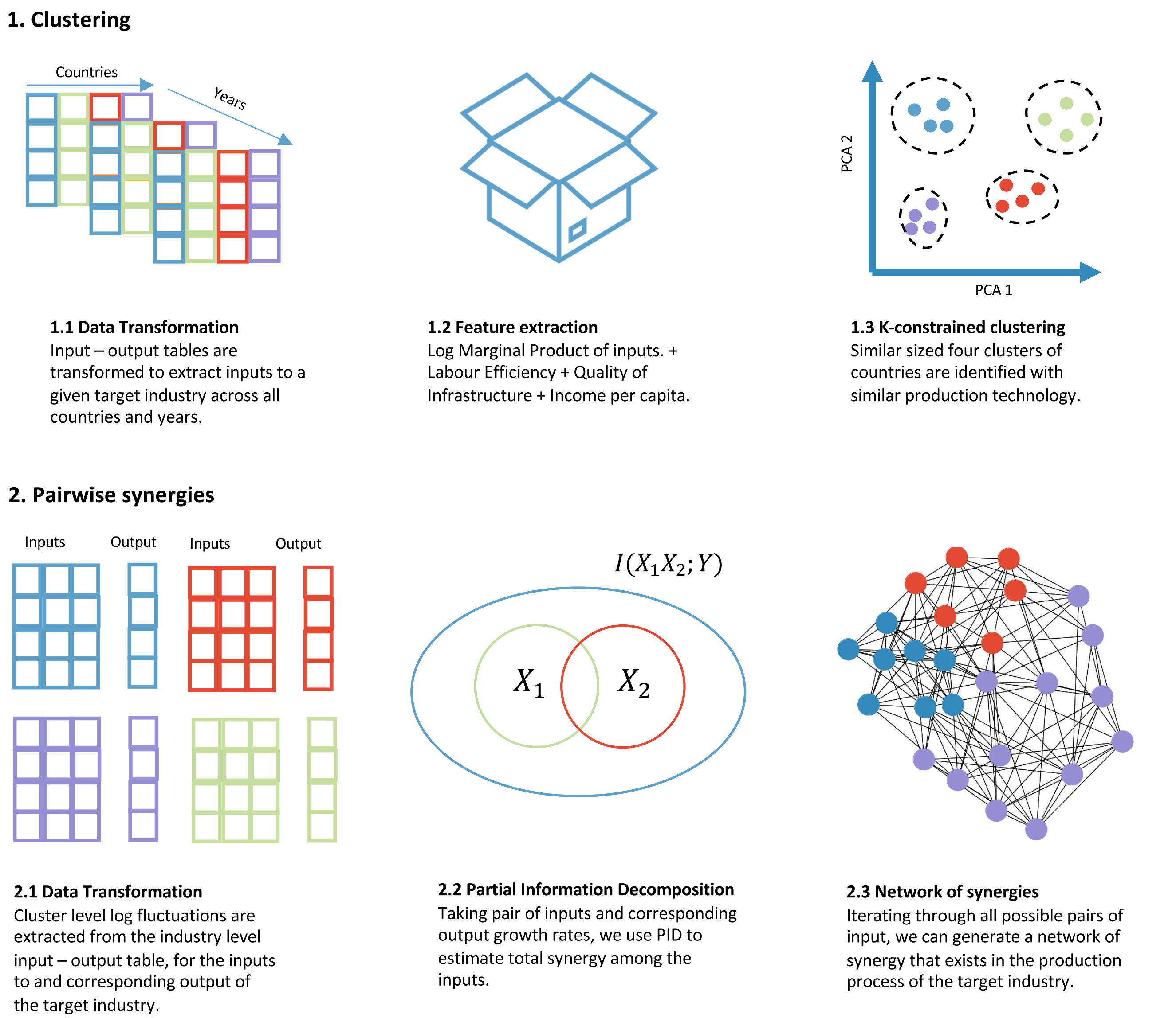

Fig. 5. The two-step workflow.

(the datasets used for estimating these auxiliary metrics are discussed in IO datasets section. Marginal product is the relative change of the output to the change in one of the inputs. This measure enables inferring the effect of each input at a first level of approximation.

To estimate these features, let us introduce some notation useful to describe the IO datasets. We denote the entries of a typical inter-country IO table as, $T^{c',c}_{i,j}$, the total value flow from industry i in country c' to industry j in country c (in USD basic prices). Then, the total input inflow from industry i to j for country c can be expressed as a sum of value flows from industry i to j over all countries c',

$$X^c_{i,j} = \sum_{c'} T^{c',c}_{i,j}. \tag{2}$$

Similarly, the total output (Y_j) of industry j in country c can be defined as

$$Y^c_j = \sum_{c',i} T^{c,c'}_{j,i} + \sum_{c'} F^{c,c'}_j, \tag{3}$$

where $F^{c,c'}_j$ represents the final demand of goods and services of the industry j of any country c' in country c. Having identified the total value of inputs and the output at the industry level, the marginal product of the input coming from industry j to industry i (in country c) can be written as,

$$MP^c_{i,j}(j) = \frac{Y^c_j(t) - Y^c_j(t-1)}{X^c_{i,j}(t) - X^c_{i,j}(t-1)}. \tag{4}$$

The median marginal product for all the available years is a feature of the inputs of a particular industry. Marginal products usually exhibit fat-tailed distributions across countries because of the characteristic output differences among the economies. Therefore, we log-transform the marginal product to normalize the features for clustering. We use log MP as the key feature for clustering countries with similar production technologies, along with the auxiliary development indicators discussed above. Finally, we employ the k-means-constrained clustering algorithm (52) to define four country clusters in a target industry, using the features described above.

We choose this constrained version of the k-means algorithm because it allows finding clusters that are balanced in size.[n] The unconstrained version of the k-clustering algorithm may introduce biases in the synergy scores due to under or over representation of certain types of countries. The hyperparameters of the algorithm, including the number of clusters, are optimized using consensus clustering.[o] In the rest of this section, we explain the estimation of the synergy score on IO datasets using PID as well as the estimation of economic complexity indices on the industry level.

Estimating the synergy score

Information-theoretic entropy measures the variability of a system by quantifying the uncertainty of being in a given state. Developing this mathematical framework, Shannon proposed a measure of dependence between two variables, commonly known as mutual information (53). The ability to assess the variability of a system—and the inter-dependency among its variables—made information theory ideal for the empirical study of a wide range of complex systems (54, 55).

Here, we adapt these concepts and tools to quantify the inter-dependencies between an industry's inputs and corresponding output. We start by briefly introducing mutual information and how it can be decomposed using PID. Finally we explain how these measures are implemented to IO datasets.

Mutual information

Mutual information quantifies the amount of information shared between two random variables. This shared information can be interpreted as a nonparametric measure of dependence between the two variables. This dependence can be quantified using the joint and marginal probability distributions of the variables. The mutual information between two random variables X and Y can be defined as below,

$$I(X; Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right), \tag{5}$$

where $p(x)$ is the probability distribution corresponding to the random variable X on the state space $\mathcal{X}$ and $p(x, y)$ represents the joint distribution of the two random variables. In essence, mutual information can be understood as the Kulback–Leibler distance between the joint distribution of the variables and the product of their marginal distributions. The measure $I(X; Y)$ becomes 0 if X and Y are independent, such that $p(x, y) = p(x)p(y)$.

For the case of Gaussian random variables, the corresponding mutual information can be written as (see (56) for a complete derivation)

$$I(X; Y) = \frac{1}{2}\log\left(\frac{det\Sigma(X)}{det\Sigma(X|Y)}\right), \tag{6}$$

where $det\Sigma(X)$ represents the determinant of the covariance matrix of X, and $\Sigma(X|Y)$ represents the conditional covariance, which can be written as

$$\Sigma(X|Y) = \Sigma(X) - \Sigma(X, Y)\Sigma(Y)^{-1}\Sigma(Y, X). \tag{7}$$

The definition in Eq. 6 can be extended to the multivariate setting.

For three or more variables, we can further investigate the nature of the interdependence quantified using mutual information. This involves quantifying higher-order effects such as synergistic interactions using the PID framework (27).

Partial information decomposition

PID breaks down the mutual information between a pair of input variables and the target output variable into *Synergistic*, *Redundant*, and *Unique* information. These three modes correspond to the amount of information that is present in the two variables collectively, common to both variables, and unique to each variable, respectively. This decomposition was first introduced in (27) and has been instrumental to study different types of interactions between random variables (57–59).

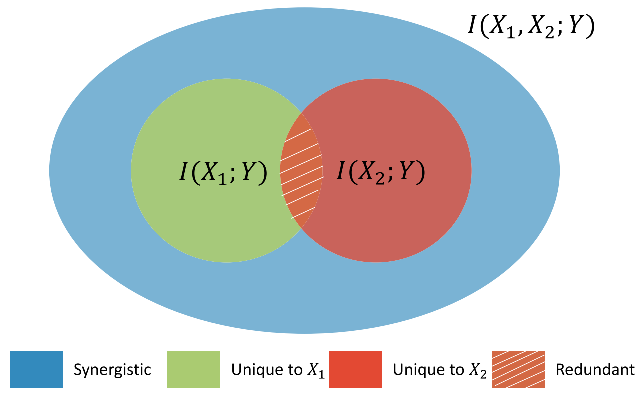

Fig. 6. Partial information decomposition.

Let us look at the case of mutual information between two input variables (X_1, X_2) and one output variable (Y). Here, the total mutual information about the output provided by the two inputs can be represented as the Venn diagram in Fig. 6.

Redundant information could come from either X_1 or X_2, as it is common to both variables, this information would be preserved even if one of the inputs was removed. Unique information comes from each input only, so it represents the unique contribution that each input makes to the output while conditioning the effect of the other input. Finally, synergistic information is the information about the target that is only accessible when both inputs are observed together. This information is lost when any one of the inputs is removed.

Let us write this decomposition as

$$\begin{aligned} I(X_1, X_2; Y) = Syn(X_1, X_2; Y) + Red(X_1, X_2; Y) \\ + Unq(X_1; Y) + Unq(X_2; Y). \end{aligned} \tag{8}$$

A general formulation of Eq. 8 for n random variables is provided by Ref. (27). The unique and the total joint information about the output provided by the inputs can be exactly calculated. However, to estimate the synergy and redundancy between the inputs, we need to assume a redundancy function. We employ the minimum mutual information redundancy estimator, which is known to work reliably for multivariate Gaussian systems (59).[p] This redundancy estimator assumes the total redundancy to be equivalent to the mutual information about the output provided by the weakest input (i.e. the input that has the least mutual information with the output). Formally, this is

$$\mathrm{Red}(X_1, X_2; Y) = \min\,(I(X_1; Y), I(X_2; Y)).$$

Following (59), using this redundancy function yields the estimator of the synergistic interaction to be

$$\mathrm{Syn}(X_1, X_2; Y) = I(X_1, X_2; Y) - \max\,(I(X_1; Y), I(X_2; Y)) \tag{9}$$

By normalizing the inputs and output variables to unit variance (for continuous measurements), we fix an upper bound on the total mutual information between the inputs and the output $I(X_1, X_2; Y)$ and thus on the synergy score $\mathrm{Syn}(X_1, X_2; Y)$ (see Eq. 9).

To remove potential numerical artifacts, we implement a bias correction procedure by performing a randomized estimation where the inputs are shuffled. The synergy score calculated using the randomized data is then subtracted from the true synergy score. This ensures that the statistical significance of the estimated mutual information by removing the effect of spurious correlations.

Synergistic interaction networks

To estimate the synergy score on IO dataset, we deploy the PID framework on the growth rate of inputs and output of each target industry, for a given cluster of countries. Firstly, the input growth rates $\hat{X}^c_{i,j}(t)$ and the corresponding output growth rate $\hat{Y}^c_j(t)$, for the year t and country c can be defined as

$$\begin{aligned} \hat{X}^c_{i,j}(t) &= \log \frac{X^c_{i,j}(t)}{X^c_{i,j}(t-1)} \\ \hat{Y}^c_j(t) &= \log \frac{Y^c_j(t)}{Y^c_j(t-1)}. \end{aligned} \tag{10}$$

For a given target industry of a cluster of countries (as clustered in Clustering countries with similar production technologies section), vectors of input and output growth rates are collated across years and countries belonging to the cluster. This procedure ensures enough observations for a reliable estimation of joint probability distribution of between input and output growth rates. Therefore, the synergy score between inputs from industries i and j corresponding to target industry k can be written using Eq. 10 as,

$$\begin{aligned} &Syn(\hat{X}_{i,k}, \hat{X}_{j,k}; \hat{Y}_k) \\ &\quad = I(\hat{X}_{i,k}, \hat{X}_{j,k}; \hat{Y}_k) - \max\left(I(\hat{X}_{i,k}; \hat{Y}_k), I(\hat{X}_{j,k}; \hat{Y}_k)\right). \end{aligned} \tag{11}$$

Using Eq. 11, synergistic interaction network S, of a given target industry k is constructed by estimating the synergy scores of all pairs of input industries i, j. Such that the adjacency matrix $S_{i,j}$ can be written as

$$S_{i,j} = \mathrm{Syn}(\hat{X}_{i,k}, \hat{X}_{j,k}; \hat{Y}_k). \tag{12}$$

Thus, each network S is a weighted undirected graph whose nodes are input industries and edges are synergy scores.

We prune the fully connected network S, to remove nonsalient edges and uncover the underlying backbone. We extract the backbone of each network by using the method discussed in (60), which builds on the well-known disparity filter (61) by automating the selection of a filtering threshold. To estimate the measures of centrality, clustering, and mixing probabilities discussed in Fig. 3A, we use the binary (unweighted) version of these backbone networks.

When performing statistical analysis in Fig. 3 on the synergy networks, we use a null model. This model generates null networks by shuffling the edges across all possible combinations of nodes. In this way, the null networks have a similar edge-weight distribution as the reference networks. The shuffled structure allows for all possible inter-sector interactions. We generate an ensemble of 1,000 null networks for every reference network to correct for sampling bias. The properties of interest such as the mixing probability, clustering, and betweenness are calculated both on the reference and the null networks. The differences between the properties in these two distributions is used in the t tests.

IO datasets

The IO data are obtained from two sources, the Eora global supply chain dataset (28) and the OECD inter-country IO tables (29). the Eora26, an aggregated IO dataset at the level of 26 industries, is built using data from national accounts of the countries and various international organizations (51). It is the largest IO dataset in terms of country coverage (181 economies). Eora26 has become standard in the study of global value chains (62) and material footprint (63). It has also been shown to be consistent with similar, but smaller, global databases (64).

In addition to the Eora26 dataset, we replicate the analysis on the IO dataset made available by the OECD.[q] The November 2021 version of this dataset includes IO monetary transactions between 66 countries across 45 unique industries in the time period of 1995–2018.

International trade:

In our validation, we calculate the indices of economic sophistication using the BACI international trade dataset (37), which is independent from Eora26. These data contain trade records between 200 countries for more than 5,000 products during the 1995–2020 period. These products are uniquely classified into 15 of the 26 industries of Eora26 through the HS92-to-ISIC correspondence tables of the UN Statistics Division.

Development indicators:

We use three auxiliary development indicators to improve our clustering procedure (see Clustering countries with similar production technologies section). The first is the World Bank's gross national income per capita indicator. The second and third are the labor efficiency and infrastructure indicators from the World Economic Forum's Global Competitive Index Report. Gross national income covers the same period as Eora26, while the other two indicators are available from 2007 to 2017.[r]

Notes

[a] Either from indigenous innovations or through the emulation or adoption of examples from other societies, countries, sectors, or industries (65)

[b] Further studies in the relatedness literature suggest that the complementarity between two economic entities also depends on the network of economic interactions that embed them (13, 30).

[c] While (4) claim that synergy lacks a formal mathematical definition in economics, this is not the case in information theory.

[d] Compelling ex post rationalizations are made to validate these measures as proxies for complementarity.

[e] To mitigate criticisms to the specification of production functions, expert surveys have been conducted to determine the degree of dependence on certain inputs (see the IHS Markit survey in Refs. (66, 67)). This information, has been incorporated in the new generation of IO models (20, 66).

[f] Recently, it has been shown that higher-level brain regions responsible for cognition and emotion have more synergistic interactions (as they combine information from multiple sensory inputs) (68). Conversely, sensory regions of the brain that carry multiple copies of the same information exhibit high redundancy. These findings suggest that the synergy score, as defined by partial information decomposition (PID), is a reliable measure of complementarity of inputs.

[g] Equation 1 can be generalized for more than two inputs. However, producing reliable estimates for more than three inputs can become too data-demanding. Nevertheless, we show in the Supplementary information that our main results hold when we estimate synergies between triplets instead of pairs.

[h] This transformation to growth rates (rather than raw values) yields normally distributed growth rates at the industry level, which

makes the data compatible with Gaussian estimators of mutual information.

[i] See Ref. (30) for a comprehensive review of the economic complexity literature.

[j] To the extent of our knowledge, most studies focus on national and regional analysis. References (35, 36) are the only studies we are aware of that look at sector level economic fitness. Industry-level calculations require mapping products into industries, which we do and explain in the Supplementary information.

[k] While co-occurrence strategies have been devised to estimate complementarities in the context of skills (69, 70), they fail to embed the output signal in the estimation step. Instead, they analyze associations between co-occurrence metrics and output data such as wages, an ex post rationalization of sorts. Furthermore, another issue with co-occurrence is that it captures both complementary and substitute interactions; for example, resilient systems tend to exhibit functionally similar behavior distributed across the system to ensure robustness to perturbations (71). These redundant/substitutable interactions help maintain the function in noisy and uncertain environments (72–74). However, synergistic interactions quantify the complementarity among parts but are more fragile (75). Co-occurrence measures weigh these interactions equally, missing the differentiation between these two kinds of interactions.

[l] With reference to a null model (described in the methods Synergistic interaction networks section), followed by a false discovery rate correction for multiple comparisons.

[m] The full list of measures and their contribution to specific factors is provided in the Supplementary information.

[n] In the Supplementary information, we show that our results are robust to the traditional unconstrained k-means algorithm.

[o] The Supplementary information explains how principal component analysis of the feature matrix is used to ensure that the marginal product is the leading feature.

[p] There is a growing number of different redundancy functions in the PID literature (57).

[q] https://www.oecd.org/sti/ind/input-outputtables.htm

[r] We find that this auxiliary information is very useful to obtain coherent technology clusters without trivially becoming the leading feature.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by Wave 1 of The UK Research and Innovation (UKRI) Strategic Priorities Fund under the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/W006022/1, particularly the "Shocks and Resilience" cross-theme within that grant; the Ecosystem Leadership Award under the EPSRC Grant EP/X03870X/1; the Economic and Social Research Council under Grant ES/T005319/2; and The Alan Turing Institute. H.R. is also supported by EPSRC Grant EP/W024020/1 as part of the Statistical Physics of Cognition project.

Author Contributions

H.R. and O.G. conceptualized the project. H.R. conducted the analysis and prepared the visualizations. O.G. performed validation tests and supervised the project. Both authors contributed to the writing and editing of the manuscript.

Preprints

This manuscript was posted on a preprint: https://arxiv.org/abs/2301.04579.

Data Availability

EORA 26 dataset is freely available for academic use for the years 1990–2015. However, data between 2016 and 2021 is available upon licensed from their website (https://worldmrio.com/eora26/). BACI Export dataset used for calculating economic complexity indices is available freely online (http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37). The OECD IO dataset is available freely on the organizations website (http://oe.cd/icio). Other open source datasets used in the analysis include, the World bank GDP dataset (https://databank.worldbank.org) and the Global competitive index published by the World Economic Forum (https://www.weforum.org/reports/the-global-competitiveness-report-2017-2018/). The information-theoretic estimations were performed using the Java Information Dynamics Toolkit, which is available on GitHub (https://github.com/jlizier/jidt) along with the K-Means Constrained clustering algorithm (https://github.com/joshlk/k-means-constrained).

References

1 Jacobs J. *The economy of cities*. Penguin Books, New York, 1969.
2 Glaeser EL, Kallal HD, Scheinkman JA, Shleifer A. 1992. Growth in cities. *J Pol Econ*. 100(6):1126–1152.
3 Frenken K, Van Oort F, Verburg T. 2007. Related variety, unrelated variety and regional economic growth. *Reg Stud*. 41(5): 685–697.
4 Brynjolfsson E, Milgrom P. 2013. Complementarity in organizations. In: *The handbook of organizational economics*. Princeton (NJ): Princeton University Press. p. 11–55.
5 Lee FLF. 2016. Impact of social media on opinion polarization in varying times. *Commun Public*. 1(1):56–71.
6 Gölpek F. 2015. Service sector and technological developments. *Procedia Soc Behav Sci*. 181(1):125–130.
7 Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. 2011. The ubiquity of small-world networks. *Brain Connect*. 1(5): 367–375.
8 Girvan M, Newman MEJ. 2002. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 99(12): 7821–7826.
9 Adamic LA. 1999. The small world web. Research and Advanced Technology for Digital Libraries: Third European Conference, ECDL'99 Paris, France, September 22–24, 1999 Proceedings 3. Springer. p. 443–452.
10 Tam Cho WK, Fowler JH. 2010. Legislative success in a small world: social network analysis and the dynamics of congressional legislation. *J Polit*. 72(1):124–135.
11 Uzzi B, Amaral LAN, Reed-Tsochas F. 2007. Small-world networks and management science research: a review. *Eur Manage Rev*. 4(2):77–91.
12 Jeong H, Mason SP, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature*. 411(6833):41–42.
13 Neffke F, Henning M, Boschma R. 2011. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Econ Geogr*. 87(3):237–265.
14 Hidalgo CA, Klinger B, Barabási A-L, Hausmann R. 2007. The product space conditions the development of nations. *Science*. 317(5837):482–487.

15 Neffke FMH. 2019. The value of complementary co-workers. *Sci Adv*. 5(12):eaax3370.
16 Dibiaggio L, Nasiriyar M, Nesta L. 2014. Substitutability and complementarity of technological knowledge and the inventive performance of semiconductor companies. *Res Policy*. 43(9): 1582–1593.
17 Castañeda G, Castro Peñarrieta L, Guerrero OA, Chávez-Juárez F. 2024. How do social capabilities shape a country's comparative advantages? Unpacking industries' relatedness. *Weltwirtsch Arch*. 160(3):1083–1116.
18 Gala P, Camargo J, Magacho G, Rocha I. 2018. Sophisticated jobs matter for economic complexity: an empirical analysis based on input-output matrices and employment data. *Struct Chang Econ Dyn*. 45(2):1–8.
19 Leontief W. *The structure of American economy, 1919–1929: an empirical application of equilibrium analysis*. Harvard University Press, 1941.
20 Diem C, Borsos A, Reisch T, Kertész J, Thurner S. 2022. Quantifying firm-level economic systemic risk from nation-wide supply networks. *Sci Rep*. 12(1):7719.
21 Zambelli S. 2018. The aggregate production function is NOT neoclassical. *Cambridge J Econ*. 42(2):383–426.
22 Gechert S, Havranek T, Irsova Z, Kolcunova D. 2019. Death to the Cobb-Douglas production function? A quantitative survey of the capital-labor substitution elasticity. Technical Report 203136. ZBW - Leibniz Information Centre for Economics.
23 Iyetomi H, Aoyama H, Fujiwara Y, Ikeda Y, Souma W. 2012. A paradigm shift from production function to production copula: statistical description of production activity of firms. *Quant Finance*. 12(9):1453–1466.
24 Constantino M, Candido O, Borges E, Silva T, Tabak B. 2019. Modeling vine-production function: an approach based on vine copula. *Physica A: Stat Mech Appl*. 531(1):121724.
25 Steward D. 1981. The design structure system: a method for managing the design of complex systems. *IEEE Trans Eng Manag*. EM-28(3):71–74.
26 Browning T. 2001. Applying the design structure matrix to system decomposition and integration problems: a review and mew directions. *IEEE Trans Eng Manag*. 48(3):292–306.
27 Williams P, Beer R. 2010. Nonnegative decomposition of multivariate information, arXiv, arXiv:1004.2515, https://doi.org/10.48550/arXiv.1004.2515, preprint: not peer reviewed.
28 Lenzen M, Kanemoto K, Moran D, Geschke A. 2012. Mapping the structure of the world economy. *Environ Sci Technol*. 46(15): 8374–8381.
29 [dataset] OECD. Oecd inter-country input-output database. 2023. Data available at: http://oe.cd/icio.
30 Hidalgo C. 2021. Economic complexity theory and applications. *Nat Rev Phys*. 3(2):92–113.
31 Felipe J, Kumar U, Abdon A, Bacate M. 2012. Product complexity and economic development. *Struct Chang Econ Dyn*. 23(1):36–68.
32 Hausmann R, Hidalgo CA, Bustos S, Coscia M, Simoes A. *The atlas of economic complexity: mapping paths to prosperity*. MIT Press, 2014.
33 Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L. 2012. A new metrics for countries' fitness and products' complexity. *Sci Rep*. 2(1):723.
34 Sciarra C, Chiarotti G, Ridolfi L, Laio F. 2020. Reconciling contrasting views on economic complexity. *Nat Commun*. 11(1):3352.
35 Pugliese E, Tacchella A. 2021. Economic complexity analytics: country factsheet. Technical Report EUR 30711 EN. Joint Research Centre, European Commission.
36 Tuğrul Çınar I, Korkmaz I, Baycan T. 2022. Regions' economic fitness and sectoral labor productivity: evidence from turkey. *Reg Sci Policy Pract*. 14(3):575–599.
37 Gaulier G, Zignago S. 2010. BACI: international trade database at the product-level. The 1994–2007 Version. Working Paper, CEPII research center.
38 Rosenstein-Rodan PN. 1943. Problems of industrialisation of eastern and south-eastern Europe. *Econ J*. 53(210–211):202–211.
39 Prebisch R. *The economic development of Latin America and its principal problems*. ECLA, 1950.
40 Hirschman AO. *The strategy of economic development*. Yale University Press, 1958.
41 Boschma R. 2017. Relatedness as driver of regional diversification: a research agenda. *Reg Stud*. 51(3):351–364.
42 Nelson RR. 1981. Research on productivity growth and productivity differences: dead ends and new departures. *J Econ Lit*. 19(3): 1029–1064.
43 Dosi G, Grazzi M. 2006. Technologies as problem-solving procedures and technologies as input–output relations: some perspectives on the theory of production. *Ind Corp Change*. 15(1):173–202.
44 Fisher A. 1939. Production, primary, secondary and tertiary. *Econ Rec*. 15(1):24–38.
45 Clark C. *Revised and reprinted in 1951. The conditions of economic progress*. MacMillan & Co. Ltd, London, UK, 1940.
46 Li C, Wang H, De Haan W, Stam CJ, Van Mieghem P. 2011. The correlation of metrics in complex networks with applications in functional brain networks. *J Stat Mech*. 2011(11):P11018.
47 Baars BJ. 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog Brain Res*. 150:45–53.
48 Dehaene S, Naccache L. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*. 79(1–2):1–37.
49 Shanahan M. 2012. The brain's connective core and its role in animal cognition. *Philos Trans R Soc B: Biol Sci*. 367(1603): 2704–2714.
50 Mashour GA, Roelfsema P, Changeux J-P, Dehaene S. 2020. Conscious processing and the global neuronal workspace hypothesis. *Neuron*. 105(5):776–798.
51 Lenzen M, Moran D, Kanemoto K, Geschke A. 2013. Building Eora: a global multi-region input–output database at high country and sector resolution. *Econ Syst Res*. 25(1):20–49.
52 Bradley P, Bennett K, Demiriz A. 2000. Constrained K-means clustering. Technical Report MSR-TR-2000-65. Microsoft Research.
53 Shannon C. 1948. A mathematical theory of communication. *Bell Syst Tech J*. 27(3):379–423.
54 Rajpal H, *et al*. 2022. Psychedelics and schizophrenia: distinct alterations to Bayesian inference. *Neuroimage*. 263(1):119624.
55 Rajpal H, *et al*. 2023. Quantifying hierarchical selection, arXiv, arXiv:2310.20386, https://doi.org/10.48550/arXiv.2310.20386, preprint: not peer reviewed.
56 MacKay DJC, Mac Kay DC. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.
57 Lizier J, Bertschinger N, Jost J, Wibral M. 2018. Information decomposition of target effects from multi-source interactions: perspectives on previous, current and future work. *Entropy*. 20(4):307.
58 Wibral M, Priesemann V, Kay J, Lizier J, Phillips W. 2017. Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain Cogn*. 112(23):25–38.
59 Barrett A. 2015. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys Rev E*. 91(5):052802.

60 Coscia M, Neffke FMH. 2017. Network backboning with noisy data. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*. IEEE. p. 425–436.
61 Serrano M, Boguñá M, Vespignani A. 2009. Extracting the multi-scale backbone of complex weighted networks. *Proc Natl Acad Sci U S A*. 106(16):6483–6488.
62 UNCTAD. 2019. Transnational Corporations. United Nations.
63 Wiedmann T, *et al.* 2015. The material footprint of nations. *Proc Natl Acad Sci U S A*. 112(20):6271–6276.
64 Moran D, Wood R. 2014. Convergence between the Eora, WIOD, EXIOBASE, and OpenEU's consumption-based carbon accounts. *Econ Syst Res*. 26(3):245–261.
65 Akamatsu K. 1962. A historical pattern of economic growth in developing countries. *Dev Econ*. 1(s1):3–25.
66 Pichler A, Pangallo M, del Rio-Chanona R, Lafond F, Farmer D. 2022. Forecasting the propagation of pandemic shocks with a dynamic input-output model. *J Econ Dyn Control*. 144(3):104527.
67 GEODIS. 2017. GEODIS unveils its 2017 supply chain worldwide survey. Technical Report.
68 Luppi AI, *et al.* 2022. A synergistic core for human brain evolution and cognition. *Nat Neurosci*. 25(6):771–782.
69 Farinha T, Balland P-A, Morrison A, Boschma R. 2019. What drives the geography of jobs in the us? Unpacking relatedness. *Ind Innov*. 26(9):988–1022.
70 Stephany F, Teutloff O. 2024. What is the price of a skill? The value of complementarity. *Res Policy*. 53(1):104898.
71 Randles M, Lamb D, Odat E, Taleb-Bendiab A. 2011. Distributed redundancy and robustness in complex systems. *J Comput Syst Sci*. 77(2):293–304.
72 Von Neumann J. 1956. Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata Stud*. 34(34):43–98.
73 Edelman GM, Gally JA. 2001. Degeneracy and complexity in biological systems. *Proc Natl Acad Sci U S A*. 98(24):13763–13768.
74 Kitano H. 2004. Biological robustness. *Nat Rev Genet*. 5(11): 826–837.
75 Luppi AI, Rosas FE, Mediano PAM, Menon DK, Stamatakis EA. 2024. Information decomposition and the informational architecture of the brain. *Trends Cogn Sci*. 28(4):352–368.