ON THE IMPORTANCE OF MANUFACTURING SECTORS FOR ECONOMIC DEVELOPMENT AND GROWTH*

ALEXANDER RADEBACH

HAUKE SCHULT

JAN CHRISTOPH STECKEL

Total word count: 7135

*The authors thank Bob Brecha, Felix Creutzig, Ottmar Edenhofer, Sabine Fuss, César Hidalgo, Michael Jakob, Gregory Nemet and Narasimha Rao for their valuable comments and discussions, as well as Mia Burger for her outstanding research assistance. This research was supported by the German Research Foundation's (DFG) Collective Research Center 1026, "Sustainable Manufacturing."

Corresponding author: Alexander Radebach, alexander.radebach@mcc-berlin.net, Torgauer Strasse 12-15, 10829 Berlin, Germany

ABSTRACT

Various attempts have been made to empirically understand the role of economic sectors in development and growth processes. In this work, we develop an innovative methodology to assess associations in a cross-country multi-sectoral dataset based on complex network-related approaches. This paper generalizes existing concepts—such as the "product space", which identifies a core-periphery pattern of the global economy using export data—by first considering total economic activities in terms of value-added data deduced from multi-regional input-output tables and, second, by assessing the relationship between inter-sectoral imbalances and overall economic growth. We find clearly distinguishable groups of sectors, mainly agricultural, industrial, services and resource extractive sectors. These are primarily linked via light manufacturing sectors. The existence of these sectoral bottlenecks, or bridges that are stable over time, allows us to conclude that (i) the buildup of specific manufacturing sectors is crucial for establishing the capabilities required for transitional growth and (ii) leapfrogging an economy's industrialized state is difficult. Along with the directionality information derived from the analysis of sectoral imbalances, our results are consistent with and expand classical sectoral ladder models towards a sectoral "climbing wall" where multiple development routes are conceivable. Those are, however, constrained by specific identifiable bottlenecks. Our conclusions have notable implications for any attempts to alleviate poverty and foster growth in light of global environmental change. *JEL* Codes: C38, C55, O11, O14, Q54

I. INTRODUCTION

Although we have witnessed remarkable success regarding poverty alleviation in recent decades, more than one billion people still live below absolute poverty levels today (Chen and Ravallion 2010) and far more lack access to basic infrastructure services, including water and energy (World Bank 2013). The majority of the world's poor lives in countries with economies that are strongly dependent on agriculture (see Figure I) and/or resource extraction.

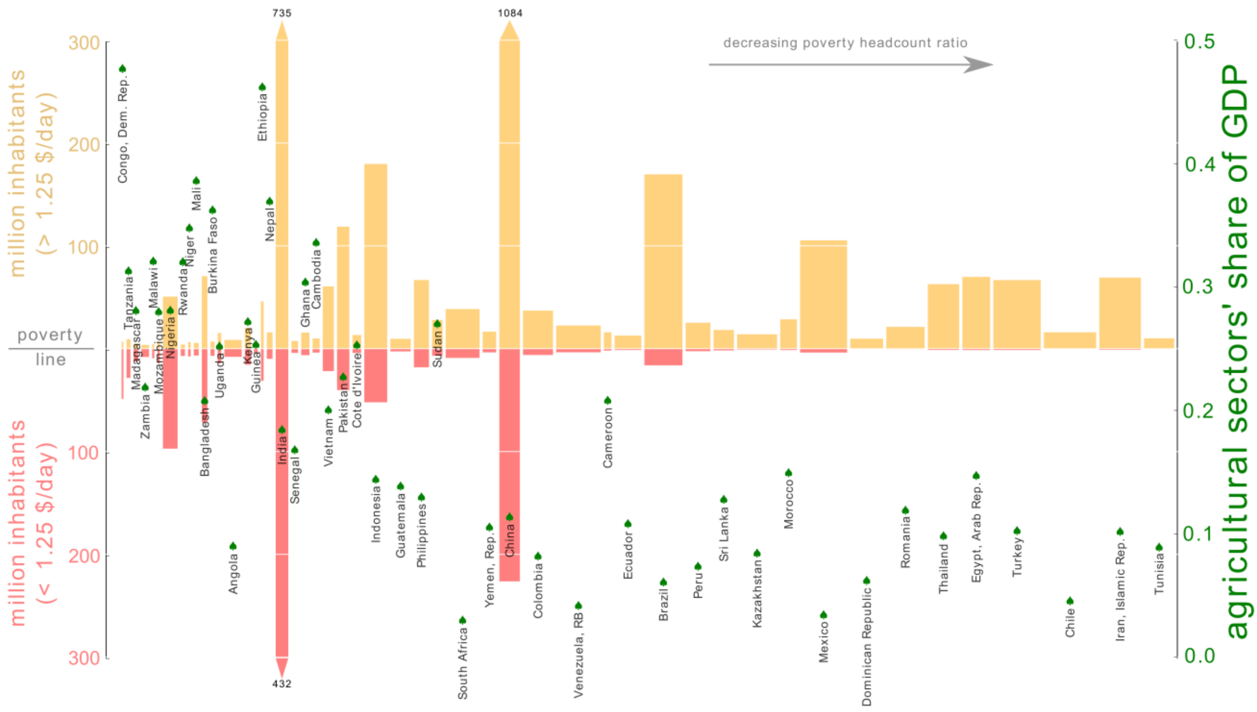


FIGURE I

Extreme poverty and agricultural share across countries

The red bars' height shows the number of people living below the poverty line of US$1.25 per day. The golden bars' height shows the remaining population. The width of the bars indicates the gross domestic product (GDP) per capita. Thus, the area of the bars depicts the total GDP. The green leaves refer to the share of agricultural sectors in GDP. The countries are sorted according to their decreasing poverty headcount ratio from the left to the right. In countries where a high percentages of the population lives in absolute poverty, agricultural sectors dominate the economy. The data were taken from the World Development Indicators database (World Bank 2014), averaged between years from 2000 to 2012 for which data are completely available per country and filtered for countries with more than 10 million inhabitants and more than 1% of the population living below the absolute poverty line. All monetary data are given in power purchase parity (constant 2011 international $).

At the other end of the spectrum, economies of countries that are commonly classified as "developed" (i.e., having high or very high income levels, as classified by the World Bank) account for the lion's share of global industrial production and provision of services. In the past, countries that have transformed from lower to higher per-capita income levels have industrialized by building physical and other infrastructure stocks. A challenging question is whether countries without significant manufacturing and service sector capacities necessarily need to go through the same structural transformations as developed countries have in the past, or whether some generalized leapfrogging might be possible.

This question gains even more relevance in light of the challenges imposed by global environmental change, as it is known that resource and energy use, as well as related greenhouse gas emissions and other environmental externalities, are largely generated by the manufacturing sectors, resource extraction and industrialized agriculture (Edenhofer et al. 2014). Conversely, global environmental change affects living conditions (Pachauri 2008), which raises the question of how economic growth can be feasible without simultaneously making global climate stabilization infeasible.

Past research has explored economic growth from various angles (Temple 1999) with emphasis on, for example, the role of capital accumulation during transitional growth and technological change for long-term growth (Solow 1956, 1994), human capital (Lucas 1988; Mankiw, Romer, and Weil 1992), research and development (Aghion and Howitt 1992; Grossman and Helpman 1991a, 1991b; Romer 1986, 1990), institutions (Acemoglu, Johnson, and Robinson 2005), political systems (Przeworski 2000) and geography (Diamond 1997; Sachs 2001). Since the studies of Albert O. Hirschman the consideration of different economic sectors and their interactions are thought to provide insights into development processes (Hirschman 1958). From a more macroscopic view, agriculture as the primary sector, industrial production as the secondary and services as the tertiary sector have been distinguished (Herrendorf, Rogerson, and Valentinyi 2014). However, recent datasets feature high sectoral resolutions that allow for a detailed investigation of compositional changes over development stages.

Furthermore, while it has long been argued *theoretically* that a better understanding of development patterns requires a more detailed understanding of disaggregated output, some recent conceptual efforts *empirically* show the complexity of national exports to be predictive for economic growth (Hidalgo et al. 2007; Hidalgo and Hausmann 2009). Furthermore, the same authors have claimed that export baskets can be used to assess the capabilities that are available in a particular economy.[1] However, it is also necessary to assume that the global trade network serves as an appropriate proxy for those assets in order to hold the abovementioned authors' theoretical argumentation based on revealed comparative advantages (Balassa 1965). To determine the (relative) level of capabilities, however, we argue that what countries are actually producing should matter more than their specific comparative advantages. Therefore, this paper's analysis is based on value-added data deduced from global multi-regional input-output (MRIO) tables, which can provide a fairly detailed picture of the global economy, even though they are more aggregated than export data. By thoroughly revising and expanding the "product space" approach (Hidalgo et al. 2007), this work aims to contribute to the understanding of the role of economic sectors for economic development.

The remainder of the paper is structured as follows: Section II describes the method and data, while Section III presents numerical results and comments on their statistical robustness. Section IV discusses the implications for the understanding of economic development at the sectoral level, particularly with respect to global environmental change, and provides some concluding remarks and questions for further research.

[1] In this study, "capabilities" shall be used in a rather broad sense by comprising all non-tradable assets that serve economic activity.

II. METHOD AND DATA

The empirical analysis of the sectoral composition of national economies has been enriched recently by the introduction of the product space (Hidalgo et al. 2007; Hidalgo and Hausmann 2009), which essentially is a square matrix of some measure of association between economic sectors, e.g., a cross-correlation matrix. More specifically, Hidalgo et al. estimate the conditional probability of above-average exports of products, given the above-average export of a chosen product, measured over the ensemble of countries. We will follow the key idea of statistical similarities of sectors, but scrutinize the approach substantially in terms of methods and assumptions.

II.A. Data

Given the ever progressing globalization of the world economy, particularly in light of vertical specialization (Hummels, Ishii, and Yi 2001), the production structure of national economies has increasingly attracted attention in terms of international trade and total production on one hand, and total output and value-added on the other hand (Koopman, Wang, and Wei 2014).

Established theory on international trade focuses mainly on the comparative advantages of national economies at the international level, expressed as export volumes. However, for the growth processes and economic development of nations, production for domestic demand will likely play an important role beyond production devoted to exports, since the latter only covers 14.0% of gross global output (see Figure II).

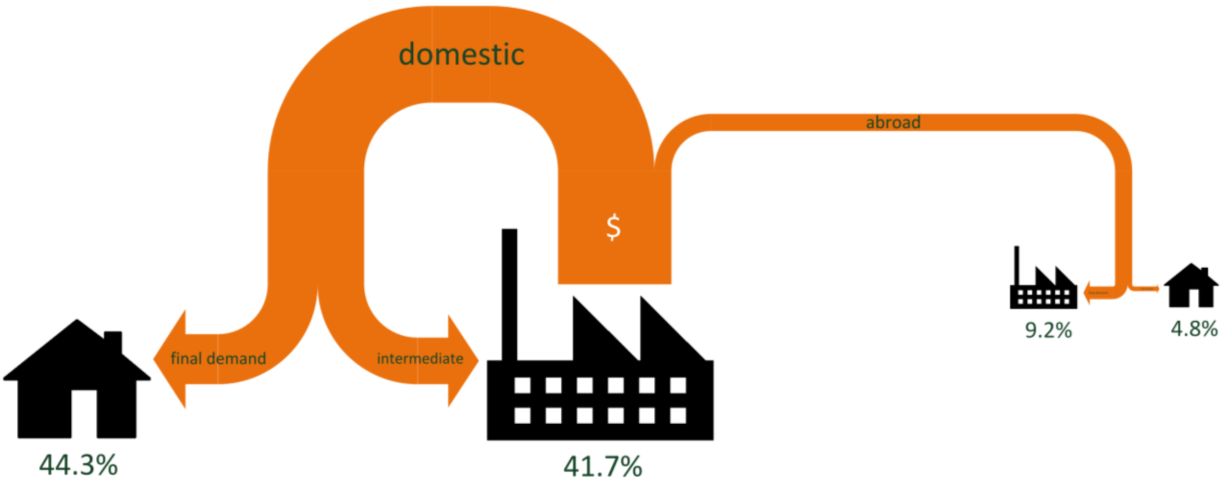


FIGURE II

Destinations of economic output on the global scale

Global shares of gross economic flows from industries' output into domestic final demand, domestic industries, foreign industries and foreign final demand. The picture is mainly identical when being based on net flows, i.e., value-added, while it varies strongly across regions and sectors. Based on the GTAP 8.1 Data Base (see below) for the year 2007.

A second asymmetry is at play, which relates to the highly different degrees of tradability of sectors (think of electricity, raw milk and education compared to coal, crops and electronic equipment) and other distorting influences, such as (politically motivated) international trade agreements and tariffs. Among other effects, this sectoral asymmetry is reflected in the observation that the exported fraction (from total output) of global sectors tends to be highly different across sectors and lower for larger sectors (see scatter points Figure III). In addition, this feature varies strongly across regions (see error bars in Figure III). All of these facts would obviously bias conclusions if international trade data were considered instead of *total* output for the purpose of empirically investigating sectoral economic development.

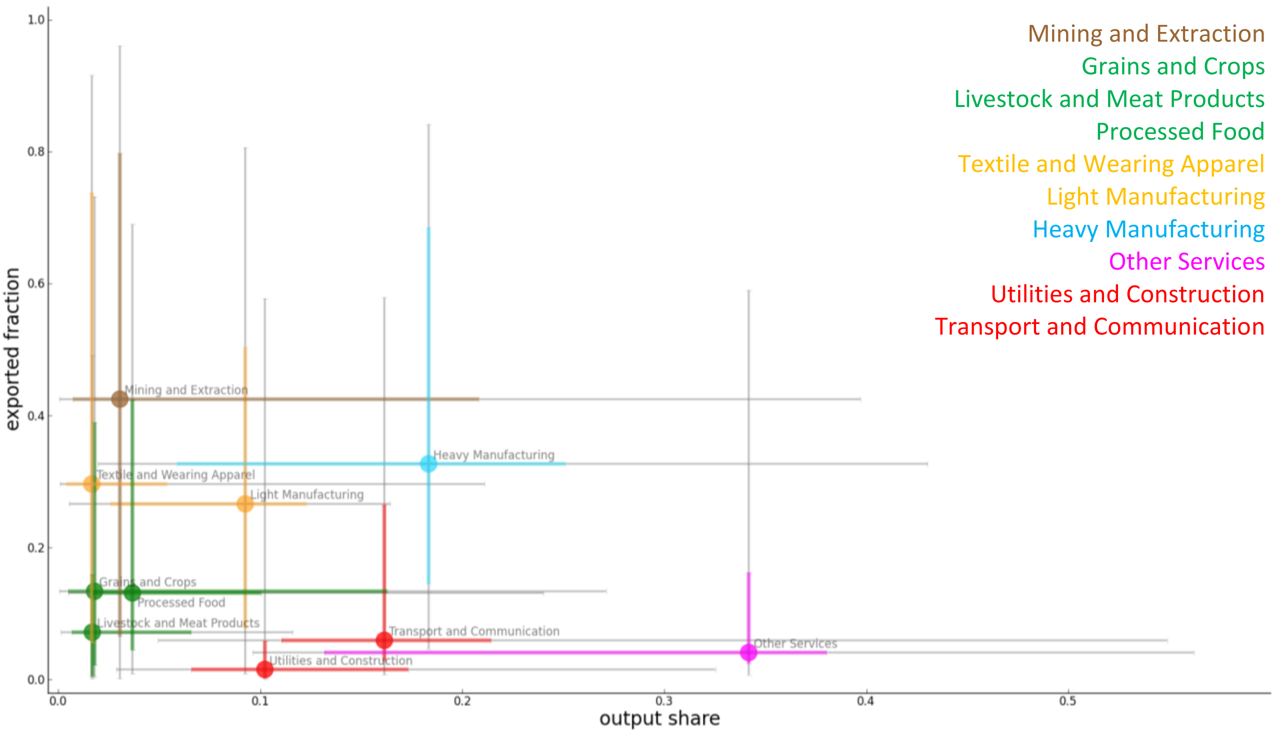


FIGURE III

Variability of export and output sizes by sector groups across regions

Fraction of output exported against sectoral output shares at the global level in 2007 for the sector groups given by the GTAP 8.1 Data Base (see the full sector list in Supplementary Table S1). The grey error bars indicate the minimum and maximum values over the ensemble of regions, which depicts an immense degree of heterogeneity among the regional sectors. The colored error bars indicates the 10^{th} to 90^{th} percentile range in the same spirit. Supplementary Figure S3 provides the full cloud of regional information for each sector group separately.

Moreover, it seems to be even more appropriate to assess sectors' strength by measuring their *value-added* instead of their *gross* output (cf. Johnson 2014) as the latter (due to multiple accounting) systematically inflates the sectors' size along global value creation chains. Analogous to the measurement of gross domestic product (GDP) the value-added $V(r, s)$ of sector s in region r based on total production can conveniently be deduced from MRIO tables by subtracting respective inputs $I(r, s)$ from outputs $O(r, s)$:

$$V(r, s) = O(r, s) - I(r, s) \quad (1)$$

Global MRIO datasets with considerably increased sectoral and regional resolutions have been published recently (Tukker and Dietzenbacher 2013). In this paper, we will build on data from sectoral inputs and outputs in (unadjusted) USD for 2007 from the Global Trade Analysis Project (GTAP) 8.1 Data Base (Narayanan, Badri, and McDougall 2012) because it offers the highest resolution (57 sectors, see Supplementary Table S1 for details) while offering a homogeneous sectoral classification across 134 regions. Also refer to the Supplementary Material for a detailed derivation of *O(r, s)* and *I(r, s)* from MRIO tables in general and from the GTAP 8.1 Data Base in particular.

II.B. Similarity of Sectors

In the spirit of the product space approach, we estimate the association between all pairs of sectors across the ensemble of regions. We consider two sectors to be similar (in this particular context) if their relative shares within the national economies correlate with each other. In other words, if two chosen sectors are relatively strong in the same group of regions and relatively weak in another group, they are considered to be positively related. In the opposite case, negative correlations occur. We stress that this nonparametric approach considers the entire set of available regions. Relying on correlation instead on conditional probability (as in (Hidalgo et al. 2007; Hidalgo and Hausmann 2009)) takes into account information from those regions that would not have met the condition and eliminates the risk of very small sample sizes after such conditioning. In any case, this cross-sectional approach implicitly assumes the given set of countries to be statistically representative in terms of their developmental stages. We validate this assumption using data across time.[2] Finally, we focus on the evaluation of the linear cross-correlation of the rank order (the so-called Spearman correlation coefficient). This is because (i) nonlinear measures of association (e.g., mutual information) would require either additional assumptions (e.g., regarding the functional relationship) or parameters, and (ii) the evaluation of the Spearman correlation coefficient is robust against outliers. Furthermore, it captures a more general feature because it tests the strength of the monotonic relationship. The Spearman correlation coefficient $P(s_1, s_2)$ of sectors s_1 and s_2 is calculated across all regions, r, as

$$\mathrm{P}(\mathrm{s}_1, \mathrm{s}_2) = \frac{\sum_r [(⟦\hat{\mathrm{V}}(\mathrm{r},\mathrm{s}_1)⟧ - \langle ⟦\hat{\mathrm{V}}(\mathrm{r}',\mathrm{s}_1)⟧ \rangle_{\mathrm{r}'}) * (⟦\hat{\mathrm{V}}(\mathrm{r},\mathrm{s}_2)⟧ - \langle ⟦\hat{\mathrm{V}}(\mathrm{r}',\mathrm{s}_2)⟧ \rangle_{\mathrm{r}'})]}{\sqrt{\left[\sum_r (⟦\hat{\mathrm{V}}(\mathrm{r},\mathrm{s}_1)⟧ - \langle ⟦\hat{\mathrm{V}}(\mathrm{r}',\mathrm{s}_1)⟧ \rangle_{\mathrm{r}'})^2\right] * \left[\sum_r (⟦\hat{\mathrm{V}}(\mathrm{r},\mathrm{s}_2)⟧ - \langle ⟦\hat{\mathrm{V}}(\mathrm{r}',\mathrm{s}_2)⟧ \rangle_{\mathrm{r}'})^2\right]}} \tag{2}$$

with $\langle \cdot \rangle_{\blacksquare} = \sum_{\blacksquare} \langle \cdot \rangle / \sum_{\blacksquare} 1$, based on rank-ordered ($⟦\cdot⟧$) intra-regional sectoral shares of value-added $\hat{V}(r, s)$, which is equivalent to an intra-regional normalization (thus bringing economies of different sizes to the same scale):

$$\hat{\mathrm{V}}(\mathrm{r}, \mathrm{s}) = \frac{\mathrm{V}(\mathrm{r},\mathrm{s})}{\sum_{s'} \mathrm{V}(\mathrm{r},\mathrm{s}')} . \tag{3}$$

The ensemble of the correlation values of all sectoral pairs s_1 and s_2 may be regarded as an association matrix, P, where the correlation coefficient $P(s_1, s_2)$ forms the matrix element at row s_1 and column s_2. This matrix is (i) square, (ii) symmetric (by definition)

$$\mathrm{P}(\mathrm{s}_1, \mathrm{s}_2) = \mathrm{P}(\mathrm{s}_2, \mathrm{s}_1) \qquad \forall \mathrm{s}_1, \mathrm{s}_2 , \tag{4}$$

[2] The reason why we do not use available time series datasets for the main analysis is because they have lower regional and/or sectoral resolutions that would lead to even less detailed results.

reflecting the fact that the symmetry of the arguments of the correlation coefficient, and (iii) its elements have values between -1 and 1.

We assess the significance of the association values by applying a nonparametric, Monte Carlo-based bootstrapping technique (Efron 1979). For this purpose, we shuffle the elements of vectors $\hat{V}(r, s_1)$ and $\hat{V}(r, s_2)$ independently and re-calculate their Spearman correlation. By doing this sufficiently often[3], we get the randomized distribution of correlation coefficients from which we take the upper and lower quantiles $P_\alpha^+(s_1, s_2)$ and $P_\alpha^-(s_1, s_2)$ that correspond to a given significance level, α, such that the fraction of α of the randomized values lie outside the two-sided quantile-based interval. This corresponds to the null hypothesis of uncorrelated vectors $\hat{V}(r, s_1)$ and $\hat{V}(r, s_2)$, i.e. $P(s_1, s_2)=0$, *while preserving the distribution of the vectors' elements*. Empirical association values within that interval are considered insignificant and are excluded from further analysis:

$$\mathrm{P}_\alpha{}^*(\mathrm{s}_1, \mathrm{s}_2) = \begin{cases} 0, & \mathit{if}\ P(\mathrm{s}_1, \mathrm{s}_2)\ \epsilon\ [\mathrm{P}_\alpha^-(\mathrm{s}_1, \mathrm{s}_2), \mathrm{P}_\alpha^+(\mathrm{s}_1, \mathrm{s}_2)] \\ \mathrm{P}(\mathrm{s}_1, \mathrm{s}_2), & \mathit{else} \end{cases} . \tag{5}$$

In general, the interval will depend on the particular pair of sectors. For the Spearman correlation coefficient, however, the sample size (i.e. the number of regions), is the only determining quantity, which remains constant in this study, thus:

$$\mathrm{P}_\alpha^+(\mathrm{s}_1, \mathrm{s}_2) = -\mathrm{P}_\alpha^-(\mathrm{s}_1, \mathrm{s}_2) = \Pi_\alpha \qquad \forall \mathrm{s}_1, \mathrm{s}_2\ . \tag{6}$$

[3] We will base the bootstrap on 10^7 randomized samples.

We chose the bootstrap approach (as opposed to, e.g., a *t*-value-based significance test) as it generalizes the significance test towards arbitrary marginal distributions of the data under consideration – instead of the special case of uniform marginal distributions. In the literature it is debated which transformation of the data is appropriate (Bahar, Hausmann, and Hidalgo 2014; Gathani and Stoelinga 2013), given highly different intrinsic scales and shapes of the marginal distributions, i.e. the intra-regional sectoral shares $\hat{V}(r, s)$.

II.C. Community Structure of the Similarity Network

The association matrix, P, (as given in Section II.B) can be interpreted as the adjacency matrix of a network, where the sectors form the nodes and their respective pairwise correlation coefficients form weighted, undirected links (with insignificant links being filtered out). Originating from empirical data, the similarity network's connectivity structure is expected to be neither fully regular nor purely random but, rather, to carry valuable information, e.g., on the nodal/sectoral communities from the link/similarity structure.

From the large variety of community detection algorithms, we choose the one based on "edge betweenness" (Girvan and Newman 2002), which takes the weights of the links (cf. Eq. (5)) into account. The main idea is that inter-community links are identifiable by their higher relevance for the shortest paths between the randomly chosen nodes. These bottleneck-like links are then considered to connect different communities and, thus, deterministically impose a division in the network. The algorithm leads to a hierarchical clustering of nodes (Kaufman and Rousseeuw 1990), which can be expressed in a dendrogram that shows the agglomerative grouping of the nodes. Given this hierarchy, a criterion is needed to select informative divisions of the network. Out of the different available measures that are capable of measuring how good inter- and intra-community connectivity information is captured by a chosen community set, we restrict the analysis to the value of modularity *m* (Newman 2004), which compares the frequency of intra- and inter-community links. High values of *m* indicate a good division of the network and allow us to choose the most informative community sets from those given by the hierarchical clustering dendrogram.

II.D. Directionality of Inter-Sectoral Connections

Due to the symmetry of the correlation coefficient (2) with respect to its arguments, our similarity network, which is defined by the adjacency matrix *P*, remains undirected. However, we are interested in the extent to which the sectoral balance of two sectors relates to the overall economic level. In other words, which one of the two sectors adjacent to a chosen link becomes relatively stronger when a national economy is growing? We proxy this tendency by *cross-sectionally*[4] evaluating the cross-correlation coefficient $D(s_1, s_2)$ of sectors s_1 and s_2 between the logarithms of the strength ratio $\hat{V}(r, s_2) / \hat{V}(r, s_1)$ and GDP per capita $G(r)$:

$$\mathrm{D}(s_1, s_2) = \frac{\sum_r \left[\left(\log\frac{\hat{V}(r,s_2)}{\hat{V}(r,s_1)} - \langle \log\frac{\hat{V}(r',s_2)}{\hat{V}(r',s_1)} \rangle_{r'} \right) * \left(\log G(r) - \langle \log G(r') \rangle_{r'} \right) \right]}{\sqrt{\left[\sum_r \left(\log\frac{\hat{V}(r,s_2)}{\hat{V}(r,s_1)} - \langle \log\frac{\hat{V}(r',s_2)}{\hat{V}(r',s_1)} \rangle_{r'} \right)^2 \right] * \left[\sum_r \left(\log G(r) - \langle \log G(r') \rangle_{r'} \right)^2 \right]}} \tag{7}$$

where

$$\mathrm{G}(r, s) = \frac{\sum_{s'} \mathrm{V}(r,s')}{\mathrm{C}(r)} \tag{8}$$

[4] In fact, in Eq. (7) we consider only those regions where the two sectors are sufficiently large, i.e. $\mathrm{V}(r, s_i) > 0.05/\sum_s 1 \quad \forall i \epsilon \{1,2\}$. This leads to different sample sizes for each sectoral pair, which induces different significance thresholds. The (adaptive) bootstrapping approach (cf. Section II.B) accounts for this.

and *C(r)* is the total population taken from the GTAP 8.1 Data Base. Analogous to the link weights, *P*, we perform a significance test that leads to a two-sided significance interval bounded by $D_\beta^+(s_1, s_2)$ and $D_\beta^-(s_1, s_2)$, which corresponds to a significance level, β. We note that the logarithmic strength ratio leads to an anti-symmetry of *D* with respect to its arguments:

$$\mathrm{D}(s_1, s_2) = -\mathrm{D}(s_2, s_1) \qquad \forall s_1, s_2 \, . \tag{9}$$

This additional information may then be used as a complementary link attribute that induces directionality as follows: If a directionality value is significantly positive for a given sectoral pair, i.e., $D(s_1, s_2) \geq D_\beta^+(s_1, s_2)$, the corresponding significant, yet undirected link $P_\alpha{*}(s_1, s_2) = P_\alpha{*}(s_2, s_1)$ between those two sectors is reduced into an uni-directed one in two steps. First, the undirected link is converted into two opposing directed links. Second, the directed link pointing from s_2 to s_1, which carries the negative value $D(s_2, s_1)$, is dropped, i.e., $P_\alpha{*}(s_2, s_1) := 0$. This leaves one directed link, which points from sector s_1 to sector s_2 and, thus, towards a (statistically) positive average effect on the overall GDP per capita when the mutual sectoral balance is altered in favor of sector s_2. In the case of a significant link (in terms of *P*), which exhibits no significant directionality information (*D*), the undirected link remains unchanged, thus expressing the absence of a significant impact on the overall GDP per capita by a changing sectoral balance.[5]

III. RESULTS

[5] We strengthen that our approach (cf. Eq. (7)) measures a correlation, i.e., its sign expresses the direction of the relationship, while its absolute value is a coeffient of determination. The latter should not be confused with the *size* of the GDP change when the sectoral balance is altered. This effect size could be measured by performing a regression instead.

III.A. Structure of the Similarity Network

Applying the previously presented method (cf. Section II.A and II.B) to input-output data from the GTAP 8.1 Data Base results in the correlation matrix shown in Figure IV. The main diagonal elements are 1 by definition (zero-lag auto-correlation), resulting in (trivial) sectoral self-links, that are disregarded in the rest of the article.

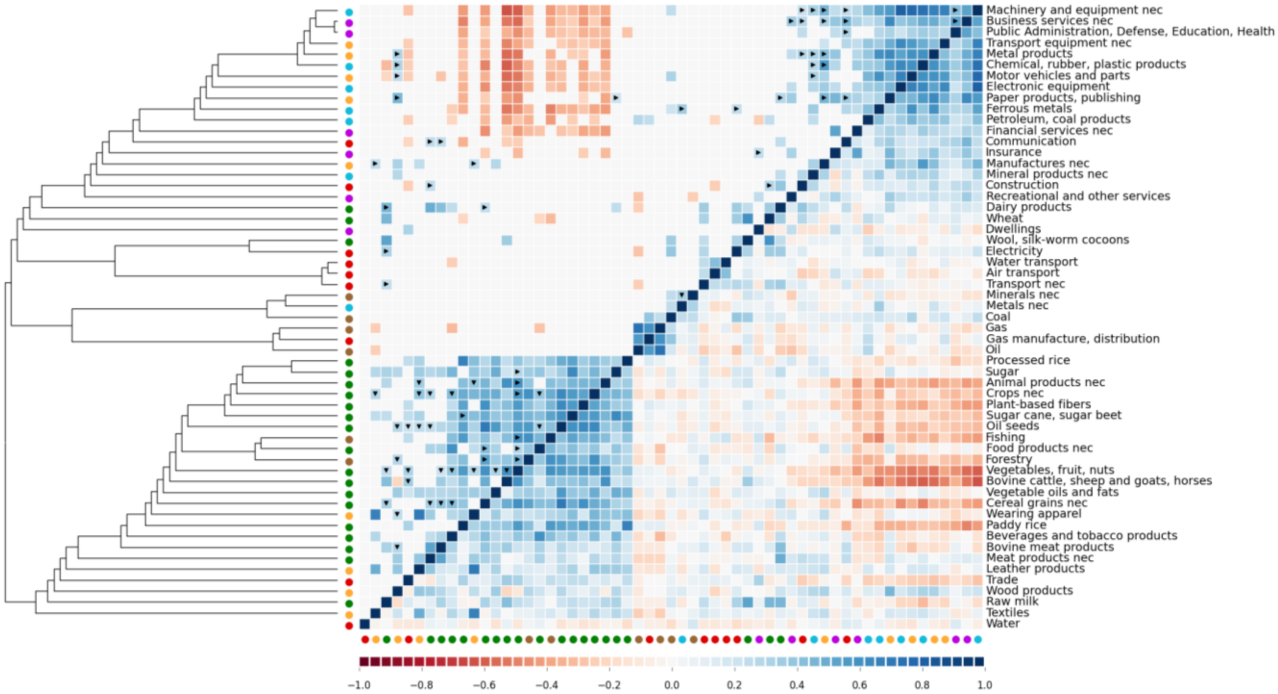


FIGURE IV

Similarity matrix of sectors and its corresponding community structure

The color-coded Spearman correlation matrix P of intra-regional sectoral value-added shares across regions based on the GTAP 8.1 Data Base for 2007. Insignificant association elements (at $\alpha=0.01$) are neglected in the upper triangle of the generically symmetric matrix. The rows and columns correspond to the annotated sectors and are sorted according to the hierarchical clustering, which is depicted by the dendrogram on the left. The colored dots represent (exogenously given) sector groups, as in Figure III. The small triangles indicate links where both, $P(s_1, s_2)$ and the directionality attribute $D(s_1, s_2)$ (at $\beta=0.01$) are significant. For a chosen link, the triangles point from the sector that gets relatively weaker to the sector that gets relatively stronger when overall GDP per capita increases.

This undirected, link-weighted network exhibits a link density—i.e., the number of present links with respect to the maximal possible number of links—of $\rho=0.312$ at a significance level of $\alpha=0.01$ (i.e., $\Pi_\alpha=0.222$). We would have expected 1% of the links to be present for the case of the null hypothesis of entirely uncorrelated sectors. However, the empirical link density of 31.2%—which consists of positively (21.7%) and negatively (9.5%) weighted links—indicates that the vast majority of links do not arise by chance, but instead incorporate valuable information. We note that at this significance level, the network fragments into two separated components. At this significance level the "water" sector has no significant connections to other sectors at all and is therefore disconnected from the rest of the network, which forms a so-called giant component.

The dendrogram to the left depicts the result of the quantitative detection of the community structure based on the positively weighted significant links while respecting the link weights (cf. Section II.C). Cutting the dendrogramm at the value of maximal modularity ($m=0.443$, which indicates a rather good division of the network; (Newman 2004); cf. Supplementary Figure S7) forces the giant component to collapse into four disjointed communities of different sizes: two dominant ones, two smaller ones. These become apparent as a remarkable block diagonal structure when re-arranging the rows and columns of the correlation matrix accordingly. The communities are also prominent when the network is embedded in a two-dimensional plane where all insignificant links as well as all links representing negative associations (i.e., $P_\alpha^*(s_1, s_2)<-\Pi_\alpha<0$) have been omitted (see Figure V). Without exception, the latter ones – as expected – connect sectors of different communities (see Supplementary Figure S5), confirming the endogenous community structure.

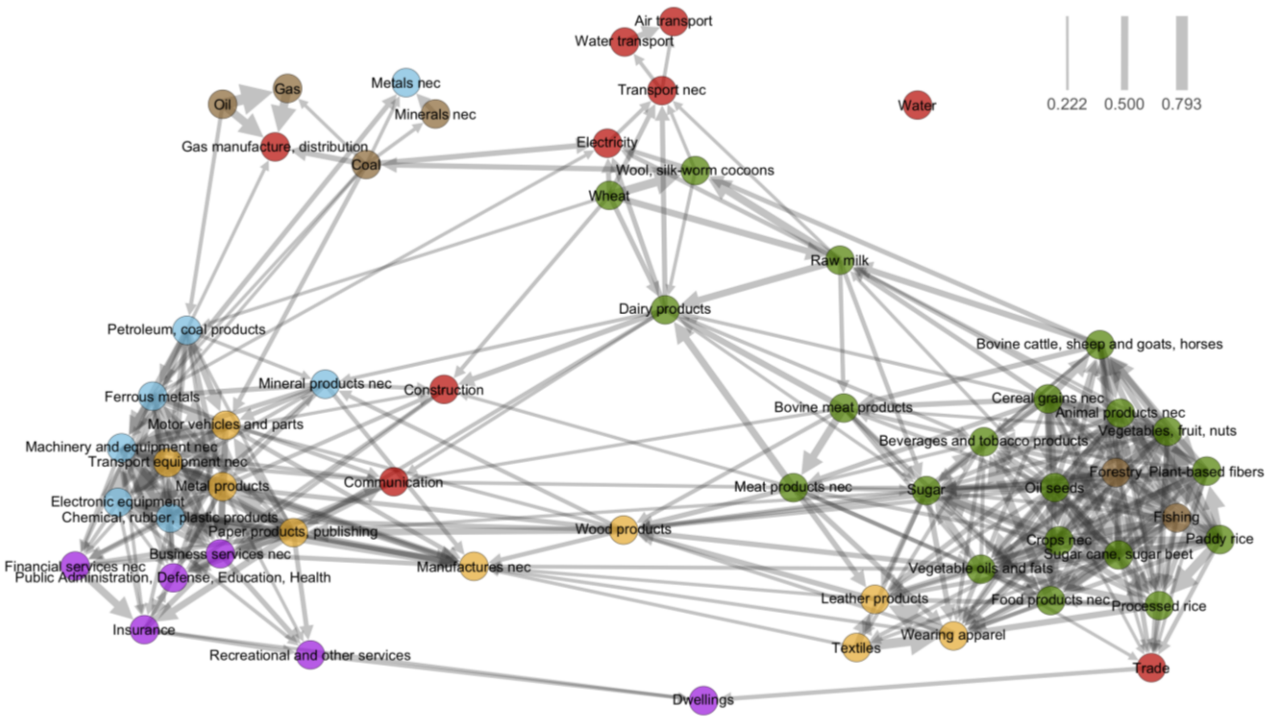


FIGURE V

The similarity network of economic sectors

Force-directed embedding (Kamada and Kawai 1989) of the similarity network of sectors based on value-added (same specifications as in Figure III). The directed links point towards the sector which gets relatively stronger when overall GDP per capita increases (cf. Section II.D). See Supplementary Figures S5 and S6 for further perspectives onto this network. The nodes were colored according to the sectoral groups taken from the GSC2 classification (cf. Supplementary Table S1). The width of the links indicates their weight $P_\alpha*(s_1, s_2)$, see legend.

The first major community consists of nearly all agricultural and food sectors including the agriculture-related extractive sectors "forestry" and "fishing", plus the textile-related industries. This community is opposed by the group of heavy and light manufacturing sectors, as well as the service sectors. The two smaller communities consist of (i) fossil fuel extraction sectors ("oil", "coal" and "gas") plus mineral and metal extraction, and (ii) transport sectors (land, water, air) plus the "electricity" and the "wool, silk-worm cocoons" sector. It is interesting that extractive resource industries are rather separated from the other sectors and form a separate community, thus highlighting their ambiguous role in development processes (van der Ploeg 2011). We emphasize that the strong agreement of the exogenously given sectoral grouping and the network's communities is non-trivial as the applied algorithms are agnostic about any sectoral classification scheme.

The hierarchical clustering algorithm necessarily assigns every sector to exactly one community, but the members of the larger two communities clearly take different positions *within* their communities (cf. Figure V). The "wood products" sector, for example, is peripheral to the (mainly) agricultural community and has comparably strong links with the nodes of both major communities, which results in a bridge-like position between the two groups. The same is true for the "coal" sector, which connects the major two communities with the resource community. Some sectors play a more prominent role with respect to inter-community connections; they are somehow situated *in between* and serve as bridges between communities. Therefore, engaging in an advanced community without having built those bridges seems difficult. We come back to this point in more detail in the discussion. The classical ladder models (Rosenstein-Rodan 1943; Rostow 1959) suggest a shift away from agricultural sectors, via light and heavy industries, towards the service sectors. In light of the similarity *network* we could now speak of a climbing wall (formed by sectors) where multiple routes are conceivable instead of a unique and linear sequence (as implied by a *one-dimensional* ladder), however, constrained by specific (sectoral) bottlenecks[6].

In order to determine whether or not the image of a climbing wall is appropriate, we tested our intuition on how the sectoral shares of countries evolve on the similarity network during growth processes. Putting the directionality information (as proposed in Section II.D) onto the previously

[6] Note that the presented findings are robust when including less significant and, therefore, lower-weighted links in the analysis or when we apply another well-established algorithm for the detection of the hierarchical clustering of nodes ("fast and greedy" algorithm, (Clauset, Newman, and Moore 2004)).

undirected links, we observe that all inter-community links point from the community containing the mainly agricultural and low-tech sectors to the community with the high-tech sectors. In particular, the links adjacent to the bridging sectors *entirely* point from the low-tech community to the high-tech industries (cf. Figures IV, V and Supplementary Figure S6). Hence, shifts in this direction imply increases in national GDP per capita.

Clearly, one might challenge a number of the links – i.e. pairwise correlation coefficients – as spurious regarding possible transitivities and/or common drivers. We emphasize that any conditioning of these correlations would already carry (implicit) assumptions on (here) unobserved characteristics (e.g. think of climatic or geographical factors, or consumption preferences). Hence, our results represent the purely multi-sector value-added-based outcome, i.e. the phenomonology of inter-sectoral connections, while the causes—that are most likely of large number and captured in exogenous data—remain the prime subject of future research. We expect different very different combinations of drivers to be responsible for the observed structure of the inter-sectoral network.

III.B. Stability over Time

Having performed a cross-sectional analysis for the 2007 data, we now assess the temporal stability of our results. Based on the Eora MRIO database (Lenzen et al. 2013) that offers annual data on a set of 26 sectors across 189 regions, we repeat the former analysis for each year from 1990 to 2011 independently and obtain a sequence of similarity networks of the sectors, which can be considered an evolving sectoral network. We then assess the temporal stability of this evolving similarity network from two angles.

First, regarding the evolving weighted adjacency matrix, we find a rather stable picture. The maximum of the absolute intertemporal link weight variation across time t at significance level α,

$$\Delta P_\alpha^{abs}(\mathrm{s}_1, \mathrm{s}_2) = \max_t\left(\left|\mathrm{P}_{\alpha,t}(\mathrm{s}_1, \mathrm{s}_2) - \langle \mathrm{P}_{\alpha,t'}(\mathrm{s}_1, \mathrm{s}_2)\rangle_{t'}\right|\right), \tag{10}$$

evaluated for all links whose weights leapfrog the positive time-independent significance threshold $\Pi_\alpha^+=0.187$ (i.e., $\alpha=0.01$) for at least one timestep, is $max_{(s1, s2)}(\Delta P_{0.01}^{abs}(s_1, s_2))=0.403$. This value occurs for the link between "fishing" and "re-export and re-import" (cf. Supplementary Figure S9b), while more than 90% of those links show $\Delta P_{0.01}^{abs}(s_1, s_2)<0.242$. The maximum of the relative intertemporal link weight variation across time t at significance level α,

$$\Delta P_\alpha^{rel}(\mathrm{s}_1, \mathrm{s}_2) = \max_t\left(\left|\left(\mathrm{P}_{\alpha,t}(\mathrm{s}_1, \mathrm{s}_2) - \langle \mathrm{P}_{\alpha,t'}(\mathrm{s}_1, \mathrm{s}_2)\rangle_{t'}\right)/\langle \mathrm{P}_{\alpha,t'}(\mathrm{s}_1, \mathrm{s}_2)\rangle_{t'}\right|\right), \tag{11}$$

evaluated for all links whose weights are above the time-independent significance threshold at all times is $max_{(s1, s2)}(\Delta P_{0.01}^{rel}(s_1, s_2))=0.726$, which occurs for the link between "fishing" and "recycling" (see Supplementary Figure S9d). More than 90% (80%) of the links whose weights surpass the time-independent significance threshold for at least one timestep show $\Delta P_{0.01}^{rel}(s_1, s_2)<2.658$ (1.480). On average, across all links that are significant at least once, the relative fluctuation is $mean_{(s1, s2)}(\Delta P_{0.01}^{rel}(s_1, s_2))=0.728$, which is mainly induced by trends across time (see Supplementary Figure S8). The variations remain close to the presented values when (i) less significant links are allowed for, i.e., higher α-values, or (ii) choosing a time-dependent significance threshold by fixing the link density in time. Thus, the intertemporal fluctuations of link weights—neither the maximal nor the mean ones, neither in absolute nor in relative terms—do not exceed acceptable levels.

Second, regarding the sectoral communities, we also find a reasonable agreement within the different years. Fixing the number of desired communities at any value between 5 and 7, where the highest modularity values occur, and quantitatively comparing the detected community sets for all pairings of years, shows that successive years typically lead to similar results. The coincidence is strong even at bigger temporal distances (see Supplementary Figure S10). Therefore, we conclude that the previously found results are sufficiently robust.

As a last indication, we embed the evolving network on the 2D plane and let it dynamically re-arrange positions (see Section S-V.C for video). The overall picture remains the same over the entire time frame, which substantiates our results solely based on 2007 data. Furthermore, although the sectoral aggregation scheme differs between the GTAP 8.1 Data Base and the Eora MRIO database, the agreement between the two networks is striking (cf. Figure V and video): An agricultural domain is connected, via light manufacturing sectors, to the advanced technological and services sectors; in addition there forms a resource extraction cluster that is barely connected to the other communities. The same conclusions can be drawn when relying on data from the World Input-Output Database (Timmer et al. 2012).

IV. DISCUSSION AND CONCLUSION

Developing from an agriculture-based economy towards increasing shares of (advanced) manufacturing and, eventually, services seems to require engagement in very particular sectors first, most of which can be classified as light manufacturing (cf. Figure V). We refine and generalize an existing framework called the product space, but in contrast with previous work that identifies a core-periphery structure (Hidalgo et al. 2007), we find several distinct communities that can largely be categorized under the agricultural, manufacturing and service sectors. Our analysis gives strong indications that these sectors develop sequentially, with each sector's performance building on the other. We also identify a resource extraction cluster that is separate from the others. While this picture is not necessarily unexpected (it largely resembles the classification of the primary, secondary and tertiary sectors) it has some interesting implications for the understanding of development processes, particularly because we can identify specific “bridging” sectors that link those communities and, therefore, can be considered crucial for structural change.

Previous literature (Hausmann and Rodrik 2003) has identified countries with similar factor endowments that specialize in different *products,* which, in our view, does not necessarily conflict with our results. Rather, in the spirit of Hidalgo et al. (2007), we hypothesize that “capabilities” understood as a broader set of societal and/or firm-level skills, or to put in the words of Rodrik (2007, p. 103): the “[acquisition of] mastery over a broad range of activities” necessary to further foster economic development and growth are developed when building up “bridging” sectors. Hence, while products can be different, the underlying capabilities are likely the same, which explains the robust depiction of bridges when looking at the more aggregated data. Developing capabilities in the bridging sectors are, therefore, a prerequisite to building up sectors that require advanced skills, e.g. high-tech industry sectors and engagement in the insurance and banking sectors. This mechanism resembles backward linkages reminiscent of Hirschman. However, these are not limited to the firm level, but extend to a broader set of factors and capabilities that are relevant to development and growth at the level of the entire economy. Moreover, acquired capabilities, which could also be understood as manifestations of knowledge in a society, can be further combined and act as important drivers of growth (analogous to Weitzman 1998).

Our results connect to the usual understanding of structural change in various ways. In the existing literature (for a detailed review, see Herrendorf et al. 2014), structural change is thought to be driven by non-homothetic preferences and differences in sectoral productivities. The latter are usually taken as given and remain largely unexplained (Herrendorf et al. 2014). In the light of our results, we argue that productivity changes might be endogenous with structural changes in the economy, as the development of certain sectors fosters the development of specific factors that increase the productivity of the entire economy. In this sense, broader societal skills that are developed when building up bridging sectors could explain large differences in productivity between agricultural and other sectors (particularly in developing countries), which do not lead to the reallocation of labor to more productive manufacturing sectors (Gollin, Lagakos, and Waugh 2014). It could also explain why structural change does not always enhance growth (McMillan, Rodrik, and Verduzco-Gallo 2014), think for example of industrial policy pushing the "wrong" sectors. It also offers an explanation for unconditional convergence of (and in) manufacturing sectors, but not the entire economy as described by Rodrik (2013). In our interpretation, necessary capabilities or skills (e.g., property rights, infrastructure and minimum institutional quality) cannot be developed before the manufacturing sectors are developed. As soon as a specific set of capabilities is established (think of the quality of institutions in line with Acemoglu et al. 2005) further diversification and, thus, the acquisition of capabilities may no longer be required. Hence, institutional and other factors that could hinder the realization of allocating labor and capital where it can be used most productively are overcome. This could also explain why countries are found to specialize in certain sectors when they reach high development levels after having previously diversified their portfolios (Imbs and Wacziarg 2003).

This interpretation requires further empirical testing. One promising method is to combine the similarity network of sectors with an empirical set of capabilities, e.g., in the form of various development indicators. Our results could be driven by the selection of databases and their sectoral aggregation schemes. It is also worthwhile to point out that our results are based purely on historic data. New bridges could emerge in the future, e.g., regarding a transition towards service economies or the invention of new processes and products. These new bridges could arise from at least two sources: Either by a changing link structure, i.e., weakening and strengthening inter-sectoral connections, or by internal specialization of existing and/or the emergence of completely new (types of) sectors. However, we are confident that our findings are robust because they are stable in cross-sectional analyses over time (despite massive globalization of the world economy during the period studied) in different available datasets and with different aggregations (for sectors and regions). Nevertheless, it would be helpful to further disaggregate economic sectors in global MRIO datasets in order to create a more detailed picture in the future.

In terms of developmental strategies, identifying robust bridges for development that connect intrinsic clusters with potential positive externalities by "crossing them" has implications for policy options. Building on a long-standing debate (for a review, see Pack and Saggi 2006), our results could support directing industrial policy towards specific sectors during certain stages of development. However, rather than simply supporting clustering (to foster Marshallian externalities) for any given set of industries where a country has a comparative advantage, it seems highly important to support sectors that have the ability to bridge differences between communities. In addition, targeted sectors or industries should be sufficiently "close" to the current structure of a given economy. However, while having the ability to identify bridging sectors could ease the identification of potential sectors and/or industries as policy targets, it does not help with estimating the amount of potential spillovers to (i) other sectors and (ii) other societal capabilities. Hence, the exact framing of industrial policy would remain unclear.

Beyond the important implications for developmental strategies, our results hint at difficulties how to deal with global environmental challenges, particularly in terms of climate change mitigation. If manufacturing sectors are needed to develop capabilities that foster growth, we should expect that large parts of Sub-Saharan Africa and South Asia, which have huge populations, will need to industrialize their economies when aiming to develop. This would, in turn, require a massive increase in energy consumption given today's state-of-the-art technologies, which in the recent past, have been supplied primarily by relatively cheap fossil fuels, particularly coal (Jakob et al. 2014; Steckel et al. 2011). Interestingly, the climate change scenarios used by the Intergovernmental Panel on Climate Change (IPCC) (Edenhofer et al. 2014) often see developing countries stabilize their levels of final energy per capita despite massive economic growth (Steckel et al. 2013). If building up domestic industrial sectors is a necessary requirement for further development, as implied by our results, this stabilization seems to be rather unrealistic, or would require massive financial and/or technological transfers to developing countries (for a discussion on related problems, see Jakob et al. in press). Thus, mitigation strategies for developing countries would need to focus on technology transfers of low-carbon technologies and energy efficient production strategies in order to avoid carbon-intensive lock-ins, while simultaneously supporting the creation of production capabilities in bridging sectors.

BERLIN INSTITUTE OF TECHNOLOGY, MERCATOR RESEARCH INSTITUTE ON GLOBAL COMMONS AND CLIMATE CHANGE, AND POTSDAM INSTITUTE ON CLIMATE IMPACT RESEARCH

REFERENCES

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "Institutions as a fundamental cause of long-run growth," *Handbook of economic growth*, 1, (2005), 385–472

Aghion, Philippe, and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 60, (1992), 323–351

Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo, "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?," *Journal of International Economics*, 92, (2014), 111–123

Balassa, Bela, "Trade Liberalisation and “Revealed” Comparative Advantage," *The Manchester School*, 33, (1965), 99–123

Chen, Shaohua, and Martin Ravallion, "The Developing World is Poorer than We Thought, But No Less Successful in the Fight Against Poverty," *The Quarterly Journal of Economics*, 125, (2010), 1577–1625

Clauset, Aaron, Marc E. J. Newman, and Cristopher Moore, "Finding community structure in very large networks," *Phys. Rev. E*, 70, (2004), 066111

Diamond, Jared *Guns, Germs and Steel*. (New York: W.W. Norton & Co, 1997)

Edenhofer, Ottmar, Ramón Pichs-Madruga, Youba Sokona et al., *Summary for Policymakers, in: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014)

Efron, Bradley, "Bootstrap methods: another look at the jackknife," *The Annals of Statistics*, (1979), 1–26

Gathani, Sachin, and Dimitri Stoelinga, "Export Similarity Networks and Proximity Control Methods for Comparative Case Studies," *Journal of Globalization and Development*, 3, (2013), 1–25

Girvan, Michelle, and Marc E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, 99, (2002), 7821–7826

Gollin, Douglas, David Lagakos, and Michael E. Waugh, "The Agricultural Productivity Gap," *Quarterly Journal of Economics*, 129, (2014), 939–993

Grossman, Gene M., and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 58, (1991), 43–61

Grossman, Gene M., and Elhanan Helpman, "Trade, knowledge spillovers, and growth," *European Economic Review*, 35, (1991), 517–526

Hausmann, Ricardo, and Dani Rodrik, "Economic Development as Self-Discovery," *Journal of Development Economics*, 72, (2003), 603–633

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi, "Growth and Structural Transformation," *Handbook of Economic Growth*, 2, (2014), 855 – 941

Hidalgo, César A., Bailey Klinger, Albert-L. Barabási, and Ricardo Hausmann, "The Product Space Conditions the Development of Nations," *Science*, 317, (2007), 482–487

Hidalgo, César, and Ricardo Hausmann, "The building blocks of economic complexity," *Proceedings of the National Academy of Sciences of the United States of America*, 106, (2009), 10570–10575

Hirschman, Albert O. *The Strategy of Economic Development*. Yale University Press, 1958)

Hummels, David, Jun Ishii, and Kei-Mu Yi, "The nature and growth of vertical specialization in world trade," *Journal of International Economics*, 54, (2001), 75–96

Imbs, Jean, and Romain Wacziarg, "Stages of Diversification," *The American Economic Review*, 93, (2003), 63–86

Jakob, Michael, Jan Christoph Steckel, Christian Flachsland, and Lavinia Baumstark, "Climate finance for developing country mitigation: blessing or curse?," *Climate and Development*, (in press)

Jakob, Michael, Jan Christoph Steckel, Stephan Klasen, Jann Lay, Nicole Grunewald, Inmaculada Martínez-Zarzoso, Sebastian Renner, and Ottmar Edenhofer, "Feasible Mitigation Actions in Developing Countries," *Nature Climate Change*, 4, (2014), 961–968

Johnson, Robert C., "Five Facts about Value-Added Exports and Implications for Macroeconomics and Trade Research," *Journal of Economic Perspectives*, 28, (2014), 119–42

Kamada, Tomihisa, and Satoru Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, 31, (1989), 7–15

Kaufman, Leonard, and Peter J. Rousseeuw *Finding groups in data: An introduction to cluster analysis*. (New York: Wiley, 1990)

Koopman, Robert, Zhi Wang, and Shang-Jin Wei, "Tracing Value-Added and Double Counting in Gross Exports," *American Economic Review*, 104, (2014), 459–94

Lenzen, Manfred, Daniel Moran, Keiichiro Kanemoto, and Arne Geschke, "Building EORA: A multi-region input-output database at high country and sector resolution," *Economic Systems Research*, 25, (2013), 20–49, v199.82

Lucas, Robert E., "On the mechanics of economic development," *Journal of Monetary Economics*, 22, (1988), 3–42

Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107, (1992), 407–437

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo, "Globalization, Structural Change, and Productivity Growth with an Update on Africa," *World Development*, 63, (2014), 11–32

Narayanan, G., Angel Aguiar Badri, and Robert McDougall. Center for Global Trade Analysis, Purdue University, 2012), "Global Trade, Assistance, and Production: The GTAP 8 Data Base," (2012)

Newman, Marc E. J., "Analysis of weighted networks," *Physical Review E*, 70, (2004), 056131

Pachauri, Rajendra K., "Climate change 2007. Synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report," (2008)

Pack, Howard, and Kamal Saggi, "Is there a case for industrial policy? A critical survey," *The World Bank Research Observer*, 21, (2006), 267–297

Przeworski, Adam *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge University Press, 2000)

Rodrik, Dani *One economics many recipes. Globalization, institutions and economic growth.* (Princeton, New Jersey: Princeton University Press, 2007)

Rodrik, Dani, "Unconditional convergence in manufacturing," *Quarterly Journal of Economics*, 128, (2013), 165–204

Romer, Paul M., "Increasing returns and long-run growth," *The Journal of Political Economy*, (1986), 1002–1037

Romer, Paul M., "Endogenous technological change," *The Journal of Political Economy*, 98, (1990), 71–102

Rosenstein-Rodan, Paul N., "Problems of industrialisation o eastern and south-eastern Europe," *The Economic Journal*, 53, (1943), 202–211

Rostow, Walt W., "The Stages of Economic Growth," *The Economic History Review*, 12, (1959), 1–16

Sachs, Jeffrey D., "Tropical Underdevelopment," *NBER Working Paper*, No. 8119, (2001), 1 – 40

Solow, Robert M., "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, 70, (1956), 65–94

Solow, Robert M., "Perspectives on growth theory," *The Journal of Economic Perspectives*, (1994), 45–54

Steckel, Jan Chr., Robert J. Brecha, Michael Jakob, Jessica Strefler, and Gunner Luderer, "Development without energy? Assessing future scenarios of energy consumption in developing countries," *Ecological Economics*, 90, (2013), 53–67

Steckel, Jan Chr., Michael Jakob, Robert Marschinski, and Gunner Luderer, "From carbonization to decarbonization? - past trends and future scenarios for China’s CO2 emissions," *Energy Policy*, 39, (2011), 3443–3455

Temple, Jonathan, "The New Growth Evidence," *Journal of Economic Literature*, 37, (1999), 112–156

Timmer, Marcel, AA Erumban, J. Francois, A. Genty, R. Gouma, B. Los, F. Neuwahl, O. Pindyuk, J. Poeschl, JM Rueda-Cantuche, and others, "The World Input-Output Database (WIOD): Contents, Sources and Methods," *WIOD Background document available at www.wiod.org*, (2012) Tukker, Arnold, and Erik Dietzenbacher, "Global Multiregional Input–Output Frameworks: An Introduction and Outlook," *Economic Systems Research*, 25, (2013), 1–19

Van der Ploeg, Frederick, "Natural resources: Curse or blessing?," *Journal of Economic Literature*, (2011), 366–420

Weitzman, Martin L., "Recombinant Growth," *Quarterly Journal of Economics*, 113, (1998), 331–360

World Bank *World Development Report 2014: Risk and Opportunity—Managing Risk for Development*. (Washington, DC: World Bank, 2013)

World Bank, *World Development Indicators,* available at: http://data.worldbank.org/indicator/

Supplementary Material for "On the Importance of Manufacturing Sectors for Economic Development and Growth"

S-I. Derivation of Regional Sectoral Value Added from MRIO Data

Let *Z(r, s, r', s')* denote the inter-industrial flow from sector *s* in region *r* to sector *s'* in region *r'*, and *Y(r, s, r')* the flow from sector *s* in region *r* into the final demand of region *r'*. These two flow matrices are the actual MRIO table. Then regional sectoral output *O(r, s)* is obtained as

$$\mathrm{O(r,s)} = \sum_{\mathrm{r'}}\sum_{\mathrm{s'}}\mathrm{Z(r,s,r',s')} + \sum_{\mathrm{r'}}\mathrm{Y(r,s,r')} \tag{12}$$

and the regional sectoral input *I(r, s)* as

$$\mathrm{I(r,s)} = \sum_{\mathrm{r'}}\sum_{\mathrm{s'}}\mathrm{Z(r',s',r,s)}. \tag{13}$$

S-II. Derivation of Regional Sectoral Value Added from the GTAP 8.1 Data Base

From the GTAP 8.1 Data Base the inter-industry flow matrix and the final demand flow matrix are obtained from the following quantities (all quantities at market prices):

VDFM(s_1, s_2, *r*)	Domestic purchases in region *r* of sector s_1 by sector s_2
VDPM(s, r)	Domestic purchases in region *r* of sector s by households
VDGM(s, r)	Domestic purchases in region *r* of sector s by government
VXMD(s, r)	Non-margin exports of sector *s* in region *r*
VST(s, r)	Margin exports of sector *s* in region *r*
VIFM(s, r)	Import purchases in region *r* by sector *s*.

While we could think of deriving the regional sectoral outputs *O(r, s)* and the regional sectoral inputs *I(r, s)* directly from those quantities, the margin exports *VST(s, r)* – that are non-zero only for the three transportation sectors (cf. Supplementary Table S1) and that deliver into the international transportation pool – need a particular treatment as they do not contain information on their use. They affect the input as well as the output matrice in a non-trivial manner (Peters, Andrew, and Lennox 2011). This is why we have to construct the entire MRIO table first and then reduce it to the input and output matrices (cf. Supplementary Section S-I). The construction of an MRIO table from the above listed quantities in turn requires certain assumptions (Peters et al. 2011).

S-III. Additional Figures and Tables

GRAINS AND CROPS
Paddy rice (PDR)
Wheat (WHT)
Cereal grains nec (GRO)
Vegetables, fruit, nuts (V_F)
Oil seeds (OSD)
Sugar cane, sugar beet (C_B)
Plant-based fibers (PFB)
Crops nec (OCR)
Processed rice (PCR)

LIVESTOCK AND MEAT
Bovine cattle, sheep and goats, horses (CTL)
Animal products nec (OAP)
Raw milk (RMK)
Wool, silk-worm cocoons (WOL)
Bovine meat products (CMT)
Meat products nec (OMT)

PROCESSED FOOD
Vegetable oils and fats (VOL)
Dairy products (MIL)
Sugar (SGR)
Food products nec (OFD)
Beverages and tobacco products (B_T)

TEXTILE AND WEARING APPAREL
Textiles (TEX)
Wearing apparel (WAP)

LIGHT MANUFACTURING
Leather products (LEA)
Wood products (LUM)
Paper products, publishing (PPP)
Metal products (FMP)
Motor vehicles and parts (MVH)
Transport equipment nec (OTN)
Manufactures nec (OMF)

HEAVY MANUFACTURING
Petroleum, coal products (P_C)
Mineral products nec (NMM)
Ferrous metals (I_S)
Metals nec (NFM)
Electronic equipment (ELE)
Machinery and equipment nec (OME)
Chemical, rubber, plastic products (CRP)

MINING AND EXTRACTION
Forestry (FRS)
Fishing (FSH)
Coal (COA)
Oil (OIL)
Gas (GAS)
Minerals nec (OMN)

UTILITIES AND CONSTRUCTION
Electricity (ELY)
Gas manufacture, distribution (GDT)
Water (WTR)
Construction (CNS)

TRANSPORT AND COMMUNICATION
Trade (TRD)
Transport nec (OTP)
Water transport (WTP)
Air transport (ATP)
Communication (COM)

OTHER SERVICES
Financial services nec (OFI)
Insurance (ISR)
Business services nec (OBS)
Recreational and other services (ROS)
Public Administration, Defense, Education, Health (OSG)
Dwellings (DWE)

Supplementary Table S1. Sectors and corresponding sector groups as listed in the GTAP 8.1 Data Base (GSC2 classification). The coloring was done by the authors for enhanced differntiation of the sector groups.

Agriculture (1)
Food & Beverages (4)

Fishing (2)
Mining and Quarrying (3)

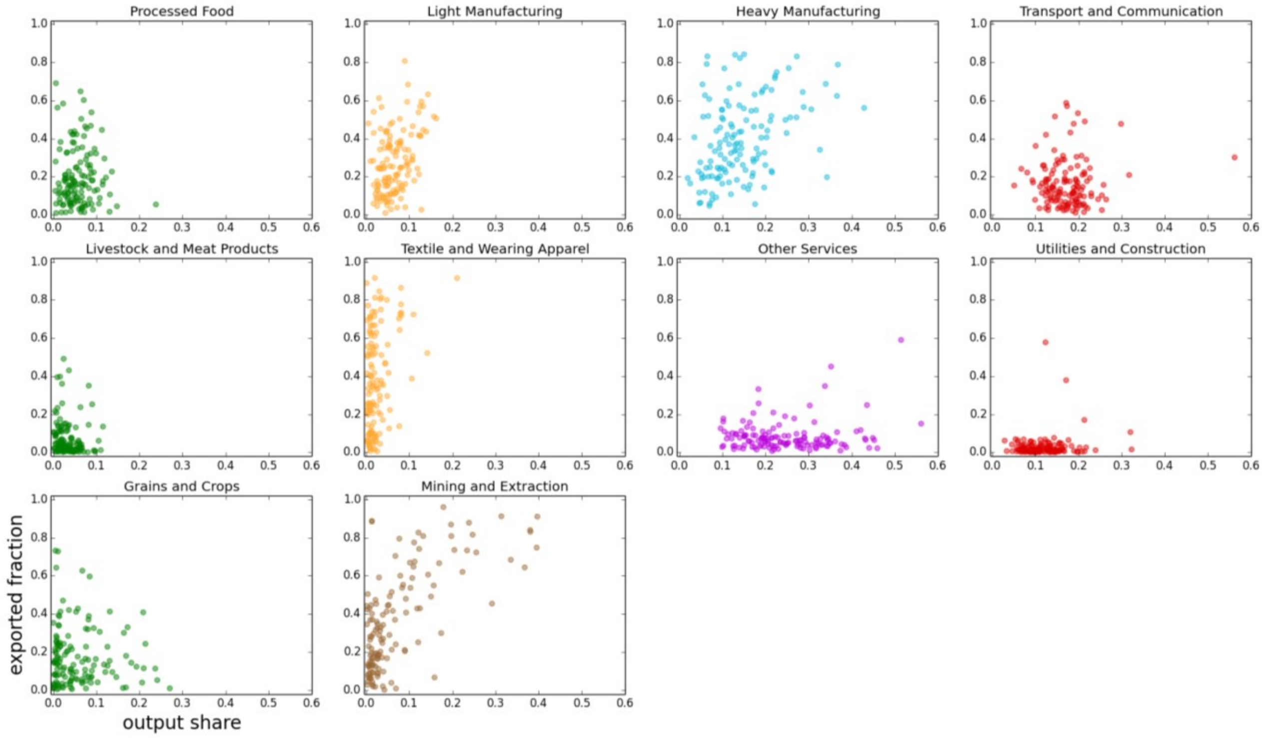
Textiles and Wearing Apparel (5)
Wood and Paper (6)
Metal Products (8)
Other Manufacturing (11)
Recycling (12)

Petroleum, Chemical and Non-Metallic Mineral Products (7)
Electrical and Machinery (9)
Transport Equipment (10)

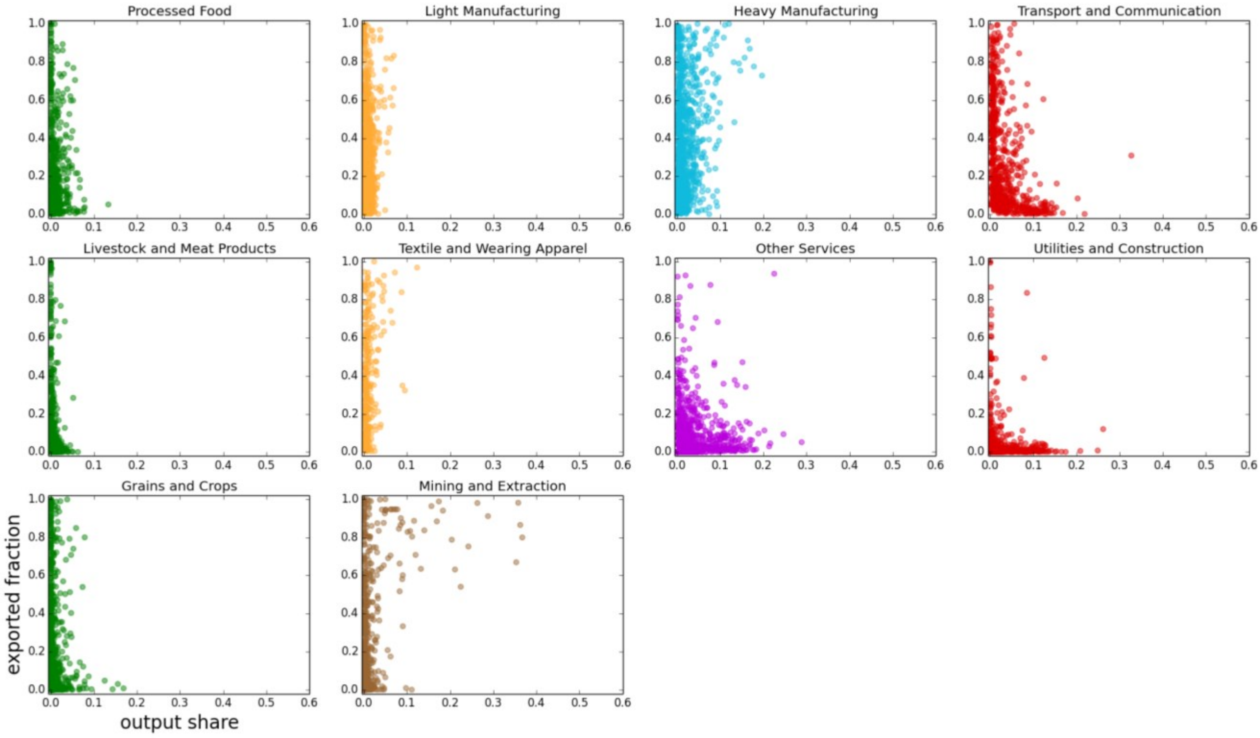
Electricity, Gas and Water (13)
Construction (14)
Maintenance and Repair (15)
Wholesale Trade (16)
Retail Trade (17)
Hotels and Restraurants (18)
Transport (19)
Post and Telecommunications (20)

Financial Intermediation and Business Activities (21)
Public Administration (22)
Education, Health and Other Services (23
Private Households (24)
Others (25)
Re-export & Re-import (26)

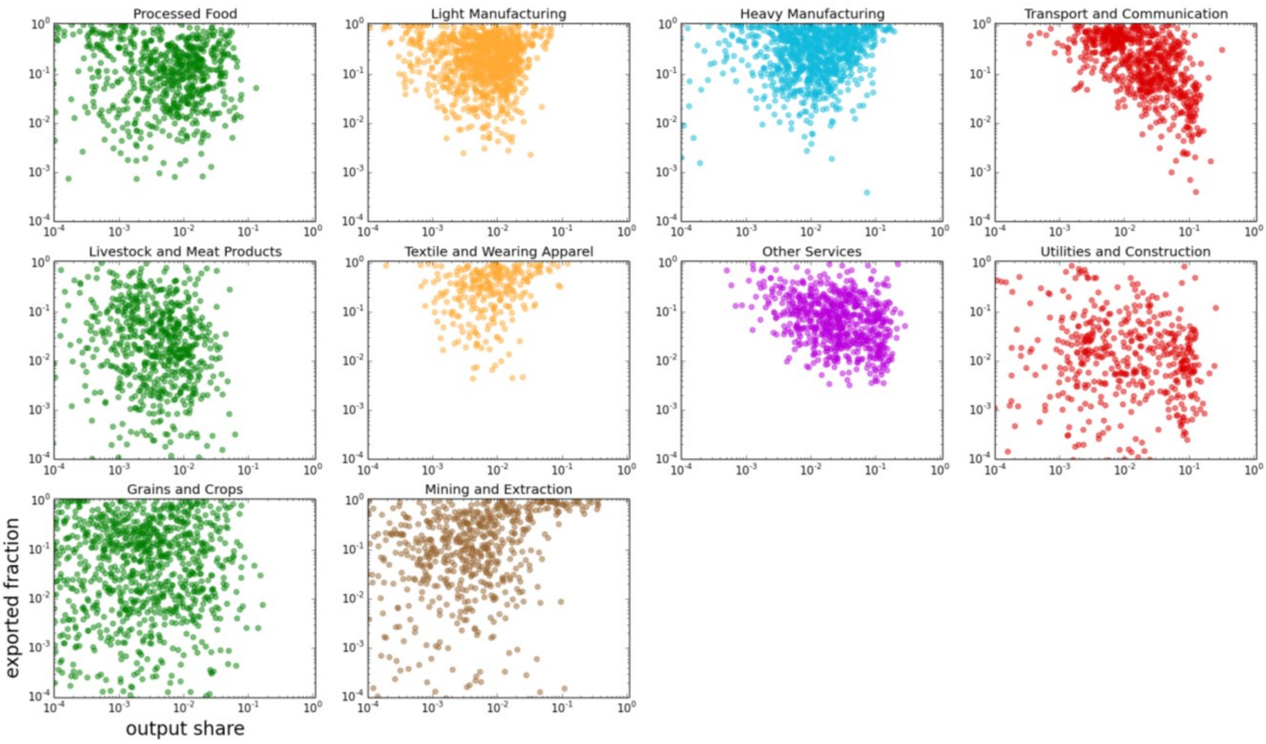
Supplementary Table S2. Sectors as listed in the Eora MRIO database. The coloring was done by the authors, following the spirit in Supplementary Table S1.



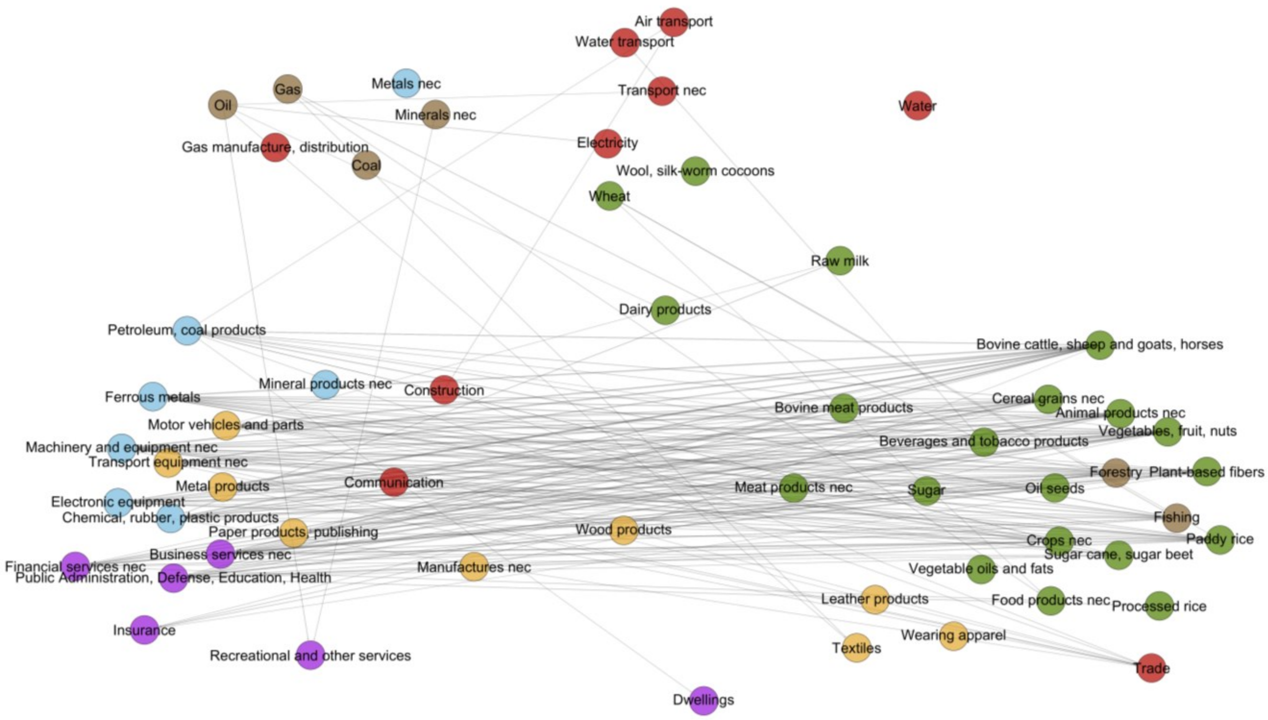
Supplementary Figure S3. The fraction of output being exported against intra-regional sectoral output share in 2007. Each tile shows a different sector group, where each scatter point represents a region. Figures S4a and S4b further resolve these points by decomposing them into the underlying sectors within that sector group. Sector groups as from the GTAP 8.1 Data Base, coloring was done according to the sector list in Supplementary Table S1.



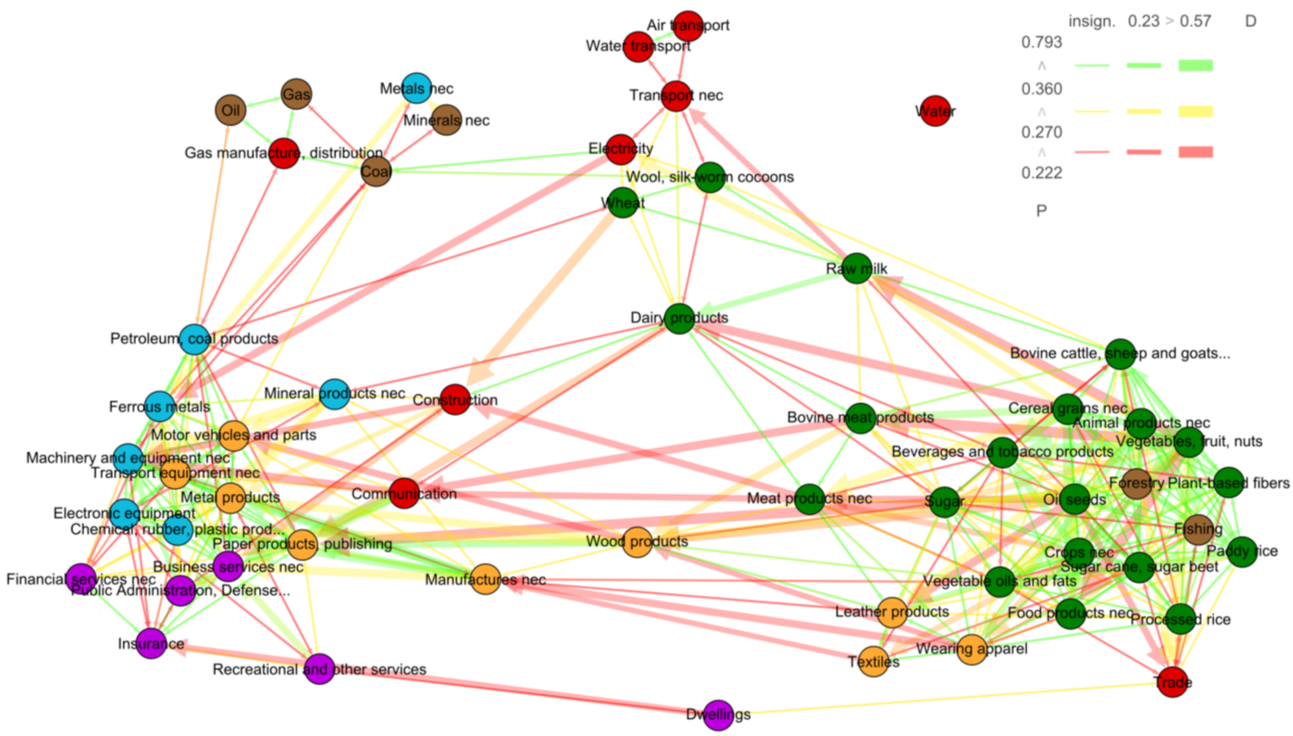
Supplementary Figure S4a. The fraction of output being exported against intra-regional sectoral output in 2007. Each tile shows a different sector group, where each scatter point represents a regional sector. Sector groups as from the GTAP 8.1 Data Base, coloring was done according to the sector list in Supplementary Table S1.



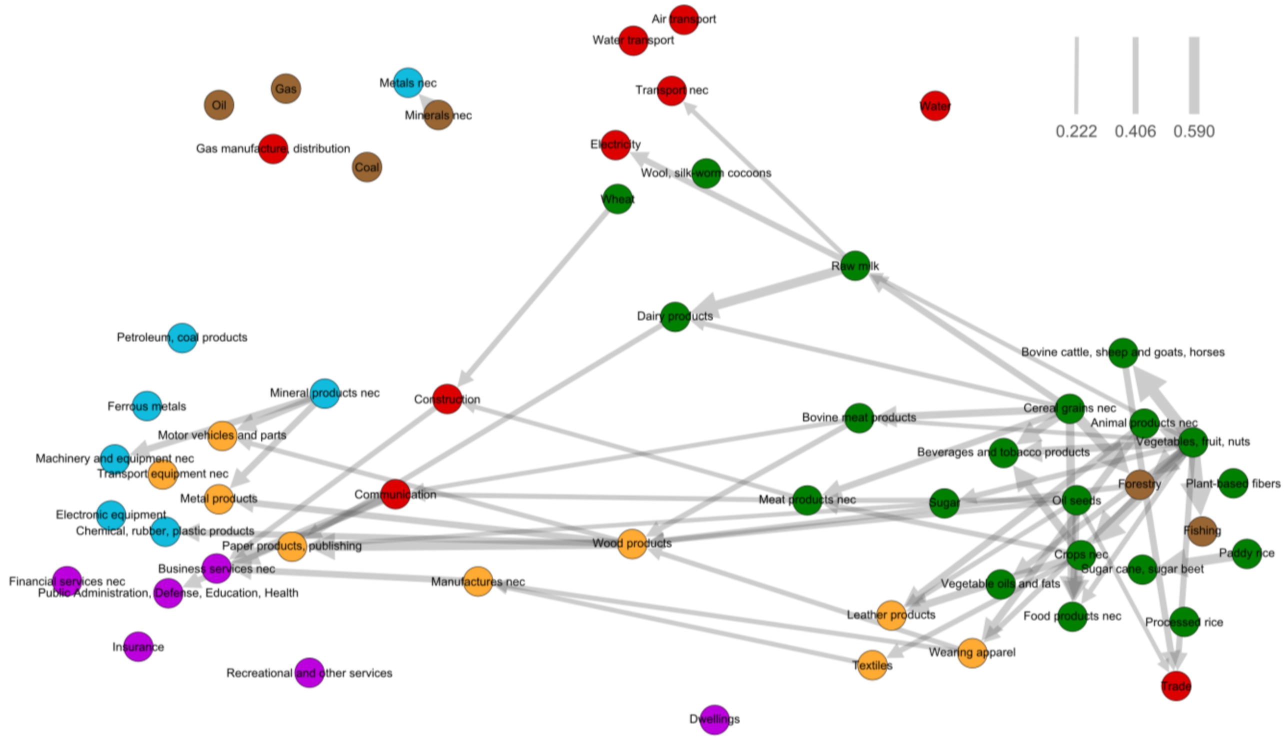
Supplementary Figure S4b. Same information as in Supplementary Figure S4a, but on a log-log scale.



Supplementary Figure S5. The significant (α=0.01), negatively weighted links of the similarity network of sectors. The placement of the nodes is kept fixed from that of Figure V.



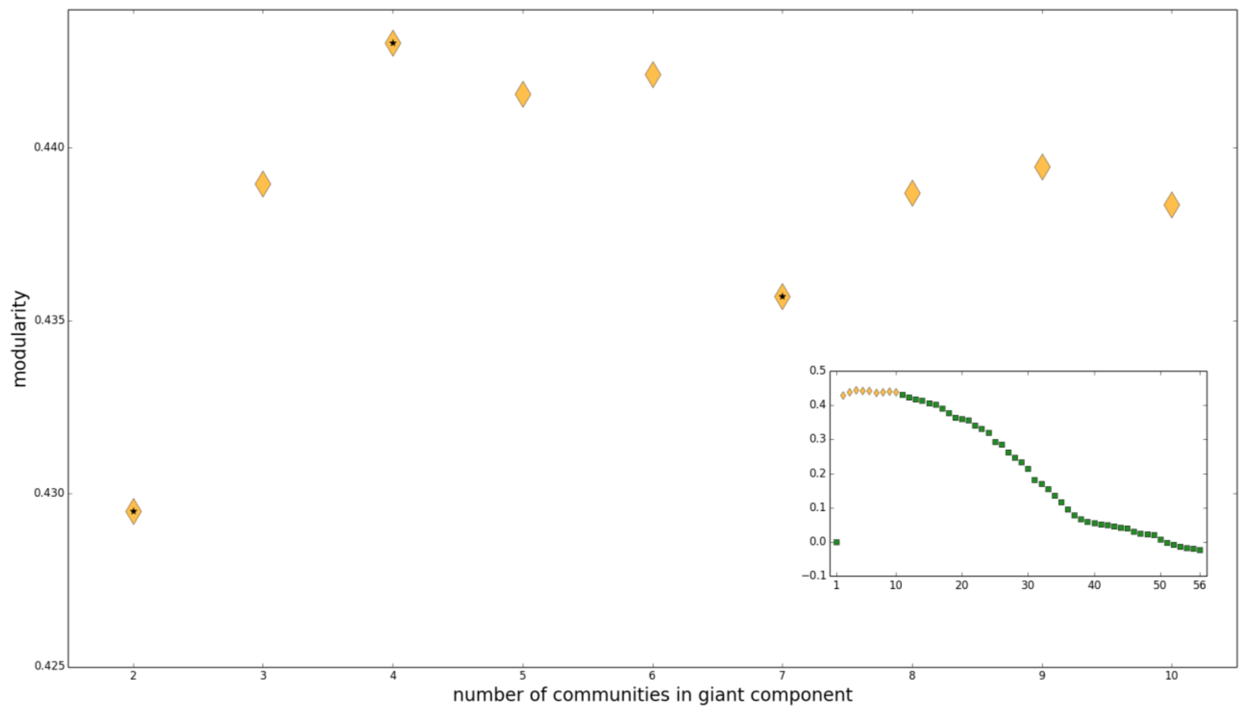
Supplementary Figure S6a. An alternative perspective onto the similarity network of sectors. Here, the width of the links indicates the directionality attribute $D_\beta*(s_1, s_2)$, while the color indicates the correlation attribute $P_\alpha*(s_1, s_2)$. Uni-directed links indicate a significance in both attributes (at α=0.01 and β=0.01). Bi-directional links (to be considered as one undirected link) indicate that only the correlation attribute is significant, while there is no significance in the directionality attribute. The placement of the nodes is kept fixed from that of Figure V.



Supplementary Figure S6b. Another perspective onto the similarity network of sectors. As in the main text, the width of the links indicates the correlation attribute $P_\alpha*(s_1, s_2)$. Only those links are shown that are significant in both attributes ($P_\alpha*$ at α=0.01 and $D_\beta*$ at β=0.01). The placement of the nodes is kept fixed from that of Figure V.

S-IV. Detailed Analysis of the Community Structure

The hierarchical clustering diagram (cf. Figure IV) allows for a more detailed analysis of the community structure. At each merge/split as indicated by the dendrogram, the corresponding community set can be evaluated regarding its information content (cf. Section II.C). In Supplementary Figure S7 we show the modularity values of all induced community sets (within the giant component, i.e. without counting the disconnected "water" node as its own community).



Supplementary Figure S7. Modularity *m* of the weighted similarity network of sectors against the number of communities within the giant component, i.e., not considering the disconnected node ("water") that forms another component/community. The modularity is calculated for each set of communities as given by the hierarchical clustering that is expressed in the dendrogram within Figure IV. Stars indicate the community sets discussed in the main text and in the SI. The inlay shows the entire data range, while the main plot focusses on the range of highest modularity values.

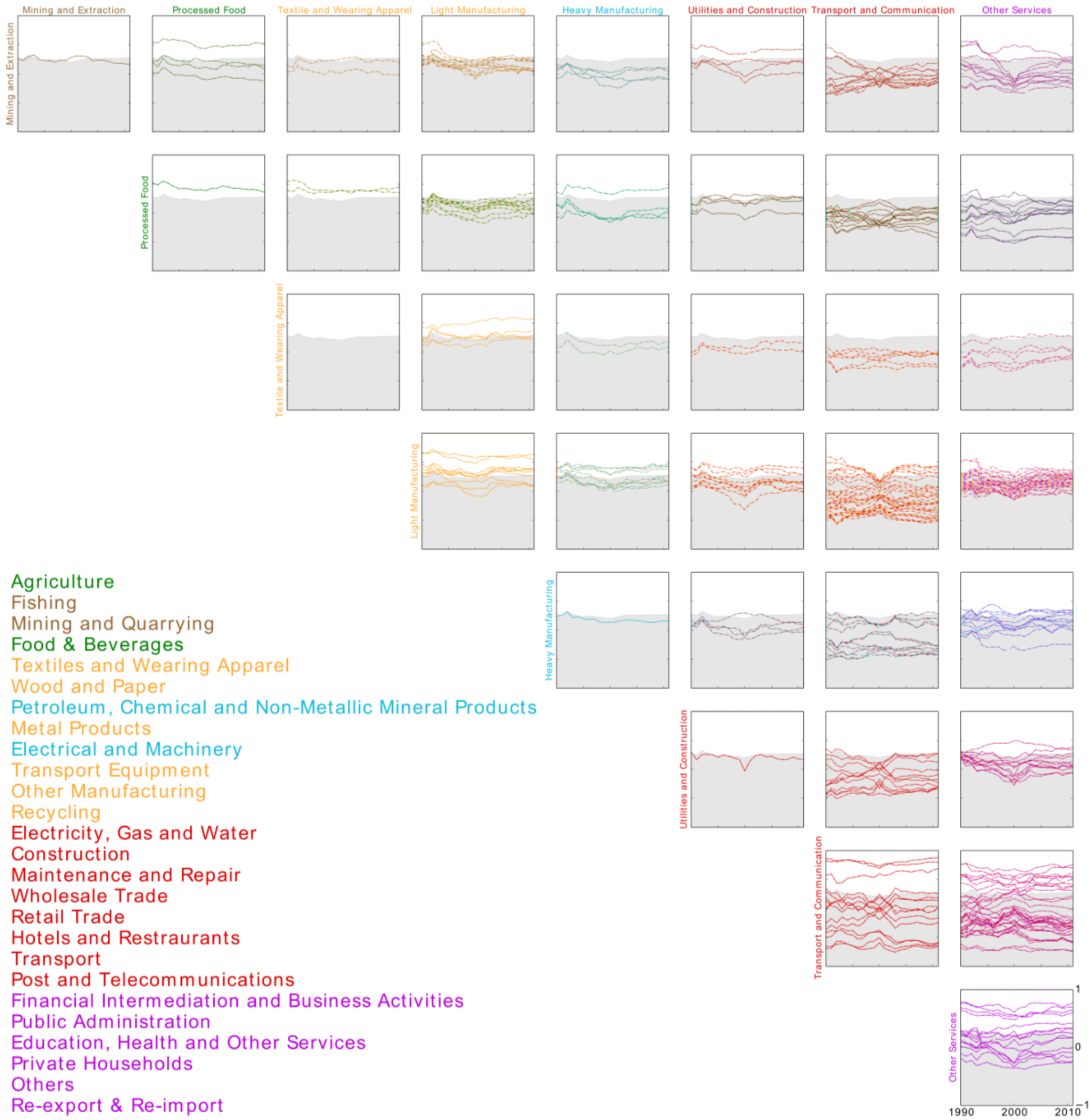
While there is no information (*m=0.00*) by choosing one community, i.e. the entire giant component as one group, the first split into who communities massively increases the modularity value (*m=0.423*, leftmost diamond). These two communities are formed by the two major communities that are discussed in the main text (cf. Section III.A) where the smaller two groups around the (i) fossil resource extraction and (ii) the transport sectors are members of the heavy manufacturing and services sectors community. As described in the main text, there is an information gain (*m=0.443*, middle diamond) when those two subsets separate from the high-tech community. At this point a maximum of information is obtained. After three subsequent splits ("dwellings", "wheat", "textiles") the modularity decreases to a local minimum (*m=0.436*, rightmost diamond) expressing the ambiguity of the lastly separated sector "textiles". This is an expression of the fact that this sector forms one of the bridges (from the mainly agricultural to the "manufactures, not else classified" sector and, thus, towards the light and heavy manufacturing community.

S-V. Detailed Results of the Robustness Analysis

In this section we provide more background information on the temporal stability of the similarity network. As described in the main text, we have repeated our cross-sectional analysis for each year in the Eora MRIO database (Lenzen et al. 2013). In the following we analyze the resulting evolving similarity network.

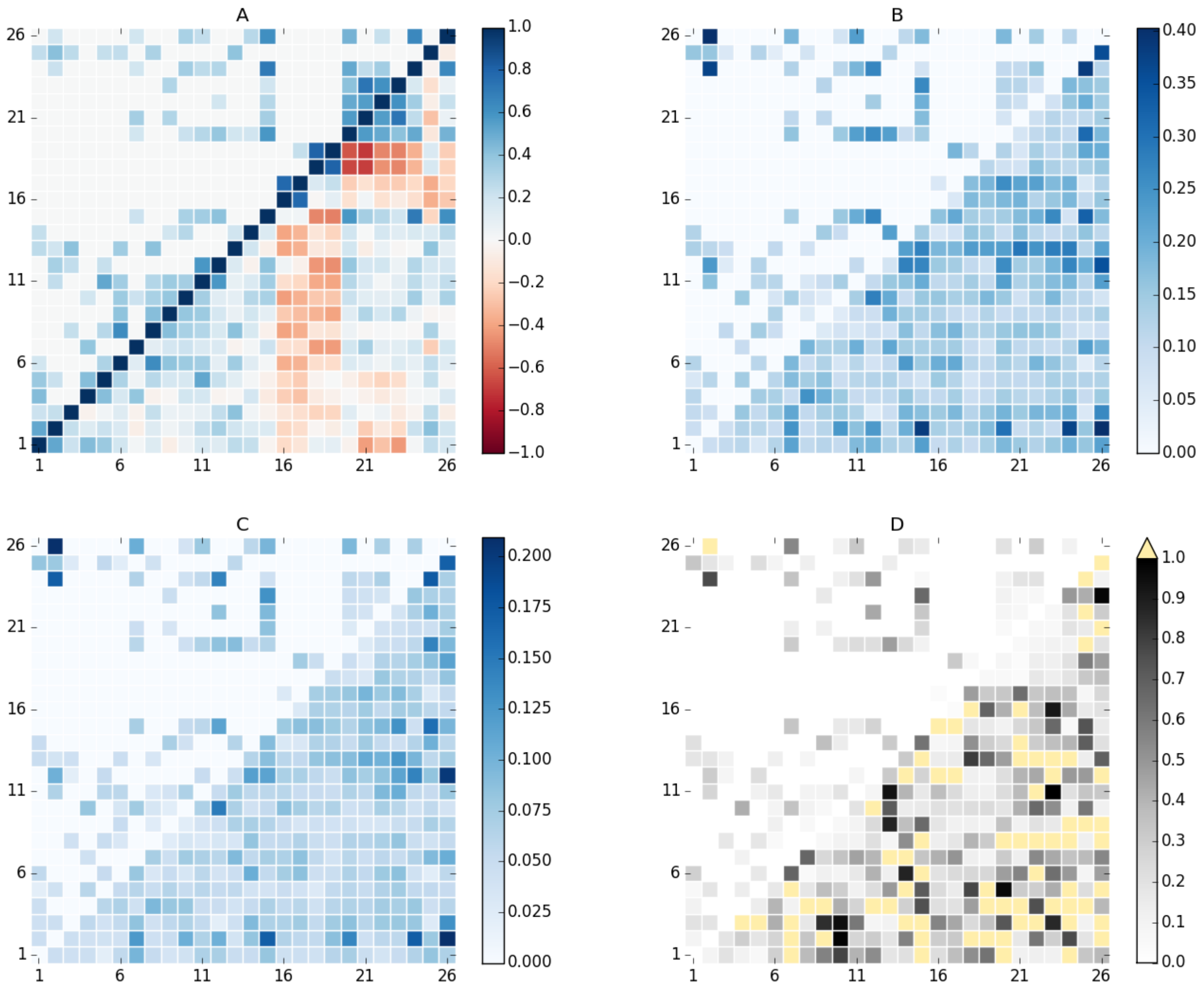
S-V.A. Temporal Stability of the Similarity Network

As the nodes, i.e., sectors remain identical across time the evolution of the network is given by the evolving link attributes. While we concentrated on the distribution of these attributes in the main text, we now want to study the evolution of all intersectoral links, i.e., sectoral pairs individually. However, we restrict ourselves to the analysis of the evolving association values, $P(t, s_1, s_2)$, which now become time-dependent as the intra-regional sectoral shares of value added, $\hat{V}(t, r, s)$, are now dependent on the time, t. As there is a large number of 325 (unordered) pairs of the 26 sectors, we group these pairs according to the exogenously given sectoral groups (as indicated in Supplementary Table S2) to which the two nodes of a link/pair belong to (see Supplementary Figure S8).



Supplementary Figure S8. Evolution of all pair-wise sectoral associations from 1990 to 2011. Each tile includes the group of links that mediates between the two sector groups (cf. Supplementary Table S2) as given by row and column, respectively. Hence, intra-group-wise links are shown in the tiles along the diagonal, while inter-group-wise links are shown in the tiles off the diagonal. Each line represents the evolution of a link over time, $P(t, s_1, s_2)$ (from 1990 to 2011). The lines are bicolorly dashed according to the two sector groups the corresponding link connects. The grey-shaded area represents the 80th percentile range of all link weights at each point in time, respectively, such that 20% of the links are outside this interval for each point in time.

While the majority of links does not fluctuate strongly across time, there are a number of exceptions, either in form of a trend along the temporal dimension or as temporary peaks/drops. There is a tendency that less stable links are adjacent to a sector either from the "utilities and construction", the "transport and communication" or the "other services" group. Among those, the most strongly fluctuating links connect to the "mining and extraction" or the "light manufacturing" group. Beyond this holistic perspective we may reduce the temporal information towards individual link's quantities that express its stability (see Supplementary Figure S9).



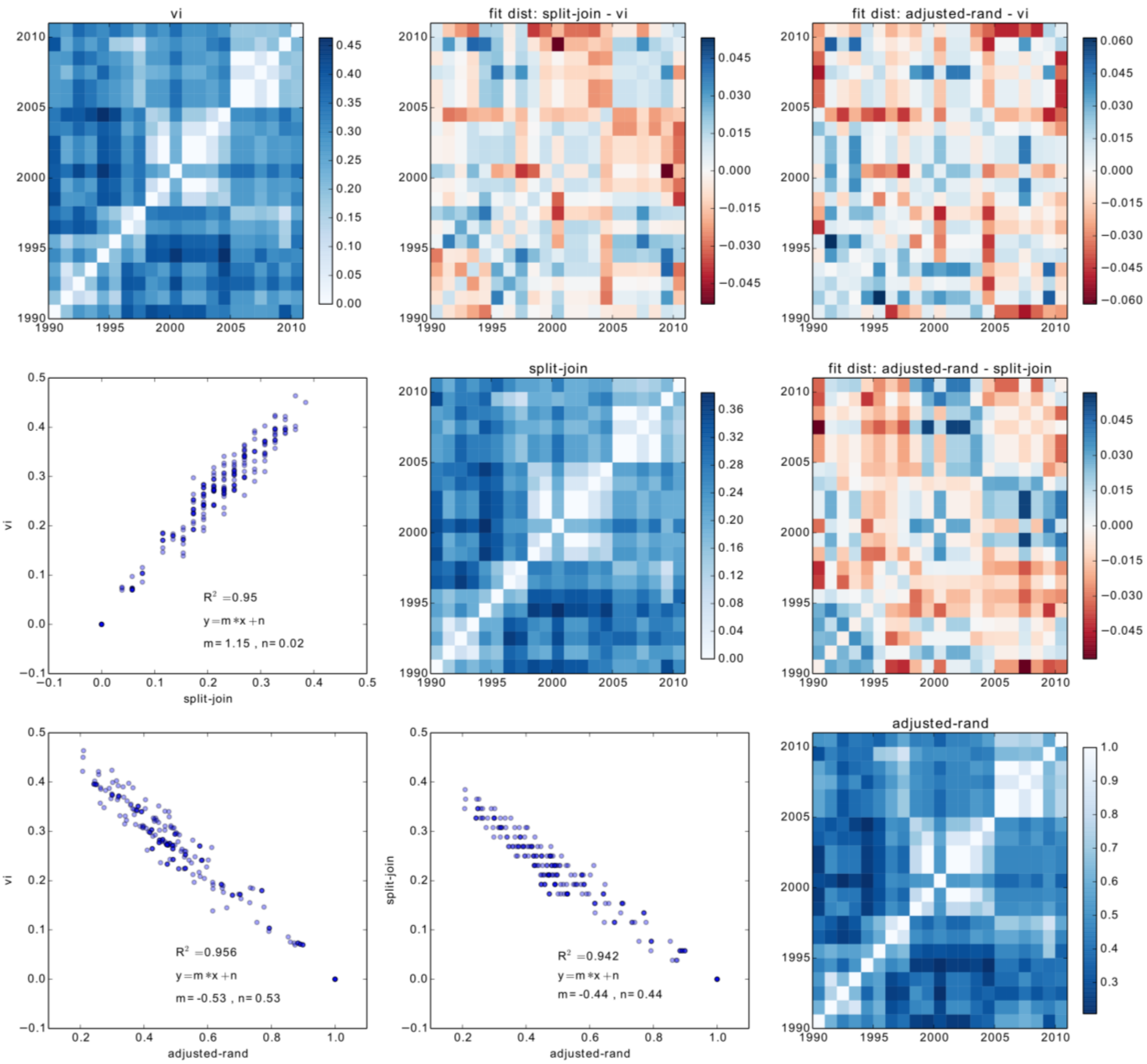
Supplementary Figure S9. Parameters of the intertemporal variation of links. Rows and columns of each subplot refer to the 26 sectors of the Eora MRIO database, and color-coded matrix elements thus refer to single links. For clarity, the upper triangle part of each subplot shows the quantities only for the strongest 20% of links (in terms of the intertemporal average), while the lower triangle gives the entire information of the underlying symmetric matrices. Order of the sectors as given in Supplementary Table S2. (A) The intertemporal arithmetic average of link weights. (B) The maximal deviation of link weight evolutions from the corresponding intertemporal average, as defined in Eq. (10). (C) Standard deviation of link weight evolutions around the corresponding intertemporal average. (D) Fluctuation strength, as defined in Eq. (11).

Beyond the arguments given in the main text, we see that in general the intertemporal standard deviation (C) as well as the maximal deviation (B) of evolving link weights is rather low for the broad majority of links. Even when rescaling each link variation with the corresponding intertemporal average, the resulting fluctuation strengths (D) are small, at least for the 20% most significant links. Still, there are some notable exceptions, which are mostly adjacent to the "fishing", the "private household" and the "re-export and re-import" sector. However, regarding our main results (a strong endogenous network hierarchy inducing communities that are connected by certain bridging sectors), these rather particular exceptions do not alter our conclusions.

S-V.B. Temporal Stability of the Sectoral Community Sets

Beyond a stability analysis on the level of individual links, we evaluate to what extent the evolving community structures that result from the evolving overall link (weight) structures are robust across time. There are two complementary approaches to this. While one can compare the entire hierarchical clustering, in form of their corresponding dendrograms, we choose to compare the induced community sets at a fixed number of communities as they are of main interest in our discussion.

For the comparison of community sets different approaches have been proposed. Here, we apply on the “variation of information” algorithm (Meila 2007), the “split-join distance” algorithm (Van Dongen 2000) and the “adjusted Rand index” algorithm (Hubert and Arabie 1985) (see Supplementary Figure S10).



Supplementary Figure S10. A two-layer comparison. First, the coherence of community sets across time is checked. Second, three different concepts to check the coherence of the evolving community sets is compared. The tiles on the main diagonal show the intertemporal comparison matrices for each of the three concepts, respectively. The tiles on the lower triangle show the pairwise comparison of two of the concepts, respectively. Within each of these tiles each scatter point's position is given by the year-year pair value of the two different concepts, respectively. The tiles on the upper triangle show the residuals when an ordinary least squares linear fit is subtracted from the scatter cloud of the chosen concept-concept pair. The parameters for the underlying hierarchical clustering is the same as in the main text (cf. Section III.B). For all comparisons we chose the clustering that yields five communities.

First, regarding the three tiles on the main diagonal, we find a strong qualitative agreement of the intertemporal community comparison matrices among the three comparison algorithms. This is confirmed by the strong correlations between the different measures (cf. tiles on the lower triangle, coefficient of determination of two chosen measures across all pairs of years is $R^2>0.94$) and the rather small residuals from an ordinary least squares fit of two measures, respectively. Without loss of generality we can thus restrict the discussion to one of the intertemporal community set comparisons, say the upper left one.

Firstly, we observe block diagonal patterns in the intertemporal plane expressing certain time windows of rather stable community sets. However, higher deviations occur when comparing years of different time windows. Since there is no a priori indication from which value on insufficient stability can be stated, we consider (i) maximum values of around 0.4 for the former two measrues, and (ii) minimum values of around 0.2 for the latter measure as reasonably small given that the former two measures are bounded between 0 (maximum concordance) and 1 (maximum deviation), and the latter measure bounded between -1 (maximum deviation) and +1 (maximum concordance). We tested this for different numbers of communities (from 5 to 7, cf. Supplementary Section S-IV) into which the hierarchical clustering result is split and came to the same conclusions.

S-V.C. Video of the Evolution of the Similarity Network

As a last perspective on the temporal stability of our findings, we embed the evolving similarity network on a 2D plane (as in Figure V) and let the nodes dynamically re-arrange their positions according to evolving edge weights as time lapses (see video, by the use of the python igraph package (Csárdi and Nepusz 2006) and Gephi (Bastian, Heymann, and Jacomy 2009)). To allow for a better comparison of the picture of the evolving network from the time-dependend but sectorally lower resolved dataset (provided by the Eora project) and the static dataset (provided by the GTAP), we here merged the "fishing" sector with the "agriculture" sector. Further, we only show links that are among the 20% strongest ones at least for 50% of the timesteps. At the very beginning we place the nodes manually to enable the reader a better look onto the sector names. After having activated the force-directed embedding (Kamada and Kawai 1989), we let the time window slide across the years.

The evolution of the network as depicted in the video shows a remarkable stability as time passes, which confirms the indications from the previous subsections and all of our conclusions from the static analysis at higher sectoral resolution.

Despite the different sectoral aggregation schemes of the static (cf. Supplementary Table S1) and the time series (cf. Supplementary Table S2) dataset the overall picture is mainly preserved. The similarity network is hierarchical, forming well separated communities where the agriculture-related sectors connect to light manufacturing industries, which in turn connect to heavy manufacturing and service sectors while the mining and extraction sectors take a peripheral position. We note that the video shows a previously not discussed bridge formed by the "petroleum, chemical and non-metallic mineral products", which is due to the very strong aggregation in the Eora MRIO database (combining ISIC Rev. 3 sub-sectors as diverse as "coke, refined petroleum products and nuclear fuel”, “chemicals and chemical products”, “rubber and plastics products” and “other non-metallic mineral products”). The same is true for the agriculture-related sectors in the Eora MRIO database, which are merged into two sectors (“agriculture” and “food & beverages”) instead of 20 different ones in the GTAP 8.1 Data Base.

REFERENCES

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," (2009)

Csárdi, Gábor, and Tamás Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, 1695, (2006)

Hubert, Lawrence, and Phipps Arabie, "Comparing Partitions," *Journal of Classification*, 2, (1985), 193–218

Kamada, Tomihisa, and Satoru Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, 31, (1989), 7–15

Lenzen, Manfred, Daniel Moran, Keiichiro Kanemoto, and Arne Geschke, "Building EORA: A multi-region input-output database at high country and sector resolution," *Economic Systems Research*, 25, (2013), 20–49, v199.82

Meila, Marina, "Comparing clusterings—an information based distance," *Journal of Multivariate Analysis*, 98, (2007), 873–895

Peters, Glen P., Robbie Andrew, and James Lennox, "Constructing an Environmentally-Extended Multi-Regional Input-Output Table Using the GTAP Databas," *Economic Systems Research*, 23, (2011), 131–152

Van Dongen, Stijn *Performance Criteria for Graph Clustering and Markov Cluster Experiments*. Technical Report, (National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, 2000)